PROCEEDINGS

OF

FIRST ASIAN TECHNOLOGY CONFERENCE IN MATHEMATICS

(ATCM)

(18 - 21 Dec 1995)

"Innovative Use of Technology for Teaching and Research in Mathematics"

The Association of Mathematics Educators
Singapore

1995

Participant Edition:
Proceedings of the First Asian
Technology Conference in Mathematics

published and distributed by

The Association of Mathematics Educators
c/o National Institute of Education
469 Bukit Timah Road
Singapore 1025
e-mail: fonghk@nievax.nie.ac.sg

The views expressed within this volume by individual authors do not necessarily represent those of the organising committee and the Association of Mathematics Educators.

Organising Committee
ATCM '95

Chairman	:	Dr Fong Ho Kheong
Committee Members	:	Dr Chong Tian Hoo, A/P Lee Peng Yee, Dr Lee Cho Seng, Dr Wang Jian Hua, Dr Ramakrishnan Menon, Dr Teo Beng Chong, Dr Rosalind Phang

International Programme Committee
ATCM

Chairman	:	Dr Yang Wei Chi
Committee Members	:	Prof John Mack, Prof Jeff Baxter, Prof Peter Jones, Prof Keith Jones, Prof Tang Ruifeng, Prof Wu Wenda, Prof Hidetsune Kobayashi, Prof Naoto Niki, Prof Kiyoshi Shirayanagi, Prof Nobuki Takayama, Prof Sheen Dongwoo, Prof Hassan b Said, Dr Chong Tian Hoo, A/P Lee Peng Yee, Dr Fong Ho Kheong, Prof Jen-Chung, Prof Simon C. Lin, Prof Jing Yu, Prof Robert Lopez, Prof Paul Wang

FORWARD

The First Asian Technology Conference in Mathematics (ATCM 95) is the first conference of this nature to be held in Asia. The Conference is organised by the Association of Mathematics Educators of Singapore, in conjunction with the Nanyang Technological University, National Institute of Education of Singapore and the Radford University of the USA.

In view of the success of conferences organised by ICTCM, the founders of the ICTCM, Bert Waits and Frank Demana, encouraged Wei-Chi Yang to organise such conference in Asia. Through some consultation and discussion with Tian-Hoo Chong, Peng-Yee Lee of the Nanyang Technological University, National Institute of Education, Singapore and Stephen Brown of TCI Software Research, the ATCM 95 was finally agreed to be first held in Singapore. The Conference Theme for the First Conference is **Innovative Use of Technology for Teaching and Research in Mathematics**.

There is always a quest for improvement in teaching mathematics at various levels, ranging from the primary to the tertiary. Enthusiasts have looked into new approaches of teaching and conducting research. Technology is one area which has great potential as technology knowledge seems to advance at greater pace than most of us expected.

The First ATCM will provide mathematics educators, computer specialists, technologist, researchers, policy makers and teachers with the opportunity to share and discuss the latest developments in their areas of specialisation. The Conference provides an avenue for the possibility of collaborating research among the participants.

There seems to be some major emphases in the current development and research in the area of technology in Mathematics. This is reflected on the papers submitted for presentation in this Conference. One would see ATCM to grow in three major areas:
(a) Pedagogy: Educators with pedagogical emphasis shall further develop the potential of technology in teaching and learning of mathematics and evaluate the impact on which it helps learner to acquire mathematical knowledge and solve problems.
(b) Computer Algebra: Researchers shall look further to investigate how to create more user-friendly software and develop algorithm efficient programs.
(c) Computational Mathematics: With the support of technology, mathematicians shall be able to make innovative conjectures and discover new theorems.

We look forward to this conference as a special occasion to exchange and consolidate ideas and practices in mathematics education with the use of technology. We hope that ATCM will go on year after year, and each year we may attract more enthusiasts and provide participants with new ideas through presentations, workshops or exhibitions.

Finally we would like to thank all the people and institutions that have provided us with excellent support to make this conference a success. In particular, special thank is given to the **Director of National Institute of Education** who sponsors us by allowing us to use the Bukit Timah Campus as the Conference venue. Special thanks are also given to the followings who have given us the financial support: the **British Council**, the **Lee Foundation**, **Association of Mathematics Educators**, **Waterloo Maple Software**, **Wolfram Research**, and **TCI Software**. We would also like to thank all exhibitors for their consents to use their technologies in the Conference. We sincerely appreciate colleagues who served on ATCM 95 organising committee, international programme committee, and as plenary speakers, workshop conductors and paper presenters.

Ho-Kheong Fong
Chair, ATCM 95
Organising Committee

Wei-Chi Yang
Chair, ATCM 95
International Programme Committee

December 1995

CONTENTS

COMPUTER ALGEBRA

Page

COMPUTATION MATHEMATICS

SEMI-AUTOMATED THEOREM PROVING

—

THE IMPACT OF COMPUTERS ON RESEARCH IN PURE MATHEMATICS

Marston Conder

Department of Mathematics, The University of Auckland

Private Bag 92019, Auckland, NEW ZEALAND

e-mail: conder@mat.auckland.ac.nz

1 Introduction

There is no doubt that the use of computers in recent years has revolutionised many branches of science, not the least of these being mathematics. Even in pure mathematics, where often quite subtle and sophisticated arguments are required for the solution of problems, computers have become an invaluable, almost indispensible tool.

In this paper I will describe in more detail some aspects of the impact of computing on research in pure mathematics, and in particular on the use of specialist software to solve mathematical problems.

I will briefly discuss computer-based proofs with reference to two famous examples: the 4-colour theorem, and the non-existence of a projective plane of order 10, and will also mention a few of the major developments within mathematics that have resulted from the influence of computing. Finally I will outline some of the ways in which I have used computer software in my own research, with the aim of illustrating the potential of experimental approaches to questions in pure mathematics.

To begin with, however, it is appropriate to make some general comments. First, it may be said that computers were originally developed to perform calculations which were essentially pure mathematics, and hence it is natural that they continue to be used in this area. On the other hand, their use will always be limited, for by Turing's 1936 answer to Hilbert's 3rd problem, there can be no *universal machine* to decide the truth or falsity of every mathematical statement.

Since the design of computers for cracking secret cyphers in World War II, major areas and directions of pure mathematics have altered considerably. The renaissance of number theory (through cryptography) is a notable example, and others include matrix algebra (resulting from extensive research on the solution of linear and differential equations) and combinatorics. More generally, we have witnessed a gradual *discretization* of pure mathematics, although not necessarily at the expense of continuous mathematics.

Computer-based proofs have become common, if not always popular, and much effort is being poured into the areas of constructive mathematics, algorithms, special-purpose mathematical software, and experimental pure mathematics. Some of these will be dealt with in the following two sections.

2 Computer proofs

In recent years a number of long-standing questions and conjectures in pure mathematics have been settled: the Four Colour Theorem, Mordell's conjecture, the Bieberbach conjecture, and of course Fermat's Last Theorem. Of these perhaps the proof of the Four Colour Theorem has been the most controversial, in that the use of a computer was necessary to complete it. Here are some observations about this and another example of interest:

Example 2.1: The Four Colour Theorem

The Four Colour Theorem (or 4CT for short) states that *only 4 colours are required to colour the regions of any plane map in such a way that every two neighbouring regions have different colours.* This was conjectured by Guthrie in 1852, and had a long history of fallacious "proofs" (and attempted proofs), until it was settled with the help of a computer in 1976.

Appel and Haken's proof [AH] came in two parts: Part I being a classification of *unavoidable configurations,* and Part II verifying the reducibility of each configuration (to show there is no minimal counterexample). Part I involved enumeration by hand of some 1400 cases, while Part II used a computer to verify reducibility in each case.

Ironically Part II caused the most controversy, with many eminent and highly-respected mathematicians raising the possibility of computer errors, yet Part I was much more prone to human error — and some say Part I has never been independently verified!

Nevertheless the 4CT is now believed to be true, and in 1994 a simpler proof was constructed by Robertson, Sanders, Seymour and Thomas [RS], replacing Part I of Appel and Haken's proof by a machine-readable and verifiable list of 633 cases.

Example 2.2: There is no projective plane of order 10

A finite projective plane of order n is an incidence structure made up of $n^2 + n + 1$ points and $n^2 + n + 1$ lines, such that any two points lie together on exactly one line and any two lines intersect in exactly one point. Such a plane is known to exist whenever n is a prime-power, however there is no *known* plane of non prime-power order n.

It was proved by Tarry in 1900 that there is no projective plane of order 6, but it then took until 1989 to show there is no projective plane of order 10. This was achieved by Lam, Thiel and Swiercz [LT], using a computer search for 19-point configurations (corresponding to codewords of length 19 in the associated binary code).

Their search required over 2000 hours of computing time, with the obvious implication of hardware errors. In fact they admit the detection and correction of such errors, but included checks in their programming so that even with one error per 1000 hours, the probability of their proof being incorrect would be at most 1 in $500,000$.

Example 2.1 indicates a change in the interpretation of "proof", where we may accept a result as being *very probably* true. In a similar vein, a general desire to understand how and why some theorems are true — rather than proving by contradiction that they cannot be false — has stimulated the growing field of *constructive mathematics.*

Along with this is a growth industry in automated reasoning (artificial intelligence), but also there has been a fundamental change of emphasis in methodology. Probabilistic and experimental techniques (using random number generation) are now common, and have even appeared in some aspects of pure mathematics.

Also with the advent of computers the need has been recognised for polynomial-time algorithms for solving problems. For a simple instance of this, note that when solving large systems of linear equations, the method of Gaussian elimination is far more efficient than Cramer's rule!

In turn new areas of mathematical research have been spawned, so much so that now one of the burning questions in mathematics concerns the relationship between problems which are polynomial-time solvable (P), and a class of those which are polynomial-time verifiable but not known to be polynomial-time solvable (NP): is $P = NP$?

3 Experimental mathematics & software

It is clear that mathematics has benefitted a great deal from the use and influence of computers. Apart from practical considerations and the wealth of new methods available, there is now a much greater understanding of many avenues of research.

Of course, computers are unlikely to ever match the ingenuity and creativity of the human mind, and quite rightly, "computer proofs" may always be viewed with some skepticism, but that should not detract from their potential to contribute in many significant ways. In particular, there are many situations in which a positive computational approach can yield new results or throw light on old problems.

Computers can be used for simulation (of systems and processes), combinatorial searches, construction and analysis of simple examples, formulation and testing of conjectures, and classification of small cases, for example. In such ways they can often provide answers that can subsequently be checked by hand, or provide a picture that points the way to a theoretical proof, as will be illustrated in the next Section.

This form of experimental approach is becoming more common (and successful) in a large number of areas, especially number theory, discrete algebra, combinatorics, numerical computation, finite geometry, low-dimensional topology, and even statistical mechanics.

Many software packages are available, including special purpose packages Magma (for discrete algebra and number theory), GAP (groups, algorithms, programming), KANT and Pari (number theory), as well as more general purpose mathematical packages such as Maple, Mathematica, and MatLab. Such packages are now widely used in teaching and research, with considerable success, in many parts of the world.

4 Some recent examples & successes

In this section I will describe three examples of ways in which I have used computer methods in my own research, to illustrate some of the potential of the approaches suggested in Section 3.

Example 4.1: hexagon-free subgraphs of hypercubes

For every positive integer n, the hypercube Q_n is an incidence structure generalising the cube to n dimensions. Its vertices are all possible n-tuples of 0's and 1's (of which there are 2^n), and any two such n-tuples are joined by an edge whenever they differ in exactly one co-ordinate.

Some years ago Paul Erdös raised the following question (which is relevant to the study of fault tolerance properties of parallel-processing architectures): Can the edges of the n-cube Q_n always be coloured using t different colours in such a way that there is no hexagon whose edges all have the same colour? By a "hexagon" is meant a circuit of length 6, such as the one with vertices $(0,0,0)$, $(0,0,1)$, $(0,1,1)$, $(0,1,0)$, $(1,1,0)$, $(1,0,0)$, and the question entails finding some t (independent of n) for which a t-colouring exists.

When I first learnt about this question, I experimented with a few possibilities for suitable colourings, with the help of the GAP package in testing them for small values of n. Eventually I stumbled on the following idea:

Consider a typical edge of Q_n, from the vertex $\mathbf{x} = (x_1, \ldots, x_i, \ldots, x_n)$ to the vertex $\mathbf{y} = (x_1, \ldots, \bar{x}_i, \ldots, x_n)$, where $\bar{x}_i = 1 - x_i$. If $\mathbf{x}$ has L 1's to the left of x_i and R 1's to the right of x_i, then let us colour the edge $\mathbf{x} - \mathbf{y}$

$$\begin{cases} \text{blue} & \text{if } L - R \equiv 0 \bmod 3 \\ \text{green} & \text{if } L - R \equiv 1 \bmod 3 \\ \text{red} & \text{if } L - R \equiv 2 \bmod 3 . \end{cases}$$

With this colouring, computation in small cases revealed no monochromatic hexagons, and then it was a relatively simple matter to prove (by hand) that for all n there are no monochromatic quadrangles or hexagons; see [C3].

Example 4.2: highly symmetric networks

A combinatorial graph (or network) Γ is said to be *symmetric* if any two ordered edges are equivalent under some symmetry of Γ, and more generally, *s-arc-transitive* if any two ordered paths of length s are equivalent under some symmetry of Γ. For example, the underlying graph of the 3-dimensional cube is 2-arc-transitive (but not 3-arc-transitive). More highly symmetric examples include the 3-arc-transitive Petersen graph (on 10 vertices) and Tutte's 5-arc-transitive 8-cage (on 30 vertices).

Several years ago Tutte proved that every symmetric finite *cubic* (trivalent) graph is at best 5-arc-transitive. Furthermore, Tutte's analysis shows that the symmetry group of any 5-arc-transitive finite cubic graph has to be a homomorphic image of a particular abstract group G_5, which may be presented in terms of generators and relations as follows:

$$G_5 = \langle\, h, a, p, q, r, s \mid h^3 = a^2 = p^2 = [p, q] = [p, s] = pqrsrs = a^{-1}paq = a^{-1}ras = h^{-1}php = h^{-1}qhr = h^{-1}rhpqr = hshs = 1 \,\rangle.$$

Conversely, any non-degenerate finite image of G_5 is the symmetry group of some 5-arc-transitive cubic graph.

Now computer methods exist for finding small images of finitely-presented groups such as G_5 (through their low index subgroups). Using such methods, Peter Lorimer and I were able to find several interesting examples of symmetric cubic graphs, providing answers to some long-standing questions; see [CL]. Subsequent identification of some of the common features of these examples was the key to the construction of an infinite family of 5-arc-transitive cubic graphs, dispelling any idea that such graphs are rare; see [C1].

Example 4.3: an unexpected isomorphism

Earlier attempts to find and analyse examples of symmetric graphs often involved the imposition of additional assumptions such as the presence of circuits whose vertices are permuted in cycles. In particular, associated with certain 4-arc-transitive graphs containing a circuit of length 12 was the group

$$4^+(a^{12}) = \langle\, h, a, p, q, r \mid h^3 = a^2 = p^2 = [p, q] = pqrqr = a^{-1}pap = a^{-1}qar = h^{-1}phq = h^{-1}qhpq = hrhr = (ha)^{12} = 1 \,\rangle.$$

This group became the subject of attention for some time following several attempts to prove it is infinite.

Again computer methods revealed some aspects of its structure, and in particular I noticed a normal subgroup of index 336 with remarkable properties. Using this subgroup I was able to construct an 8×8 matrix representation of $4^+(a^{12})$, and further computation showed that modulo small primes $p \equiv 2, 3$ and 5, these 8×8 matrices generate a group of order $2p^3(p^3-1)(p+1)$, which happens to be twice the order of the 3×3 matrix group $SL(3, p)$.

In turn this observation led to the following theorem, which can be proved by hand (but which was discovered as a result of computer experimentation):

The group $4^+(a^{12})$ *is isomorphic to* $SL(3, \mathbb{Z}).C_2$, *the group of all* 3×3 *integer matrices of determinant* 1 *extended by its inverse-transpose automorphism.*

For the details, see [C2]. Incidentally, the reason underlying this unexpected isomorphism has been shown by Peter Neumann to have a connection with finite projective planes; but that is another story!

References

[AH] K. Appel and W. Haken, Every planar map is four colorable, Parts I & II, *Illinois J. Math.* **21** (1977), 429–567.

[C1] M. Conder, An infinite family of 5-arc-transitive cubic graphs, *Ars Combinatoria* **25A** (1988), 95–108.

[C2] M. Conder, A surprising isomorphism, *Journal of Algebra* **129** (1990), 494–501.

[C3] M. Conder, Hexagon free subgraphs of hypercubes, *J. Graph Theory* **17** (1993), 477–479.

[CL] M. Conder and P. Lorimer, Automorphism groups of symmetric graphs of valency 3, *J. Combinatorial Theory Ser. B* **47** (1989), 60–72.

[LT] C. Lam, L. Thiel and S. Swiercz, The non-existence of finite projective planes of order 10, *Canadian J. Math.* **41** (1989), 1117–1123.

[RS] N. Robertson, D. Saunders, P. Seymour and R. Thomas, The four-colour theorem, *preprint*.

INFORMATION TECHNOLOGY - THE *VIRTUAL REALITY* OF THE SCHOOL MATHEMATICS CLASSROOM

David C. Johnson
King's College, University of London
Cornwall House Annex, Waterloo Rd.
London SE1 8WA, ENGLAND
e-mail: djohnson@bay.cc.kcl.ac.uk

Abstract

Where are we today? Achievements and research over the past 30 years would suggest significant progress has been made - *virtual reality*. However, what is the reality of school mathematics classrooms in the UK and US? The recently completed ImpacT research in the UK and the US component of the International IEA Computers in Education Study indicate both countries have a long way to go before the potential is realised on any large scale basis. These works also provide some insights in regard to setting realistic medium and long term goals - and providing support for the key participants in this endeavour, teachers and pupils. However, there are no easy or quick solutions; changes in education take time and are highly dependent on people, priorities and resources.

Personal Reflections on the *Virtual Reality*, 1965-1995

I propose to begin this paper by sharing some personal reflections. The themes for these are based on my experiences in computing in school mathematics as a classroom teacher, teacher educator and researcher, 20 years in the US and 18 in the UK. The decade of my initial work was the 1960's, and this was the period of the first small scale, but relatively intensive use of computers in school mathematics classrooms. I note here that changes in mathematics education have a tradition of being influenced by culture and context as well as the available technology - one needs only to look at early National Council of Teachers of Mathematics (NCTM) Yearbooks for documentation of this (see, for example, the Third Yearbook, 1928, Selected Topics in the Teaching of Mathematics, with chapters on 'functional thinking, dynamic symmetry, introductory calculus as a high school subject, etc). However, the *virtual reality* of the changes is often well ahead of the real world of classrooms. Changes on any large scale take time.

In the narrative which follows I will also attempt to give some feel for the pace of research and development on/in the role and use of computers in school mathematics. I will highlight some selected events and publications and apologise if I have omitted an event you may feel was even more important than those used here. I can only say my choices are intended to be illustrative and not definitive.

The 60's and 'early' 70's

It was in 1962 that I took my first computer science courses - these were primarily numerical analysis and the focus was on programming in machine code. What was this experience like (see Figure 1)?

Figure 1. 1962 - Computer Science - Two terms

Context: Octal coding, punch tape, UNIVAC 1103 (valves or vacuum tubes)
Most significant task: Write an octal program to produce decimal output 1-20
Focus: Overcoming constraints in instructing the machine to complete a task

For me this was an interesting experience and my examination marks were high - however never once was I able to get a program to run successfully (fortunately I had a partner who loved to 'live in the computer room' and debug a program or wait while some machine problem was sorted out). If anyone had asked me at this time about the possibility of using the computer to study school mathematics I'm afraid I would have thought the person crazy. On the other hand, it was only a year later, 1963/64, that I was using computer programming with pupils aged 12-18 as part of the mathematics curriculum in the University of Minnesota High School (UMHS), the University laboratory school.

We had 'discovered' FORTRAN and were transporting programs for card punching and running on the University computer (now up-graded to a more reliable and up-to-date IBM). This was a bit of a nightmare however as the error rate was high, between 60-70% - not the least due to inaccuracies in the preparation of format statements. As an aside here, let me note that another teacher and I elected to help the pupils with these, and what happened? Yes, the error rate went to 80-90%. Ah well, we soon had help from a new development - a set of materials developed by Robert Smith at CDC, 'Card FORTRAN without Keypunch'. It was also the case that we also now had some school machines, the Bitran 6 (a six binary digit machine, with one for sign, what kind of numbers could we work with?), and the CDC Bendix G-15 (one of the last of the valve machines developed, designed to sell for $60K, taken over and sold to schools at $15K, with annual maintenance at about the same amount). But the Texas Instruments developments with the transistor meant that changes in hardware were becoming dramatic and fast - at least they felt fast then, but this was just the 'tip of the iceberg' yet to come in the next 30 years. I could reminisce more on these early experiences, but time and space just do not permit. Suffice it to say, we were ready to welcome the next development as school computing still suffered from a focus on getting the machine to execute a particular task, with a successful debugged program the goal, rather than the use of the program to explore big ideas in school mathematics.

In the Spring of 1964 we were fortunate enough to establish contact and link with Dartmouth University in New Hampshire. We now had access their GE time-share computer with the computer language BASIC. Time-share meant we could access the Dartmouth computer from the classroom or school laboratory in Minnesota. The language BASIC was developed by John Kemeny, mathematician, and Tom Kurtz, computer scientist, for use by students in the University - the goal being easy access and use by those other than computer scientists. While the early versions of the language eventually came in for considerable criticism, for those of us attempting to implement ideas of algorithmics (the design and analysis of algorithms) as a way of describing mathematical ideas this was an almost unbelievable move forward. The focus in school mathematics computing was no longer one of merely programming the machine to complete an often trivial task, but rather now offered a new way to view mathematical ideas as dynamic procedures - e.g., 'primeness' is a procedure for testing whether or not a number is prime, not merely a definition, and a circle as movement - move a little turn a little - not just a 'locus of points'. This is not to say we made substantial progress in re-thinking mathematics as much of the work in the late 60's was still focused on supporting the 'traditional' curriculum.

The years 1964 on through the early 70's were exciting ones for me. I was the director of the Computer Assisted Mathematics Program (CAMP), a R&D project which produced supplementary references for supporting the teaching and learning of secondary school mathematics - five books (Hatfield & Johnson, 1968, Walther & Johnson, 1969, LaFrenz & Johnson, 1969, Kieren & Johnson, 1969, and Katzman & Johnson, 1970) and computing activities for inclusion in a sixth (a full course text, Wisner, 1973). The CAMP project represented the first major attempt to integrate computing into the teaching and learning of a substantial portion of the whole of the secondary school mathematics curriculum. It was also during this time that I had the opportunity to work with colleagues on a major NCTM Committee - the Computer-Oriented Mathematics Committee[1]. This Committee produced three influential publications for the NCTM - Computer Facilities for Mathematics Instruction (1967), Introduction an Algorithmic Language (BASIC) (1968), and a chapter, 'The role of electronic computers and calculators' for the NCTM Yearbook on Instructional Aids in Mathematics (1973).

The CAMP research produced results which along with R&D activities elsewhere in the US and UK demonstrated the exciting potential of computing in school mathematics (and of course, there were the beginnings of exciting work in other subjects, most notably simulations in school science and social studies or history and geography). Of note here was the fact that developments in school mathematics at that time were generally of two types - the machine as a tutor or 'drill master' (e.g., see Suppes, Jerman and Brian, 1968) and pupil's controlling or using the machine through programming. Both areas had strong support, but my personal interest was in the latter and this was also the position espoused by a number of individuals and publications (see Figure 2).

Figure 2. R&D - programming - people and publications - late 60's and 'early' 70's

Richard Andree in Oklahoma, a book (1967) focusing on exploring mathematical/computing ideas through programming
Bill Dorn & Gary Bitter at Denver, Colorado (1970, 1972). research and publications supporting and extending mathematical ideas through programming
Wally Feurzig and Seymour Papert (1969, 1971) Logo programming and 'turtle geometry' with primary and secondary school pupils
School Mathematics Project (SMP) in the UK (1971) - a booklet for secondary school teachers providing examples from mathematics for computer exploration
Larry Hatfield & Tom Kieren (1972) - research results from the CAMP project on computer-assisted problem solving through programming (see also previous CAMP publications)
Thomas Foster (1972) and Ed Andersen (1977) doctoral theses on problem solving and programming in secondary schools (I supervised their work)
Tom Dwyer (1974 and 1975) reports on project SOLO, pupil's developing projects through in-depth investigations and programming
The IFIP 1977 working conference in Varna, Bulgaria (Johnson & Tinsley, 1978), on the implications of computing for secondary school mathematics - contributions from individuals representing 17 different countries.
David Johnson & Robert Harding (1979) - problem solving research results at Cambridge University, high ability students exploring mathematical ideas (based on the developments of the CATAM project, 1976)

There was some conflict in this period as schools needed to take decisions on how best to make use of this limited resource - a common situation was that if a computer was available, this was often only one computer (terminal) in a department. The 'drill and practice' role for raising standards on conventional tests vs. programming for extending the power of pupils for investigating mathematical ideas, much of which was considered new, and by some a 'fad' which would soon 'go away'.

Arguments and debate - computing, curriculum, pedagogy and teachers

There were strong arguments even in the group of mathematicians and mathematics educators supporting the inclusion of computing/programming in the school (mathematics) curriculum. These ranged from

- the role of machine code - yes there were those who felt this should be the first introduction to computing, even with the availability of languages such as BASIC and Logo; through to
- the role of flow charts;
- defining fundamental concepts, in particular should these include those associated with the hardware; and of course
- the programming language itself - FORTRAN and/or ALGOL as these were those used in the real world of business, or BASIC or Logo (and advocates for each of

these debating amongst themselves) or another similar educational derivatives, in terms of the fact that these were developed for use by pupils/students; and finally,
- the inclusion of algorithmics, the design and analysis of algorithms, as a strand in the school mathematics curriculum.

The debate in regard to the last two points was, and is still, fascinating and could easily provide the basis for a full paper. Those interested in the current debate might wish to consider the position espoused in a more recent paper (Johnson, 1992).

Other concerns relative to curriculum, pedagogy and teachers included

- the value of programming for its own sake as distinct from the role in exploration and problem-solving, the latter difficult to support with limited resources;
- the allocation of the resource to skill learning through D&P in a time of 'back to basics';
- the role of computer literacy and awareness courses - to enable the school to provide all pupils with some, often minimal, exposure to the power and potential of the technology in a period of limited resources; and
- the demands on teachers to take on board both the new technology and approaches and the need for time to reflect on these developments.

The last of these was a particularly serious problem then, the reality of school mathematics classrooms in the 60's and 70's, and, as will be shown later in this paper, remains a major concern even now, 25 years later.

I might digress here and also add two quotations from this period:

> "...in the lifetime of today's student the use of computers will become as much a part of everyday life as the telephone or automobile."
>
> "In five years time student access to computers in school will be as common and natural as access to, and use of, pencil and paper."

The first of these was made by John Kemeny in the early 60's, and he was clearly correct in his view of the world of business, industry and leisure. I made the second in an interview for a magazine article published in 1968 and hasten to note that this was also the view of many of my colleagues. Of course we all know this is still far from the truth, even today after the accelerated time-line for developments which were to come in the 80's and early 90's.

The 80's and early 90's

A period of exponential growth in developments in hardware, software and exciting ideas. We see the availability of 'user friendly' hardware and software. The advent of the micro in the 70's has progressed to include enhanced capabilities - memory, graphics, colour, speed, etc. - which enabled the development of powerful utilities for school mathematics, both generic and domain specific tools, for example:

- spreadsheets, including graphics and other representations
- graphics packages
- data base packages, including data analysis facilities
- modelling utilities
- symbolic manipulators
- geometric supposers

Logo became available on a micro - so the 80's became a *decade of movement* - but how far have we moved with Logo? And we now have graphics calculators, palmtop (wallet size) and laptop computers which provide pupils with power, portability and **personal technology**. And these have been expanded to include peripheral devices, an exciting 'add on' to the graphics calculator is that of 'data loggers' and the availability of a computer interface for the calculator. Collecting and analysing real data becomes a (*virtual?*) reality.

The late 80's and early 90's also brought CD ROM, 'telematics' (including, for example, Internet and WorldWide Web), and 'hypermedia', as demonstrated with the fascinating work of APCOT (Apple Classrooms of Tomorrow) which provides an illustration of what might be achieved in an Information Technology (IT) rich environment. We certainly do have some indications of the *virtual reality* and I hasten to add, this is reality in some classrooms, but in how many, or how common is this phenomena?

Before leaving this period, let me share a few personal experiences - see Figure 3. This is a very limited and selected listing of some chronological events I found to be helpful in formulating ideas, e.g., people and documents which confirmed and/or aided in extending my own thinking; a number of which I also had a participating role (these are denoted with an *).

Figure 3. Some selected chronological events in the 80's and 90's

Papert's Mindstorms (1980) - a look at what might be!

NCTM report on the Maryland Computing and Mathematics conference in 1982* (Fey, 1982) and the 1984 Yearbook (Hansen & Zweng) - some wide ranging discussions on the future potential.

The Pendley Manor Conference in the UK* (1983, journal publication, DES/MEP, 1985) - a forward looking position paper, particularly the discussion of algorithmics (of course this reflects my bias and I was a participant).

What is called the 'Fletcher brown booklet' (1983) - carrying on from Pendley Manor.

MicroMath published by the Association of Teachers of Mathematics (ATM) - the first issue in 1985 (more on this later).

The development of mathematics teacher in-service packs by the UK Micro-electronics in Education project (MEP) - the Secondary In-service Pack (Waddingham & Wigley, 1985)* and the Logo Primary Pack (Straker, 1985), both widely distributed and guidance provided which covered a range of potential activities.

While I have had many doctoral students conduct research in the area of school mathematics and computing/programming - one of particular note here is the work with Logo conducted by Richard Noss (1985)*, which provided further confirmation of the potential.

The group involved in the Pendley Manor conference and the MEP packs carried on to produce the booklet Will Mathematics Count? (CET, 1987)* a look at where we were and what needed to be done.

Another fascinating Logo book by Sylvia Weir, Cultivating Minds (1987).

An exciting IFIP conference and publication, Informatics and the teaching of mathematics (Johnson & Lovis, 1987)* - a look at computing and school mathematics ten years on (building on the 1977 conference - publication in 1978, see Figure 2), with consideration give to a range of computer uses.

The NCTM 1991 Yearbook, Discrete mathematics across the curriculum K-12 (Kenney & Hirsch, 1991) - note the particularly important (in my opinion) discussion of the role of algorithmics.

UNESCO report The influence of computers and informatics on mathematics and its teaching (Cornu & Ralston, 1992)* - an update on the conference and conference report with the same name held in Strasbourg in 1985.

The publication of Learning Mathematics and Logo (Hoyles & Noss, 1992).

A chapter, 'Technology and mathematics education' (Kaput, 1992) in the major research reference edited by Douglas Grouws, Handbook of Research on the Teaching and Learning of Mathematics.

The book by Schwartz, Yerushalmy & Wilson, The Geometric Supposer: What is it a case of? (1993).

The new Nuffield Advanced Mathematics*, Hugh Neill, Director and General Editor - I was the project Grant-holder and Faculty associate - the books were published in 1994 and 1995. Possibly the first full two-year A-level course which fully integrated the graphics programmable calculator and other computer software utilities into the teaching and learning activities, i.e., one must use the technology to do the course, and in which algorithmics and modelling also each plays a central role.

A return to MicroMath - one feature of the journal has been that of special issues in which a substantial portion of the publication has been devoted to a particular theme or potential contribution of a particular type of software, e.g.,

Data bases/Handling data (1989, 1991, 1994)
Numerator - and modelling (1990)
Spreadsheets (1990)
Graphics calculator (1991, 1995)
Cabri Geometre (1992)
Teaching Algebra, with all the tools (1993)
Teaching and Learning Geometry - Cabri, Geometry Inventor, Geometer's Sketchpad (1995)

Government data collection, reported in the press, indicates that pupil-computer ratios for primary and secondary schools are 18:1 and 10:1 in 1995, having changed from 107:1 and 60:1 a decade ago (1985).

STOP! WHAT DOES THIS ALL MEAN?
WHAT IS THE REALITY OF THE CLASSROOM?

Classroom Reality - Findings from Recent Research

Background

The developments indicated earlier are expanded on in the popular press - claims are made that "notebook computers are producing positive results", "(a technology conference has) clearly demonstrated how quickly IT has become part of the fabric of most schools", and that there is a "dramatic growth of CD ROM in education". Schools in the UK and US have had substantial inputs from governmental sources to facilitate the acquisition of hardware and software and the provision of in-service support. Concern has therefore been expressed that it would seem important to confirm and assess the situation in schools.

One approach to such an assessment is to conduct large scale surveys of availability of resources, e.g., numbers of machines and pupil/machine ratios, and aspects of access and use in school subjects, often presented as percentages of total computer use distributed across, for example, maths, science, native language, etc.. In the UK such work has been carried out on a regular basis by the Department of Education, DES, now the Department for Education, DfE. At the international level we have the multi-national IEA Computers in Education Study (Pelgrum & Plomp, 1991), and, of particular relevance to this paper, Anderson's (1993) report on the US component of this work. Anderson reports on a large scale survey (over 69,000 pupils, aged 10, 13, and 16, in 2,500 schools) conducted in 1992 (as a follow-up to a similar study conducted in 1989). The report provides information on, for example, the numbers (percentages) of teachers in 'computer-using schools' who use computers in class 'on at least several occasions during the year'; percentages reported for these schools were typically around 50%. The schools also had a pupil/computer ratio of about 7:1, and note is made of the fact that since 1989, U.S. schools have increased their inventory of computer units by as much as 50%. In making international comparisons, Anderson states that "in ... peripherals and networks, the US has at the very least comparable and often larger inventories than the other countries (in the IEA survey)" (p. 27)

The US study also placed emphasis in the research upon using computers in learning different subjects such as science, geography, languages, etc.. "In the U.S., about half of school computer time is spent on learning about computers and the other half on learning other subjects with the help of computers" (p. xvii). But what does this 'learning other subjects' really mean? Some aspects of this question are discussed by Lundmark in the chapter 'pupils 'opportunity to learn with computers', but the evidence reported is primarily that of 'counts' of responses to items in the survey questionnaires (as opposed to, say, actual observations of classes and pupil assessments). An important consideration in the discussion is the fact that pupils were counted as a 'substantial user' of IT, if they used software six or more times in the school year, i.e. six or more times in a period of 150 school days. In her conclusions, Lundmark notes that

> "The question of how far computer instruction has been integrated across the school subjects curriculum is noteworthy. A little more than one-fourth of all US students say they used computers in none of the traditional subjects (included in the survey) and a little more than one-fourth say they used computers in only one of those ... subjects during the year. ... it seems that the use of computers as instructional tools has not advanced very far across the curriculum." (p. 70)

A further source of evidence for discussing learners and IT in school subjects, are the numerous studies conducted in different educational contexts, for example, multimedia learning environments or aspects of group work in classrooms, or studies which focus on the effects of IT on pupils' learning within a specific conceptual domain in a curriculum area. Niemiec and Walburg (1992) report on results from over 250 studies on the effects of IT on pupils' learning of particular skills and concepts, with the overall results generally supportive of the experience.

Evidence can also be found in books, monographs and conference proceedings, which summarise and synthesise the research and theoretical perspective supporting classroom activity linked to a particular type of use (e.g., simulations) or the role of powerful generic software tools.

While such reports add important information to our understanding of the potential, what can we say about the impact of IT on teachers and pupils activities and achievements across subjects and over time in the world of 'typical' schools? Has the acquisition of resources been 'cost effective' in terms of pupils achievements, or are the results found in the literature only possible in restricted and highly supported contexts? The UK Department for Education (DfE) attempted to address this question through a large scale longitudinal investigation - the ImpacT study.

The ImpacT research (England and Wales)

The ImpacT study, "An evaluation of the impact of Information Technology on children's achievements in primary and secondary schools" (Watson, 1993), was carried out by a team of researchers in the Centre for Educational Studies, King's College London. The focus of the work was on pupils' learning and classroom activity involving IT in four school subject areas - mathematics, science, geography, and English - at three age levels - 8-10, 12-14, and 14-16, designated here in terms of the initial year, i.e., year 4 (Y4), Year 8 (Y8) and year 10 (Y10). Each age cohort, with some exceptions, was followed for two years.

The work was designed to extend our understanding from earlier research to include longitudinal effects within school subjects, cross-subject considerations of general aspects of classroom use of IT, and the provision and use of hardware and software resources. These were integrated to address a range of issues encompassing learning, pedagogy, and school organisation. The ImpacT results from the main component parts were linked to enable the research team to address three main questions:

- Pupils' Learning: did IT make a contribution?
- Pedagogy and Practice: what can we say about the planning and practice of teaching to incorporate IT?
- Schools' Organisation: what were the demands of IT on the schools?

My intent here is to focus on an aspect of the first question: the impact of IT on mathematics learning, although findings from the second question will be used in the discussion relative to how we might move forward. Details on the research design and methodology are given in

Johnson, Cox, and Watson (1994) and some selected outcomes are reported in Johnson (1995a). Full details on all aspects of the work are given in the main report (Watson, 1993).

The study included data collected from over 2300 pupils from 87 classes in 19 LEAs, distributed throughout England and Wales. The classes were chosen as matched pairs, all classes being nominated for their good teaching and curriculum delivery, but one of each pair made regular use of IT (HiIT), while the other (LoIT) did not. Three kinds of data were collected:

- An assessment of pupils' achievement of specific learning tasks and skills, through two administrations of specially designed **subject-focused assessments** to the matched pairs of classes. The assessments were administered as posttests twice during the study, once during the first year and again near the end of the second year (the mathematics tests are designated as **MR1** and **MR2** in the discussion which follows). Additionally, some of the pairs, and some of the HiIT classes were the focus of eight topic-specific mini-studies, and all pupils took a final test for IT concepts and skills. Soon after the classes had been chosen, all pupils also took a general reasoning test; to be used as a pretest in the research and to provide a check on how well matched the two classes in each pair were (this is designated as **AH1**, primary, and **AH2**, secondary).

- Five in-depth longitudinal **case studies** in HiIT classes focused on classroom processes and pupil interactions. Classrooms were observed, pupils and teachers were interviewed and documentary evidence was gathered to illuminate classroom realities. Qualitative analysis was based on those themes and issues that emerged from the data *across* the five studies.

- IT resourcing and use was monitored throughout by the regular returns of **questionnaires and data sheets** from the teachers and pupils in each class. Hardware and software provision, pupils' IT use in ImpacT subjects and across all subjects, and pupils' extra-mural use of IT were analysed descriptively by classes, age-cohorts and subjects.

In the case of mathematics the 'assessment of achievement' and 'IT resourcing and use' data were from matched pairs of classes in each of the three age groups – six, eight, and eight, representing three or four matched pairs in each group, Y4, Y8 and Y10 respectively. In addition a comparative HiIT-LoIT mini-study was conducted – this involved the role of Logo programming in the study of angles (see Johnson, 1995a, for a report on this study). As indicated above, the case studies were treated as a whole with the focus on general themes across the five studies (one of which was a secondary school mathematics classroom).

Selected results - pupils' learning. Within the limitations of the study, the answer to the first research question was yes, **IT did make a contribution to the learning of mathematics, but the contribution was not consistent across age bands or classes within an age band.** Group data for each age cohort are given in Table 1. When the results were adjusted for the differences in ability of the classes the Y4 and Y10 HiIT classes scored significantly better ($p<0.05$) on both administrations of the mathematics assessments.

Table 1 - Lo- and HiIT Group Performance: **AH1, MR1** and **MR2**,

Y4		**AH1**		**MR1**	
Group	n[a]	mean	sd	mean	sd
HiIT	70	24.41	8.21	8.20	3.29
LoIT	93	26.33	7.69	7.51	3.45
				MR2	
HiIT	59	24.61	8.57	10.39	4.00
LoIT	69	26.35	7.85	9.26	4.00
Y8		**AH2**		**MR1**	
Group	n	mean	sd	mean	sd
HiIT	90	63.18	13.40	18.46	6.20
LoIT	95	58.85	12.47	16.87	5.20
				MR2	
HiIT	62	64.10	12.89	23.10	6.09
LoIT	60	57.98	13.14	21.50	5.98
Y10		**AH2**		**MR1**	
Group	n	mean	sd	mean	sd
HiIT	86	62.12	11.48	19.33	4.59
LoIT	69	58.39	9.40	15.45	3.52
				MR2	
HiIT	68	63.28	11.63	21.75	4.56
LoIT	42	58.79	8.91	19.55	4.11

a Pupils with both **AH** pretest and a corresponding **MR1** or **MR2** posttest score

While the results on the mathematics reasoning assessment for the Y8 cohort indicated no significant differences, the 'Angles' mini-study conducted with two classes in this age-group did yield highly significant differences on both an immediate 'angles' posttest and a retention test given 8 months later (Johnson, 1995b). The results provide strong support for the inclusion of Logo turtle geometry programming, both during the study of angles and angle relationships and as an on-going activity in school mathematics at this level.

However the generally better performance of the HiIT groups mask considerable variation between the groups, and the main proportion of the increase in scores came from a small number of the HiIT classes. **Access and use in these classes suggested there may well be some minimum threshold of IT use for any detectable impact on mathematics attainment.** This was particularly noticeable in the Y8 Hi- and LoIT matched pair involved in the angles mini-study (this HiIT class also out-performed *all* other classes in the cohort on **MR2**, but this performance was offset by the lower performance of the other HiIT classes in the cohort).

Table 2 below provides an indication of frequency of access of pupils in each of the ImpacT mathematics classes. For the ImpacT sample *across all four subjects and age groups*, very few class median scores for IT use in subject *for any one term*, on a scale of 0-4, were 3 or higher, where 3 represented pupil use of the computer in the main subject at a frequency on average of once each week, and 2, for example, represented use between three and five times per term. Returns from 70 classes were included in the analysis, and of these, nine attained a median score of 3 or above for *at least one term* - these comprised seven from mathematics and two from English. From these data **it would appear that the notion of frequent use of IT in ImpacT classes throughout the country during the period of the research was still far from being achieved.** [And the issue here is are we really coming close to what might be required?]

Table 2 - Median Scores[a] for Class IT Use in Mathematics; Y4, Y8 and Y10

		LoIT			HiIT			
Class Number		2	4	6	1	3[b]	5	7
Y4	Max.	3	0	0	3		2	2
	Min.	0	0	0	0		1	0
Y8	Max.	0	0	0	3	3	3	1
	Min.	0	0	0	0	1	3	1
Y10	Max.	0	0	0	3	2	3	1
	Min.	0	0	0	0	0	1	1

a The IT use entry is on a scale of 0-4 with 3 being once per week on average. Maximum and minimum are for school terms with the highest and lowest median values.

b There were six classes included in the analyses for Y4.

Information was also collected on the software used in subject. Further information on the mathematics cohort can be found in Cox (1993). A main observation here is that in general the software used was of an 'investigatory' nature, including programming in Logo. One conclusion from these data, when taken with the other information, was that the IT activity in certain classes made a significant contribution to the pupils' achievements. Further, as indicated previously, the results suggested that there may well be some minimal threshold of access, both frequency and over time and type of classroom activity for such a contribution to become apparent within the assessment schemes now utilised in classrooms, or even nationally.

The main focus of the case study research was on classroom processes. The case study aspect of the research must be qualified in terms of the approach and method in that while the data collection was rigorous and detailed, the analyses were designed to provide exemplification rather than generalisations. Selected observations from the case study classes suggested some important considerations for pupil use of IT which tended to support and extend results from the other assessments, for example:

- computers were found to be good motivators which heightened pupils' interest and enjoyment and were also seen to have a positive effect upon the status of the subject;
- computers aided concentration by focusing pupils' attention on the work in hand and as a result some pupils and teachers believed that the standard of work produced was of a higher quality than it would have been otherwise;
- opportunities to work in an open-ended way enabled pupils to become involved in more complex and challenging learning situations beyond that typically experienced.

Further, some of the failures to detect any contributions through the use of IT may be attributed to some of the problems encountered by pupils in the case study classes:

- difficulties in using a particular software package;
- inability to work effectively in a collaborative environment.

Selected results - pedagogy and practice. The answer to the question regarding the planning and practice of teachers mainly involved a consideration of classroom management and organisation and teaching styles along with aspects of hardware and software availability and use. **The results from the case studies and other aspects of the field study indicated quite clearly that any contribution was dependent upon a range of factors, the most important being that of the role of the teacher.**

Teachers' responsibilities were found to demand careful attention to organisation and management, in particular the effective use of collaborative or group work. Further, effective use of IT represented substantial demands in terms of knowledge and understanding of, and familiarisation with, a variety

of software in order to integrate the activity, in philosophical and pedagogical terms, with a larger scheme of work.

The use of more general purpose software, e.g. spreadsheets, databases, and programming, placed additional demands on the teacher, beyond that of becoming familiar with the use of the more complex software, to include more reflection on the nature of the subject and the potential role of such software in enhancing processes and understanding.

Discussion

The ImpacT research has provided evidence of a significant contributions of IT to pupils' learning of mathematics. However, the issues and problems regarding changes in the educational system are complex and multi-faceted. Enhancing educational opportunity through IT is no different, in that any strategy must take into account a range of needs and issues and the fact that achieving the goals will take time. The research showed that in spite of a number of commendable efforts and a sustained national strategy to support the implementation of IT in education, people at all levels still need more help in formulating clear policies and strategies; this should go beyond focusing on particular aspects of issues and problems and provide a comprehensive and long term view to take full advantage of the potential impact of IT on pupils learning. The reality is still a long way behind the *virtual reality.*

footnote

1. The membership of the NCTM Computer-oriented Mathematics Committee for the 1967 and 1968 publications included the following people: Robert L. Albrecht, William F. Atchison, Sylvia Charp David C. Johnson, Bruce E. Meserve, John O. Parker, Dina Gladys S. Thomas (for the 1967 and 1968 publications). The membership for the Yearbook chapter was changed with Walter Koetke replacing Sylvia Charp.

References

Andersen, E.D. (1977) An evaluation of Teacher In-service Program on Computer-extended Mathematics Grades 7-12. Unpublished Doctoral Dissertation, University of Minnesota.

Anderson, R.E. (Ed) (1993) *Computers in American Schools 1992: An Overview.* A National Report from the International IEA Computers in Education Study. Minneapolis (Minnesota): University of Minnesota (Social Sciences).

Andree, R.V. (1967) *Computer Programming and Related Mathematics.* New York: John Wiley & Sons.

Bitter, G.G. (1970) Effect of computer applications on achievement in a college introductory calculus course. Unpublished doctoral dissertation, University of Denver.

CET (Council for Educational Technology) (1987) *Will Mathematics Count? - Computers in Mathematics Education.* London: CET.

Cornu, B., & Ralston, A. (Eds.) (1992) *The influence of computers and informatics on mathematics and its teaching.* Science and Technology Education Document Series #44, Paris: UNESCO. [An update of the earlier ICMI Study Series book by the same name, published by Cambridge University Press, 1986.]

Cox, M.J. (1993) Information Technology resourcing and use. In D.M. Watson (Ed.) *The ImpacT Report: An evaluation of the impact of information technology on children's achievements in primary and secondary schools.* King's College London (for the Department for Education), 27-60.

DES/MEP (1985) Mathematics and microcomputers: A Pendley Manor Report. *Mathematics in School,* **14**, 2, 26-28.

Dorn, W.S. (1970) Computer-extended instruction: an example. *The Mathematics Teacher,* **63**, 147-158.

Dorn, W.S., Bitter, G.G. & Hector , D.L. (1972) *Computer Applications for Calculus.* Boston: Prindle, Weber & Schmidt.

Dwyer, T.A. (1974) Heuristic strategies for using computers to enrich education. *International Journal for Man-machine Studies,* **6**, 137-154.

Dwyer, T.A. (1975) Soloworks: Computer based laboratories for high school mathematics. *School Science and Mathematics*, **3**, 1, 93-99.

Feurzeig, W. et al (1969) Programming languages as a conceptual framework for teaching mathematics. Report No. 1889. Final report of the first 15 months of the Logo project. Cambridge (Massachusetts): Bolt, Beranek and Newman.

Feurzeig, W. et al (1971) Programming languages as a conceptual framework for teaching mathematics. Report No. C615. Cambridge (Massachusetts): Bolt, Beranek and Newman.

Fey, J.T. (Ed.) (1984) *Computing & Mathematics: the impact on secondary school curricula.* Reston (Virginia): NCTM.

Fletcher, T.J. (1983) *Microcomputers and Mathematics in Schools: A discussion paper.* London: Department of Education and Science.

Foster, T.E. (1972) The Effect of Computer Programming on Student Problem-Solving Behaviours in 8th Grade Mathematics. Unpublished Doctoral Dissertation, University of Wisconsin.

Hansen, V.P., & Zweng, M.J. (Eds.) (1984) *Computers in Mathematics Education.* Reston, Virginia: National Council of Teachers of Mathematics.

Hatfield, L.L., & Johnson, D.C. (1968) *CAMP first course.* Glenview (Illinois): Scott, Foresman & Co.

Hatfield, L.L., & Kieren, T.E. (1972) Computer-assisted problem solving in school mathematics. *Journal for Research in Mathematics Education*, **3**, 2, 99-112.

Hoyles, C., & Noss, R. (Eds.) (1992) *Learning Mathematics and Logo.* Cambridge (Massachusetts): The MIT Press.

Johnson, D.C. (1992) Algorithmics in school mathematics: Why, what and how?. In I. Wirszup & R. Streit (Eds.) *Developments in School Mathematics Education Around the World - Volume Three.* Reston (Virginia): NCTM, 330-345.

Johnson, D.C. (1995a) The reality of learners' achievements with IT in the classroom. In D. Watson & D. Tinsley (Eds.) *Integrating Information Technology into Education*, London: Chapman and Hall, 73-83.

Johnson, D.C. (1995b) Logo and angles - a focused investigation with pupils aged 12-13. In Y.J. Katz (ed.) *Computers in Education: Pedagogical and Psychological Implications.* IFIP Transactions. Sophia: Bulgarian Academy of Sciences Central Library, 178-187.

Johnson, D.C., Cox, M.J., & Watson, D.M. (1994) Evaluating the impact of IT on pupils' achievements. *Journal of Computer Assisted Learning*, **10**, 3, 138-156.

Johnson, D.C. & Harding, R.D (1979) University level computing and mathematical problem-solving ability. *Journal for Research in Mathematics Education*, **10**, 1, 37-55.

Johnson, D.C., & Lovis, F. (Eds.) (1987) *Informatics and the teaching of mathematics.* Amsterdam: Elsevier Science Publishers B.V.

Johnson, D.C., & Tinsley, J.D. (1978) *Informatics and mathematics in secondary schools: impacts and relationships.* Amsterdam: North-Holland.

Kaput, J.J. (1992) Technology and mathematics education. In D.A. Grouws (Ed.) *Handbook of Research on Mathematics Teaching and Learning.* New York: Macmillan (for NCTM), 515-556.

Katzman, P.W., & Johnson, D.C. (1970) *CAMP geometry.* Glenview (Illinois): Scott, Foresman & Co.

Kenney, M. & Hirsch, C. (Eds.) (1991) *Discrete mathematics across the curriculum, K-12.* Reston, Virginia: National Council of Teachers of Mathematics.

Kieren, T.E., & Johnson, D.C. (1969) *CAMP intermediate mathematics.* Glenview (Illinois): Scott, Foresman & Co.

LaFrenz, D.E., & Johnson, D.C. (1969) *CAMP algebra.* Glenview (Illinois): Scott, Foresman & Co.

MicroMath - A journal of the Association of Teachers of Mathematics (ATM) - address MicroMath, ATM, 7 Shaftesbury Street, Derby DE23 8YB, UK.

NCTM (1928) *Selected Topics in the Teaching of Mathematics - Third Yearbook.* New York: Teachers College, Columbia University.

NCTM Computer -oriented Mathematics Committee (1967) *Computer Facilities for Mathematics Instruction.* Reston (Virginia): NCTM.

NCTM Computer-oriented Mathematics Committee (1968) *Introduction to an Algorithmic Language (BASIC*. Reston (Virginia): NCTM.
NCTM Computer-oriented Mathematics Committee (1973) The role of electronic computers and calculators. In E.J. Berger (Ed.) *Instructional Aids in Mathematics - Thirty-fourth Yearbook.* Reston (Virginia): NCTM.
Niemiec, R.P., & Walburg, H.J. (1992) The effects of computers on learning. *International Journal of Educational Research,* **17**, 1.
Noss, R. (1985) *Creating a Mathematical Environment through Programming: A Study of Young Pupils Learning Logo.* Doctoral Dissertation. Chelsea College, University of London. London: Institute of Education, University of London.
Nuffield Foundation (Neill, H. Gen. Ed.) (1994/5) *Nuffield Advanced Mathematics.* [Five books with eight additional options, including for example, *Discrete mathematics, The history of mathematics*, and *Mathematics, music and art.*] Harlow, Essex (UK): Longman.
Papert, S. (1980) *Mindstorms: children, computers, and powerful ideas.* New York: Basic Books.
Pelgrum, W.J., & Plomp, T. (1991) *The Use of Computers in Education Worldwide.* Results from the IEA Computers in Education Survey in 19 Education Systems. Osford (UK): Pergamon Press.
SMP (School Mathematics Project) (1971) *Computing in mathematics: Some experimental ideas for teachers.* Cambridge (UK): Cambridge University Press.
Schwartz, J.L., Yerushalmy, M., & Wilson, B. (Eds.) (1993) *The Geometric Supposer: What is it a case of?* Hillsdale (New Jersey): Lawrence Erlbaum Associates.
Straker, A. (Ed.) (1985) *Logo In-service pack* (a set of booklets prepared by the MEP Primary Project team). Winchester (UK): MEP National Primary Project (King Alfred's College).
Suppes, P., Jerman, M. & Brian, D. (1968) *Computer-Assisted Instruction: Stanford's 1965-66 Arithmetic Program.* New York: Academic Press.
Waddingham, J. & Wigley, A. (1985) *Secondary Mathematics with Micros In-Service Pack.* London: MEP/CET.
Walther, J.W. & Johnson, D.C. (1969) *CAMP second course.* Glenview (Illinois): Scott, Foresman & Co.
Watson, D.M. (ed.) (1993) *The ImpacT Report: An evaluation of the impact of information technology on children's achievements in primary and secondary schools.* King's College London (for the Department for Education).
Weir, S. (1987) *Cultivating Minds.* New York: Harper and Row.
Wisner, R.J. (1973) *Elements of probability.* Glenview (Illinois): Scott, Foresman & Co.

TEACHING MATHEMATICS THROUGH SYMBOLIC COMPUTATION [1]

Zaven A. Karian
Department of Mathematics and Computer Science
Denison University
Granville, Ohio 43023
Karian@Denison.edu

Symbolic computing systems or computer algebra systems are sophisticated mathematical software packages that combine numeric, graphic and symbolic computation in a unified working environment. Such systems have been available for almost 30 years; however, their use in mathematics instruction has been a relatively recent phenomenon. By today's standards, early versions of symbolic computing systems were quite expensive and rather awkward to use. As recently as a decade ago, a dedicated single-user symbolic computing system (hardware and software) cost about $US40,000; today far more sophisticated versions can be obtained for less than $US2,000. The general availability of this type of software has begun to revolutionize both the teaching as well as the the practice of mathematics. Witness the rich diversity of publications related to symbolic computation by professional organizations (see [2], [3], and [5] for relevant publications from the Mathematical Association of America). Commercial publishers have also become active by releasing general and specialized texts (e.g., [1] and [4]) on symbolic computation.

The purpose of this presentation is to introduce you to the symbolic computing system *Maple* and to show, through examples taken mostly from introductory undergraduate mathematics, the variety of pedagogic uses that can be made of *Maple*. Of course, the use of symbolic computation is not limited to introductory courses. At our institution we use *Maple* as we teach calculus, linear algebra, differential equations, numerical analysis, and probability and statistics.

1. Basic Features of Symbolic Computing Systems

I was advised by the organizers of the conference not to assume that the participants would be familiar with *Maple* or with any symbolic computing system. Accordingly, I will start by describing the most elementary features of *Maple*. To begin with, and to no one's surprise, *Maple* has basic arithmetic capabilities. To understand the *Maple* interaction shown in Figure 1, you need to know that

- the ">" symbol indicates input to *Maple* (i.e., what the user enters) and portions not marked with ">" are responses to the preceding input.

[1]Supported by grants from the Fund for the Improvement of PostSecondary Education (FIPSE grant P116B30079) and the W. M. Keck Foundation.

- `ifactor` stands for integer factorization; `isprime` inquires whether the given argument is prime; consistent with standard mathematical notation, `3!!!` is the factorial of the factorial of 3 factorial; `nextprime` gives the first prime following the argument; `Pi` represents π; and `evalf` converts an expression to floating-point form—in particular, `evalf(Pi,500)` gives the first 500 digits of π.

Figure 2 shows how algebraic expressions may be manipulated *symbolically*. In this interaction ":=" is used to attach names to expressions. Thus, on the first line the expression $1 + 1/n + 1/m$ is given the name `a` and henceforth, `a` will represent this expression.

```
> 3*5+12;
                    27
> ifactor(12345678900987654321);
      (3)^2(11)(57920960187043)(2153)
> isprime(12345678900987654321);
                   false
> isprime(57920960187043);
                   true
> 3!!!;
260121894356579510020490322708104361
                     ⋮
000000000000000000000000000000000000
> nextprime(10**10);
              10000000019
> nextprime 20!;
          2432902008176640029
> Pi;
                    π
> evalf(Pi);
              3.141592654
> evalf(Pi,500);
3.14159265358979323846264338327950288
                     ⋮
4956735188575272489122793818301194912
```

Figure 1. Some Numeric Features

```
> a := 1 + 1/n + 1/m;
```

$$a := 1 + 1/n + 1/m$$

```
> b := 1 + 1/n - 1/m;
```

$$b := 1 + 1/n - 1/m$$

```
> a/b;
```

$$(1 + n^{-1} + m^{-1})\,(1 + n^{-1} - m^{-1})^{-1}$$

```
> simplify(a/b);
```

$$\frac{nm + m + n}{nm + m - n}$$

```
> c:=sin(x)**2+sin(x)**3*cos(x)
   +sin(x)*cos(x)**3-sin(x)*cos(x);
```

$$\sin(x)^2 + \sin(x)^3 \cos(x) + \sin(x)\cos(x)^3 - \sin(x)\cos(x)$$

```
> d := sin(x)*cos(x);
```

$$\sin(x)\cos(x)$$

```
> simplify(c/d);
```

$$\frac{\sin(x)}{\cos(x)}$$

```
> q := x**2 + x + 1;
```

$$x^2 + x + 1$$

```
> solve(q=0, x);
```

$$-\frac{1}{2} + \frac{1}{2}I\sqrt{3},\ -\frac{1}{2} - \frac{1}{2}I\sqrt{3}$$

```
> subs(x=1, q);
```

$$3$$

Figure 2. Some Symbolic Features

Symbolic manipulations, such as differentiation and indefinite integration, are also possible in *Maple*. The first part of the *Maple* interaction given in Figure 3 differentiates a polynomial. This is followed by a set of commands that plot a trigonometric expression and its derivative. The output associated with the `plot` command is shown in Figure 4.

```
> f := (5*x**2 + 2*x - 8)**5;
```

$$(5x^2 + 2x - 8)^5$$

```
> g := diff(f, x);
```

$$g := 5(5x^2 + 2x - 8)^4(10x + 2)$$

```
> a := sin(x)**2 + cos(x);
```

$$\sin(x)^2 + \cos(x)$$

```
> b := diff(a, x);
```

$$b := 2\sin(x)\cos(x) - \sin(x)$$

```
> plot({a, b}, x=-2*Pi..2*Pi);
```

Figure 3. Some Graphics Features

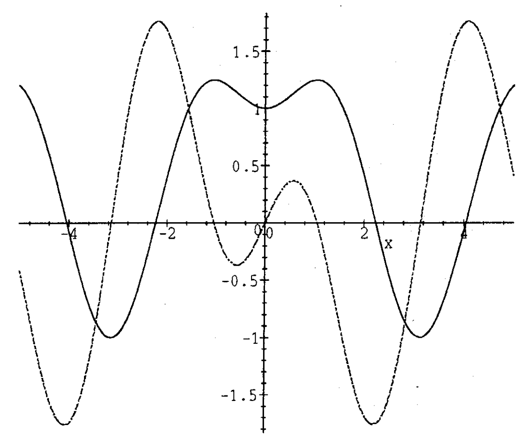


Figure 4. Plot Output

2. Symbolic Computation in Mathematics Instruction

It should be clear, even from the very brief introduction given in the preceding section, that symbolic computing systems put a great deal of power at our disposal. Our task is to use this enormous power to address some of the more challenging problems in mathematics education. Can we help students think about a problem in several ways—analytically, algebraically, geometrically? Can we encourage them to experiment and discover results even if they cannot yet prove or disprove what they may discover? Can we get our students to become less passive and explore ideas on their own?

2.1. Animation

Simple Animations. Animation can be used to understand, among other things, the effect that a parameter has on an equation. One could ask a mathematically unsophisticated student to make some observations regarding the value of the parameter a and the shape of the graph of $f(x) = ax^2$. With *Maple* available to the student, observations of the influence of a on the graph of $f(x)$ becomes a simple matter. `with(plots);`

loads the plotting routines (more on this in Section 3) and `animate(a*x**2, x=-2..2, a=-1..1);` produces animated graphs (moving pictures) of $f(x)$ for values of a varying between -1 and 1. Similar observations can be made about other functions such as $\sin(a+x)$ or $\sin(ax)$. In all cases, after students have observed the impact of the parameter, they should be expected to give justifications for their observations.

Tangent Lines. For a more interesting example, suppose we have a differentiable function $f(x)$ and we consider the manner in which secant lines connecting $(a, f(a))$ and $(x, f(x))$ approach the tangent line to $f(x)$ at $(a, f(a))$ as x approaches a. The user-written procedure `TangLine` (more on user-written procedures in Section 3) can be used to observe the convergence of the secant lines to the tangent line as x approaches a. For $f(x) = 2x^3 - 9x^2 + 12x + 2$ and $a = 3$, for example, the animation would be achieved by first entering $f(x)$ as an expression and then using `TangLine(f, 3, -1..5);`.

Taylor Polynomials. For a final example of the use of animation, we consider the approximation of a function $f(x)$ by its Taylor polynomials. To study the way Taylor polynomials of $f(x)$ approach $f(x)$, a sequence of Taylor polynomials could be animated over the graph of $f(x)$. Suppose $f(x) = \sin(x)$ and we want Taylor polynomials based at $a = \pi/2$. The *Maple* interaction given in Figure 5 will produce the graph of $f(x)$ with an animation of Taylor polynomials of degree $0, 2, 4, \ldots, 22$. The choice of $f(x)$, a, or the degrees of the polynomials can easily be adjusted for other situations.

Notice the use of ":" instead of ";" in Figure 5. When ":" is used at the end of a statement, all computations are done but the output associated with the computations are suppressed. The command `TS := taylor(sin(x), x= Pi/2, n):` returns a truncated taylor series (with remainder term). When `TP := convert(TS, polynom):` is applied to this result, the remainder is discarded and a Taylor polynomial of degree $n-1$ is produced. In the loop structure of Figure 5, as `k` assumes values $1, 2, \ldots, 12$, `P.1, P.2, ..., P.12` become data structures representing the graphs of $\sin(x)$ and a Taylor polynomial approximating $\sin(x)$. Eventually, `display` plots in sequence (i.e., animates) all of the accumulated graphs. Four of these polynomials (of degree 0, 6, 12, and 18) are shown in Figure 6.

```
> with(plots):
> for k from 1 to 12 do
>     TS := taylor(sin(x), x=Pi/2, 2*k-1):
>     TP := convert(TS, polynom):
>     P.k := plot({sin(x), TP}, x=-3*Pi..4*Pi, -1.5..1.5):
> od:
> display([seq(P.k, k=1..12)], insequence=true);
```

Figure 5. Taylor Polynomials for $\sin(x)$

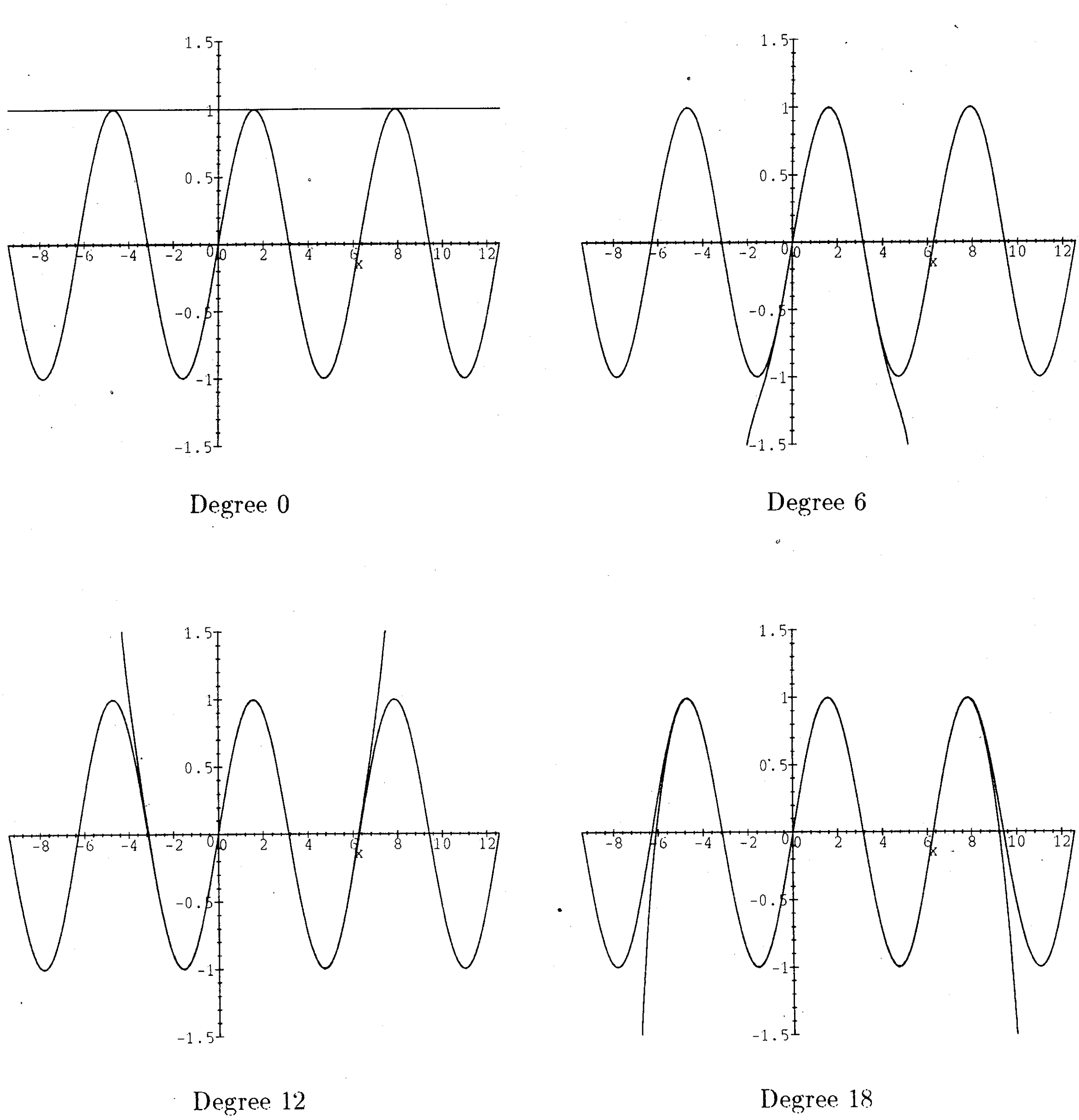


Figure 6. Frames from the Animation of the Taylor Polynomials

2.2. Discovering Mathematics

We know from experience that students learn better if they take an active part in developing the ideas that they are studying. Unfortunately, this is generally quite difficult to do in undergraduate mathematics. Most students are not ready to "discover" theorems or even offer conjectures. However, encouragement and guidance from the instructor and availability of *Maple* can change that.

Markov Matrices. Suppose, before studying the properties of Markov matrices in class, we ask students to explore some aspects of these matrices. We could, for instance, give them a specific 3×3 or 4×4 Markov matrix, M, and ask them to make some conjectures about the powers of M. Are the powers of M themselves Markov? Can anything be said about $\lim_{n\to\infty} M^n$? If X is a probability vector, what can be said about MX, M^2X, M^3X, ... and $\lim_{n\to\infty} M^nX$? For

$$M = \begin{pmatrix} 0.5 & 0.25 & 0.375 \\ 0.375 & 0.25 & 0.25 \\ 0.125 & 0.25 & 0.375 \end{pmatrix}$$

and $X = (0.3,\ 0.3,\ 0.4)^T$, the *Maple* commands given in Figure 7 can get this investigation going.

The linear algebra package of *Maple* is loaded first (the `with(linalg):` command) and then portions of this package such as `matrix` and `multiply` are used. The `multiply` command computes the product of the matrices that are given as its argument. The entries of a matrix are specified with the `matrix` command; the first argument is the number of rows, the second is the number of columns, and the numbers inside the brackets are the entries of the matrix, listed in rows.

A Matrix and the Adjoint of its Adjoint. For a more challenging exercise one could ask students to investigate the relationship between adj(adj(A)) and A, where A is an $n \times n$ matrix of real numbers (for a fuller discussion, see the paper by Auer and Muller in [2]). A natural starting point is to look at specific 2×2 matrices. This quickly leads to the conjecture adj(adj(A)) $= A$ which is true for $n = 2$ but not true in general.

The next step would be to consider the general 2×2 case (shown at the beginning of Figure 8—lines 2 and 3). This also leads to adj(adj(A))$= A$. Going on to 3×3 matrices, the student discovers that the expressions in adj(adj(A)) (i.e., output of line 5 of Figure 8) have become considerably more complicated. Although *Maple* helps with the calculations, to most students the results exhibited by *Maple* will seem intractable. Some students may need a bit of help here—perhaps they can be encouraged to look at det(A) or to factor the entries of adj(adj(A)). With this type of a hint, they are likely to conclude that for $n = 3$,

$$\text{adj}(\text{adj}(A)) = \det(A)A.$$

Again, this conjecture is valid for a specific n but is not true in general.

The expressions representing entries in adj(adj(A)) get even more complicated for $n = 4$. But factoring and looking at det(A), leads to adj(adj(A))=det$(A)^2A$. The general result, not at all simple to observe, is

$$\text{adj}(\text{adj}(A)) = (\det(A))^{n-2}A.$$

```
> with(linalg):
> M:=matrix(3,3,[.5,.5,.375,.375,
  .25,.25,.125,.25,.375]]);
> M2 := multiply(M, M);
> M3 := multiply(M, M2);
> M4 := multiply(M2, M2);
                ⋮
> X := matrix(4,1,[0.3,0.3,0.4]);
> X1 := multiply(M, X);
> X2 := multiply(M2, X);
> X3 := multiply(M3, X);
                ⋮
```

Figure 7. Markov Matrices

```
> with(linalg):
> A := matrix(2,2,[a,b,c,d]);
> B := adj(adj(A));
> A :=
matrix(3,3,[a,b,c,d,e,f,g,h,i]);
> B := adj(adj(A));
> for i from 1 to 3 do
>    for j from 1 to 3 do
>      factor(B[i, j]);
>    od
> od
                ⋮
```

Figure 8. A and adj(adj(A))

2.3. Open-Ended Explorations

Our ambition for our students is to develop them into independent mathematical thinkers. Unfortunately, most of our students do not share this vision and as is the case with most lofty objectives, we fail more often than we succeed. Nevertheless, this is a noble and worthy goal. One of the steps in making students independent is to introduce them to interesting (not necessarily difficult) problems and ask them to *investigate* the problem. This is very different from what we typically do when we teach mathematics: direct them to specific known results by posing detailed questions. Here is the type of problem I am thinking of.

Circular Coordinates. For a given point (x, y) in 2-dimensional Euclidean space, consider the two circles centered at $(x, 0)$ and $(0, y)$ with respective radii $|x|$ and $|y|$. These circles intersect at $(0,0)$ and at (u, v) (in degenerate cases (u, v) will be $(0,0)$). The coordinates u and v are called the circular coordinates of the original point (see [6] for a more detailed discussion). Without providing them with a list of questions, we can ask our students to study, individually or in a small group, the mapping $(x, y) \to (u, v)$. Initially, most of them will be at a loss—their intuition will be limited and they will not know what questions to ask. They will discover that asking interesting questions is as important as as answering questions asked by someone else.

After some initial difficulties most undergraduate mathematics students, with the help of the graphics features of *Maple*, should be able to pose some questions: What sets are invariant under this mapping? What do the images of some simple objects such as straight lines look like? (Figure 9 shows the images of parallel straight lines and Figure 10 shows the images of straight lines with a common intercept.) How might the images of straight lines be characterized? Can the idea of this mapping be generalized to 3-dimensional space, perhaps to n-dimensional space?

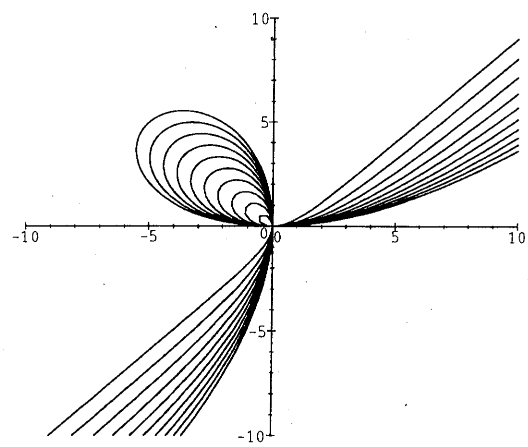


Figure 9. Images of Parallel Lines

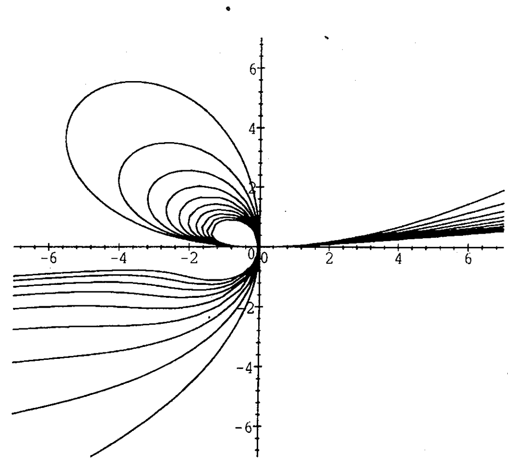


Figure 10. Images of Lines with Common Intercept

3. Some Advanced Features

The examples of the preceding section should have given you some sense of the general capabilities of symbolic computing systems. Actually, most symbolic manipulators in use today can be used in connection with a broad spectrum of mathematical and scientific problems. *Maple*, for example, has packages that deal with special areas of mathematics (we used the plotting package in Section 2.1 and the linear algebra package in Section 2.2—Figures 8 and 9). The packages listed in Figure 11 should give you an idea of the more special features of *Maple*.

Most modern symbolic computing systems also have fully developed programming languages embedded within them and all commands that are used during an interaction follow the syntax rules of that language. It is possible, therefore, to extend the capabilities of the software by adding to it user-written procedures that are of special interest. This was done in connection with the animation of the tangent lines in Section 2.1. A listing of the `TangLine` procedure is given in Figure 12 for those who are curious about the *Maple* programming language structure. The language, with its Pascal-like syntax, is built on the *Maple* platform and it derives its strength from the presence of the existing *Maple* procedures.

numapprox	combinat	DEtools	difforms	Gauss
GaussInt	geom3d	geometry	grobner	group
liesymm	linalg	logic	networks	np
numtheory	orthopoly	padic	plots	powseries
projgeom	simplex	stats	student	totorder

Figure 11. *Maple* Packages

```
TangLine := proc(expr:algebraic, a:numeric, R:range)
   local var, Expr, x, DE, slope, line, b, A, B, Line:

   with(plots, animate):

   var := op(select(type, indets(expr), name)):
   if nops([var]) = 0 or nops([var]) > 1 then
      ERROR(`The first argument must have exactly one variable in it`) fi:

   Expr := subs(var=x, expr):
   DE := diff(Expr, x):
   A:= subs(x=a, Expr):
   slope := subs(x=a, DE):
   line := A + slope*(x-a):

   B:=subs(x=b, Expr):
   Line := A + (B-A)*(x-a)/(b-a):
   animate({Expr, line, Line}, x=R, b=a..rhs(R)):

end:
```

Figure 12. *Maple* Procedure for TangLine

References

[1] Heck, A., *Introduction to Maple*, Springer Verlag, New York, 1993.

[2] Karian, Z. A. (Ed.), *Symbolic Computation in Undergraduate Mathematics Instruction*, MAA Notes Number 24, The Mathematical Association of America, 1992.

[3] Karian, Z. A. (Ed.), *Proceedings of the Denison Conference on the Use of Symbolic Computation in Undergraduate Mathematics*, The Mathematical Association of America, 1992.

[4] Karian, Z. A. and Tanis, E. A., *Probability and Statistics Explorations with Maple*, Prentice-Hall Inc., 1995.

[5] Smith, D. A., Porter, G. J., Leinbach, L. C. and Wenger, R. H. (Eds.), *Computers and Mathematics*, MAA Notes Number 9, The Mathematical Association of America, 1988.

[6] Tanis, E. A. and Kuivinen, L., "Circular Coordinates and Computer Drawn Designs", *Mathematics Magazine*, vol. 52, no. 3., 1979.

BETTER TEACHING AND LEARNING IN ELEMENTARY MATHEMATICS WITH THE SUPPORT OF DATA GENERATING AND PROCESSING SYSTEM

Ki Hyoung Oh, President
The Society of the Study of Human Education
202-1109 Yoowon 2cha Apt., Dangsan Dong 2Ka,
Youngdeungpo Ku, Seoul, Korea, 150-045
FAX 82-2-632-2771, TEL.82-2-679-2223

Introduction

The title "Better Teaching and Learning in Elementary Mathematics" is the focus point of the equalization of excellent human education. Equalization of excellent human education will contribute to the peaceful and equal co-existence and co-prosperity of all peoples in the global community. Addressing this issue is generally recognized as one of the most difficult problems facing many nations, including affluent societies. I have been pursuing this task since 1955 at the requests and with the support of a number of national and international, governmental and non-governmental agencies and organizations. This paper aims to present a comprehensive approach to this task based on research and development which have been conducted by the author and his colleagues for the last 40 years. [OH1, pp.9-35]

This approach has been pursuing improvement of the quality of teaching and learning, the efficiency and productivity of management in education and industries related to educational resources and services, and the creativity of research and development related to human education. The core of this approach is an automation system of job performances in generating and processing data to be carried out mainly by high level professionals engaging in education. We call it Data Generating Processing and System (DGPS). DGPS is an integrated system structurally related to all fields which can facilitate human education including technology. [OH2, pp.107-110]

This paper will be focused on the teaching and learning of elementary mathematics, based on the actual experience in Korean elementary schools, with sample illustrations. Later at the workshop, the processes of teaching and learning, and generating and processing data will be demonstrated.

I. The Basic Concepts related to the Illustrations

The basic concepts which are directly related to the illustrations are 'teaching and learning', 'learning characteristics of learner', 'contents of learning' and 'goals of education'.

I.1 Steps in Teaching and Learning

In elementary school mathematics education, the teaching is the role of teachers. They help learners learn. This role is performed by guidance, provision of means, and supervision. Learning is the activities of the learner. He is the 'subject' of learning activities, and not an 'object' of teaching activities. In such learning, the first step of learning is to learn by learner as teacher teaches. This is learning in teaching-learning process, namely 'learning in teaching-learning' ('L in T-L). The second step of learning is to stabilize learning based on the 'L in T-L', namely 'stabilizing learning' ('S-L'). The third step of learning is mastering stabilized learning, namely 'mastering learning' ('M-L'). The fourth step of learning is to apply what a learner has learned to the elective areas of learning,(higher level) namely 'elective learning' ('E-L'). In the 'L in T-L', the direct teacher's role, as a whole, is essential. But in S-L, M-L and E-L stages, the individual learner learns independently with the means properly provided. DGPS supports learner's independent learning in S-L, M-L, E-L, through the system including technology, with the indirect guidance and supervision of teacher. [OH3, pp.40-45]

I.2 Learning Characteristics of Learner

In school education where learners learn according to curriculum, the learning characteristics of a learner are defined as follows: [OH4, pp.227-241]

LCH = F(S. D. R.)

LCH : Learning Characteristics of Learner
F : Functions of learning of learner
Functions are classified into three, performance function(p) facilitation function (f) and creation function (c).
S : Starting point of learning by a learner
D : Ability of learning by a learner in terms of difficulty level
R : Time spent in learning of learner

I.3 Sequenced Contents of Learning

Contents of learning are sequenced by school levels, grades, terms (two terms per year in Korea), units and learning themes. In case of elementary school mathematics, contents of learning are sequenced, for instance, 6th grade, first term, 4th unit, 5th learning theme. This is coded as 610405. Learning themes are further classified into three sub-learning themes, namely common learning (CL), previous learning (PL) and elective learning (EL) themes. Common learning themes are those contained in the text book which are generally classified into four levels. The first level is coded as C-3. The second level, C-2; the third level, C-1; the fourth level, CT. The previous learning themes are sub-learning themes preceding the next learning theme according to the sequences of learning themes. They are coded as P-1, P-2, P-3, P-4 etc.. The elective learning themes are those which provide learning tasks for application of common learning themes sequenced in terms of difficulty level. They are coded as E-1, E-2, E-3, E-4 etc.. [OH5]

I.4 The Goals of Human Education

The goals of human education are classified into three. The first goals are ideal aims which clarify 'education for what and why'. They serve as the basic direction of education, basic criteria for determining overall scope and classifying it into areas of educational activities. They further serve as basic principles to be applied to the processes of educational activities. They are defined in the related laws, regulations and curriculum. The second goals are content objectives which clarify 'what to educate'. Based on the ideal aims, they serve as criteria for determining scopes and classifying them into detailed areas of learning contents such as elementary mathematics. They further serve as criteria for sequencing detailed areas of teaching and learning such as grade-term-unit-theme (610405) including the principles to be applied to the processes of teaching and learning. They are defined and described in the curriculum and the teachers manuals and illustrated in the textbooks. The third goals are progress targets which clarify 'how much to educate' based the sequences within the criteria and framework of ideal aims and content objectives.

The importance of such specifications of goals of human education is to make clear 'why', 'what' and 'how much' to teach and learn, so that both the teacher and the learner can be consistently aware of them. Such awareness in the processes of teaching and learning facilitates development of the intellectually, morally, aesthetically, religiously sound personality of learners. In addition, it promotes human health as a whole and normal and able life and activities of learners as members of communities in all fields at all levels. [OH6, pp.56-57]

II. Illustrations of Better Learning by a Sample learner

This is to illustrate DGPS, of which technology is an integral part. DGPS supports improving the quality of teaching and learning. Here a sample learner is illustrated in the actual processes of teaching and learning. He has 5 weeks experience of learning with a teacher who has one year experience of teaching supported by DGPS. The sample learner is one of 44 learners in his class. He has learned well 6104 and descriptions will be made stage by stage. [OH7]

II.1 Stages in Teaching and Learning illustrated

In the first stage of regular teaching and learning, the sample learner selected E-2, responded to 10 learning tasks, time spent per task was 26 seconds, performance index was 100. He selected E-2 based on the result of the previous class hour. The result of previous regular T-L, 15 in number of learning tasks, 35 seconds in time spent per task, 93 in performance index. He was not satisfied with the result. He decided to make improvement after regular T-L of previous class hour through homework. He actually made improvement by reducing time spent per task from 35 seconds to 26 seconds and by raising performance index from 93 to 100.

In the second stage of regular teaching and learning in classroom, the teacher starts to teach C-3, that is, sub-learning theme of common learning in group relationship with learners. The teacher teaches to help learners 'L in T-L'. The sample learner completed 'L in T-L' of C-3 before the teacher stopped teaching 'L in T-L'. He started 'S-L' and continued to 'M-L'. He learned 7 learning tasks, spent 39 seconds per task with performance index 86. Through similar processes, he completed the rest of the sub-learning themes of common learning up to CT. The results were, in C-2, 7 tasks-73 seconds per task-100 performance index, in C-1, 15 tasks-24 seconds per task-80 performance index, and in CT, 8 tasks-32 seconds per task-63 performance index, respectively. Since the passing standard of performance index is 80, his 'M-L' in CT was considered not good enough.

In the third stage of regular teaching and learning, The sample learner passed P-1, identifying that he had no learning deficiency.

The fourth stage of the regular T-L in classroom is for individual independent elective learning. By that time over 90% of the learners in the class learn independently, selecting sub-leering themes suitable to him/her. While most of the learners are learning independently, the teacher helps directly those who have not completed 'L in T-L'.

The sample learner actually selected C-1 and CT in individually independent learning to make up M-L in common learning. He made improvement, in C-1, 5 tasks-17 seconds per task-80 performance index, reducing time spent per task from 24 to 17 seconds. In CT, 5 tasks-29 seconds per task-100 performance index, reducing the time spent per task from 32 seconds to 29 seconds and raising performance index from 63 to 100. With this improvement, he continued to learn E-1. 5 tasks-39 seconds per task-80 performance index. Thus he can master one stage beyond the text book level. And he challenged E-2, but the time before the end of regular class hour was not sufficient for him to master E-2.

In the fifth stage of regular T-L in classroom, the teacher is engaged in overall analysis of T-L, the interpretation of the T-L processes, and to identify what are the things to be further improved. Each individual learner identifies tasks for improvement, application and creative activities to be carried out after regular school teaching and learning activities. The sample learner decided to improve learning E-2, and he received suitable home work materials.

II.2 Achievement of Goals

The sample learner has achieved the goals of education quite well, with balance and consistency in the regular classroom teaching and learning. In the first stage, he identified achievement of home work following 610404 as E-2. In the second stage, he passed the standard C-3, C-2 and C-1, but he did not pass CT. On the third stage, he identified that he had no deficiency for learning 610406. In the fourth stage, he improved CT and, in addition, he passed E-1. In the fifth stage, he planned to study E-2 in homework with the materials provided by the school.

With reference to the functions of learning(F) as the first item of characteristics, he started learning with suitable help of teacher in 'L in T-L', that is, he started learning with performance function(p). And he learned independently S-L, M-L, E-L, that is, he continued learning with facilitation function(f). Further he laid foundation for creative learning by extending to E-L, namely with creation function(c).

With reference to the starting point of learning(S), he had no deficiency. With reference to the ability of learning in terms of difficulty level(D), he was ready to learn E-2. The time spent per task of sub-themes of common learning(R) is 28.2 seconds, 8 tasks per stage(CLN) and 96.5 of performance index(CLP) in average per stage. He learned concepts of learning themes from C-3 to E-1 in the process of M-L, so he could write those learning tasks in written form, He had learned altogether 75 learning tasks in 33.25 minutes independently using the system during the regular class hour of 40 minutes. All through the learning process, he made no error in the utilization of system(US) and made no error in the selections of suitable learning tasks(SUL).

II.3 Learning Characteristics of the Learner in Code

These lengthy descriptions of 610405 are expressed below in a simple form using code.

LCH-24=Fc(S-0, D-2.00, R-28, CLN-8, CLP-96.5, CF-el) US-100, SUL-100

Such descriptions are produced usually in unit bases per class for improving school teaching and learning, and a copy of the unit 6104 is presented in table 1 and 2. In the copy, several more items are added. The first item is ID which is identification code designated to learners. In order to protect privacy of the learners, the number system and code systems are designed in such a way that the specific learner may not be exposed in the data processing processes. The second item is R which expresses time spent in learning per task. R in the copy is expressed by index in which 1.00 is average of the class. In case of the sample learner, R is 0.65 which expresses his time spent per task is 65% of the class average. The third item is CF, concept formation learning. 'el', 'cl' and '0' expresses concept formation learning of 'EL', 'CL' and no concept formation learning. The fourth item is percentile which is added in cases of IQ, S, ED, R. CLN and CLP. The fifth item is a correlation chart which includes correlation between IQ, S, ED, R and CLP.

II.4 Brief Analyses and Comments

As indicated in the correlation chart, correlation between all the variables are high with level of significance $P < .001$, ranging from .9121 to .6054. The high correlation empirically suggests that all the learners achieved self-realization quite high. From the normative view, the higher are the correlation the higher the level of self-realization of the learners.

Table 1. Learning Characteristics of Sample Class:6104

ID	IQ	F	S	ED	R	CLN	CLP	CF	US	SUL
av.			-0.61	1.37	1.00	8.89	89		99	92
01	116: 36.4	f	-0.33: 47.7	1.50: 52.3	0.85: 34.1	10.6 : 18.2	92: 45.5	cl	100	100
02	91: 88.6	p	-1.14: 90.9	0.14: 93.2	1.33: 93.2	4.62:100.0	76: 95.5	0	92	80
03	97: 77.3	p	-0.43: 52.3	0.71: 75.0	1.01: 59.1	9.32: 45.5	87: 81.5	0	100	87
04	115: 40.9	f	0.00: 27.3	1.71: 43.2	0.76: 22.7	10.3 : 27.3	93: 38.6	el	100	100
05	116: 36.4	f	-0.29: 45.5	1.71: 43.2	0.85: 34.1	8.80: 52.3	93: 38.6	cl	100	88
06	112: 52.2	f	0.00: 27.3	1.43: 56.8	0.74: 18.2	9.38: 43.2	88: 77.3	cl	100	100
07	110: 59.1	f	-1.86: 93.2	1.29: 61.4	0.93: 47.7	6.10: 93.2	88: 77.3	0	100	100
08	133: 2.3	c	0.00: 27.3	3.14: 2.3	0.63: 4.5	11.4 : 11.4	97: 4.5	el	100	100
09	96: 79.5	p	-0.29: 45.5	0.43: 81.8	0.88: 38.6	9.42: 40.9	87: 81.8	0	100	94
10	95: 81.8	p	-2.00: 95.5	0.00: 97.7	1.52: 97.7	7.44: 81.8	76: 95.5	0	100	83
11	120: 22.7	f	-0.14: 36.4	2.00: 29.5	0.78: 27.3	11.1 : 13.6	93: 38.6	cl	100	100
12	130: 4.5	c	-0.29: 45.5	2.86: 6.8	0.61: 2.3	11.9 : 4.5	95: 25.0	el	100	100
13	117: 31.8	f	-0.14: 36.4	1.71: 43.2	0.75: 20.5	10.6 : 18.2	90: 59.1	cl	100	100
14	73:100.0	p	-2.43: 97.7	-0.57:100.0	2.32:100.0	5.83: 95.5	66:100.0	0	100	71
15	106: 61.4	f	-1.00: 84.1	1.71: 43.2	1.17: 75.0	8.20: 68.2	95: 25.0	cl	100	100
16	101: 75.0	f	0.00: 27.3	1.14: 70.5	1.27: 86.4	8.21: 65.9	89: 65.9	cl	100	90
17	102: 70.5	f	-0.57: 61.4	1.57: 50.0	1.11: 70.5	9.20: 47.7	88: 77.3	cl	100	91
18	120: 22.7	f	-0.57: 61.4	1.57: 50.0	1.21: 81.8	7.18: 86.4	91: 52.3	cl	100	100
19	125: 11.4	f	-0.29: 45.5	2.14: 22,7	0.69: 13.6	8.33: 63.6	89: 65.9	cl	100	100
20	110: 59.1	f	-0.71: 68.2	2.00: 29.5	0.70: 15.9	11.4 : 11.4	96: 11.4	cl	100	100
21	103: 68.2	f	-1.14: 90.9	1.14: 70.5	1.02: 61.4	7.65: 77.3	86: 86.4	cl	96	80
22	119: 27.3	f	0.00: 27.3	1.14: 70.5	1.22: 84.1	8.68: 54.5	90: 59.1	cl	100	100
23	118: 29.5	f	0.00: 27.3	2.14: 22.7	0.93: 47.7	10.4 : 22.7	95: 25.0	cl	100	100
24	128: 6.8	c	0.00: 27.3	2.43: 11.4	0.65: 9.1	12.3 : 2.3	96: 11,4	el	100	100
25	83: 97.7	f	-1.00: 84.1	0.14: 93.2	1.30: 90.9	7.02: 88.6	83: 88.6	0	100	93
26	84: 95.5	f	-0.86: 75.0	0.14: 93.2	1.12: 72.7	9.88: 34.1	88: 77.3	0	100	80
27	124: 13.6	f	0.00: 27.3	2.50: 9.1	0.66: 11.4	10.2 : 29.5	93: 38.6	cl	100	100
28	89: 90.9	p	-1.00: 84.1	0.14: 93.2	1.18: 77.3	8.01: 72.7	80: 90.9	0	100	94
29	103: 68.2	p	-3.00:100.0	0.00: 97.7	1.29: 88.6	5.04: 97.7	73: 97.7	0	100	82
30	113: 47.7	f	-0.14: 36.4	2.00: 29.5	1.07: 68.2	8.01: 72.7	94: 27.3	cl	100	100
31	114: 43.2	f	-0.57: 61.4	1.86: 31.8	0.94: 54.5	8.52: 56.1	95: 25.0	cl	100	91
32	121: 18.2	f	-0.57: 61.4	2.14: 22.7	0.65: 9.1	10.5 : 20.5	92: 42.5	cl	100	100
33	104: 63.6	f	-1.14: 90.9	1.14: 70.5	1.21: 81.8	7.48: 79.5	91: 52.3	cl	100	100
34	88: 93.2	f	-1.00: 84.1	0.57: 79.5	1.05: 65.9	8.46: 61.4	91: 52.3	cl	100	92
35	122: 15.9	c	0.00: 27.3	2.14: 22.7	0.81: 29.5	11.4 : 11.4	95: 25.0	el	100	100
36	115: 40.9	f	-0.86: 75.0	1.57: 50.0	0.96: 56.8	9.75: 36.4	89: 65.9	cl	100	100
37	112: 52.3	f	-0.43: 52.3	2,14: 22.7	0.94: 54.4	8.96: 50.0	95: 25.0	cl	100	100
38	119: 27.3	f	0.00: 27.3	1.71: 43.2	0.89: 40.9	10.1 : 31.6	96: 11.4	cl	100	100
39	92: 84.1	f	-0.86: 75.0	0.71: 75.0	1.04: 63.6	8.58: 56.8	90: 59.1	0	100	86
40	91: 88.6	p	-0.71: 68.2	0.29: 84.1	1.39: 95.5	6.19: 90.9	86: 86.4	0	100	85
41	113: 47.7	f	-0.14: 36.4	1.29: 61.4	0.86: 36.4	9.73: 38.6	93: 38.6	cl	100	100
42	101: 75.0	f	-0.71: 68.2	0.57: 79.5	0.94: 54.5	7.31: 84.1	88: 77.3	0	83	71
43	127: 9.1	c	0.00: 27.3	2.86: 6.8	0.78: 27.3	10.3 : 27.3	98: 2.3	el	100	100
44	110: 59.1	f	0.00: 27.3	1.43: 58.6	0.93: 47.7	7.93: 75.0	92: 45.5	cl	100	100

Table 2. Correlation:6104

	IQ	S	ED	R
S	.6054 **			
ED	.9121 **	.6748 **		
R	-.7467 **	-.6683 **	-.7570 **	
CLP	.7196 **	.7802 **	.8390 **	-.7811 **

** P < .001

III. DGPS Support to improve The Quality of Teaching and Learning

DGPS supports improving the quality of teaching and learning in school education within the integrated system of total human education which is a sub-system of the society. Within such context, DGPS supports improvement of the quality of teaching and learning in mathematics through Education Service Automation Network for School Teaching-Learning in Elementary Mathematics. We call this DGPS supported ESANET-STL-EM system. Basic units of the system are classroom systems of schools in which file-servers and terminals for learners are networked by Local Area Network (LAN) system. In the present stage, the Central Support Service Center (CSSC) supports area service centers to support school units through Wide Area Network (WAN) system. Such systems reduce the burden of teachers as much as possible and help teachers concentrate on performing their professional jobs. For teachers, such a system improves teaching, on the one hand, and improves the learning of learners in 'L in T-L', 'S-L', 'M-L' and 'E-L' on the other. In a word, they provide means of teaching and learning, and they help both teachers and learners to concentrate on teaching and learning. Thus automated systems support better teaching and learning through data generating and processing.

(1) DGPS provides sufficient learning tasks of sub-learning themes suitable to each learner. 15 tasks per sub-learning themes from the first step to fourth step, except in the third step, each with 3 tasks per sub-learning theme. This is performed by ESA Learning Tasks Generation (ESA-LTG) system which is an integral part of DGPS. In addition, DGPS provides printed materials for teachers to prepare lesson plans and teaching for 'L in T-L', and, for learners, out of regular class learning in 'S-L', 'M-L', 'E-L'. This is performed by ESA Automated Manuscript Writing (ESA-AUTOMAWS) system which also is an integral part of DGPS. Both systems are supported by Data Generating Base (DGB) of DGPS. For the illustrated case of 6104, 76,768 tasks are necessary to be generated so that any learner has enough tasks which are suitable to him/her in 7 class hours. In addition, 5,736 tasks covering all sub-learning themes in printed materials are generated. Both systems are to generate 82,504 tasks in all. This is the primary level generating data.

(2) With the provision of the learning tasks for 6104, learners of the illustrated class actually learned with 14,261 tasks provided by ESA-LTG system and 3,942 tasks provided by ESA-AUTOMAWS. As soon as the learner learns with each task, the system checks the result of the learning giving both audio and visual signals indicating 'correct' or 'no'. In case of 'no', the system provides an opportunity to find where the error is, to correct the error, and to check whether one made the right correction or not. This is designed to serve as a hint for recalling what a learner learned in 'L in T-L. All through the processes, data are generated and installed as data indicating what sub-themes, how many tasks per each sub-systems are learned, and how much time is spent in each task classifying right responses and errors. In case 6104, the number of cases were 14,261 in all. Those data are generated and installed as generated data. These are the secondary generating and primary processing of data.

(3) Those data generated, in the form of learning tasks, in the form of results of each learner, and in the form of installed generated data, are processed at the end of regular T-L in class in the form of summaries of results of each learner. They are usually in the form of a summary of all five steps, to be basic hints for planning by both teacher and learners in the following teaching and learning. These are the tertiary generating and secondary processing of data.

(4) Those data generated and processed in the school units are collected and stored in the DGB of CSSC through WAN. The CSSC processes collected data of learning units as illustrated earlier and provides them to the schools. School teachers and administrators use them for training provided by the CSSC to improve teaching and administration creatively. The provision of supports through such systems are essentially for teachers and administrators to provide the best possible means to improve both teaching and learning. Such provisions support teachers to concentrate on their own professional activities including teaching in 'T in T-L'. And for individual learners to proceed through well balanced learning activities including S-l, M-L and E-L independently with indirect guidance of their teacher with well grounded bases. The processing of the data by CSSC is primary level generating data at CSSC for creative job performance by all professionals related to education.

(5) DGPS supports cooperation among school teachers and administrators. It further supports professionals in related fields of education including professionals in educational administration, management of industries of educational resources and services, research and development in education. DGPS supports those professionals through processing and generating data without additional input process.

IV. Improvement of Educational Administration and Management

The role of administration and management is to promote and facilitate job performances of educational professionals in schools. And the role is performed by consultation or counseling, provision of personnel and material resources and supervision of school teachers and administrators. Since school teachers and administrators are professionals, they perform their professional jobs corresponding to the job performance environments, including provision of means, resources and economic support. DGPS supports educational administration and management with focus on provision of means and conditions improving job performance environments.

Under the current situation in Korea, most of the classes in urban areas from grade 1 to 12 are organized with around 40 learners, due to population migration trends. In rural and remote areas, classes are smaller, with less than 20, and some in multi-grade with 2 to 3 grades. The large classes are in advantaged areas and smaller classes are in disadvantaged areas. Consequently there is a serious educational gap between large classes in advantaged areas and smaller classes in disadvantaged areas. In one school with DGPS support in a disadvantaged area, achievements of goals are at the same level as those in urban areas. This suggests that the number of learners per teacher can be increased up to 80-100% in comparison with advanced countries. This means DGPS may improve raising efficiency in educational management.

V. Improving Management of Industries related to Education

According to the views of human education, education is pursuing values acceptable universally by developing and utilizing the best possible means. In other words, improving management of production, circulation and consumption of educational means is to contribute to the realization of the values. Based on such views, DGPS is designed to support reducing unit costs of educational means including S/W, H/W and services. Thus DGPS improves educational environments with the same amount of educational expenses.

In case of S/W, through Authoring Support System, which is one of the sub-systems of DGPS, it supports reducing expenses to one fifth in average. In case of H/W, DGPS supports, through ESANET-STL system, reducing expenses to one third in comparison with consuming PCs available in the market. For educational services, DGPS supports reducing expenses at least to one tenth. In other words, DGPS directly supports raising efficiency and productivity in educational management and indirectly pioneering the industries related to education.

VI. Facilitation of Creativity in Research and Development

Equalization of excellent human education is very important and it should be pursued consistently in spite of difficulties. A sense of mission and responsibility made me and my colleagues pursue it for the last 40 years. And it inspires us with continuation of R & D and demonstrations for coming decades.

It is imperative to improve creatively the quality of teaching by training creative school professionals and educational administrators and managers of industries directly related to human education. DGPS supports the improvement of the quality of teaching, even in classes with more than 40 learners in urban areas as well as small and multi-grade classes in disadvantaged areas. DGPS supports improving educational environments by creative educational administration through reducing unit costs of educational resources and services. For these creative developments in human education, DGPS facilitates creativity in research, development and demonstration for all professionals engaging in human education.

VII. Perspectives

Looking toward the coming age of cyberspace and the global village, we can assure ourselves that we can equalize excellent human education. With elementary school mathematics as a case for illustration, I have attempted to describe raising the quality of teaching and learning, school management, management of industries related to education and creativity of R & D in education through DGPS. Though it is limited to school elementary mathematics, considering the fact that one hundred million children are entering 1st level schools every year, equalizing excellent education in elementary mathematics can be a good starting point. The author, the representative of copyright holders, has been preparing for the dedication of royalty of DGPS supported ESANET-STL-EM basic system. He is determined to assure that the dedicated system is utilized only for the equalization of excellent human education that contribute to the equal and peaceful co-existence and co-prosperity of global community members. [OH8]

References

[OH1] K. H. Oh, *The Basic Tasks of Humanization of Education*, Seoul:SSHE, 1991, pp.9-35.

[OH2] K. H. Oh, *Human Education Development Plan:1994.3-2000.2,(School Year 1995)* Seoul:SSHE, 1995. pp.107-110.

[OH3] K. H. Oh, *Manual for DGPS Supported Normative and Empirical Research in Education*, Seoul: SSHE, 1994. pp.40-45.

[OH4] K. H. Oh, *"Suitable Learning Task Approach to Teaching Learning"*, Proceedings of Pre-congress Conference, The IVth World Congress of Comparative Education Societies, 1980, pp.227-241.

[OH5] K. H. Oh, "*Education in School and Democracy*", The Journal of Educational Research, Korean Society for the Study of Education, Vol. 18, No. 1, 1980.

[OH6] K. H. Oh. *Principles of Human Education*, Seoul:Kyoyookkwahaksa, 1983, pp.56-57.

[OH7] K. H. Oh, *Manual for DGPS Supported Normative and Empirical Research in Education*, Seoul:SSHE, 1994.

[OH8] K. H. Oh, *Procedures for the Dedication of ESANET-STL-EM*, Seoul:SSHE, 1994.

Computer Algebra – Where we are?, Where we go?

Tateaki Sasaki

Institute of Mathematics, University of Tsukuba
Tsukuba-shi, Ibaraki 305, Japan

Abstract

We list some epoch-making events in computer algebra in **1**, point out prominent features and thema of algebraic systems in **2**, prominent features and thema of algebraic algorithms in **3**, and mention where we will or should go next in **4**.

0 Brief summary of the talk

After the birth in about 1960 [Appl, Systm], computer algebra has attained astonishing development in these 35 years. By middle 1970's, elementary operations on polynomials and rational functions, such as GCD and factorization [Factri], have already been made sufficiently efficient, and remarkable advancement has been attained also for most mathematical operations that non-mathematics course students learn, such as indefinite integral [IntDE] and infinite series. After 1980, algorithm research went to higher mathematics, such as ideal and ring theory [Book], and is now building a wonderful realm there. The fruit of algorithm research has immediately been taken into computer algebra systems, and many powerful and useful general-purpose systems have been constructed so far: REDUCE and MACSYMA representing 1970's - 1980's, and Axiom, Maple and Mathematica representing late 1980's - 1990's [Systm]. These systems are now used by not only researchers but also teachers and students. The application areas of computer algebra are now ranging from mathematics and physics to most of theoretical science and engineering and even to economics [Appl]. Summarizing the conventional computer algebra, we may say as follows.

$$\begin{aligned}\text{[Computer Algebra so far]} &= \text{[Mathematics]} + \text{[Efficiency]}\\ &\Rightarrow \textit{Computerization of Mathematics}\end{aligned}$$

Then, where will computer algebra go ? Undoubtedly, computerization of mathematics will go further, which is, although satisfactory from the viewpoint of mathematics, never satisfactory from the viewpoints of users. The reason is lack of practicality. For example, for algebraic equation, conventional computer algebra tries to find algebraic numbers as exact solutions, which is usually quite difficult and requires much time. However, for many users, algebraic numbers are useless; they want to know approximate values of the solutions. In applications, getting approximate answers quickly is often much more important than getting exact answers after many hours of computation. Practicality is a leading principle in numeric computation, and being guided by this principle, numerical analysis has built a unique realm. Recent researches of the author's group revealed that the principle of practicality also leads computer algebra to a new and fruitful world which we call *approximate algebra* [AppAlg]. Therefore, the author proposes the following new way.

$$\begin{aligned}[\text{Computer Algebra from now}] &= [\text{Math.}] + [\text{Effic.}] + [\text{Practicality}] \\ &\Rightarrow \textit{Infrastructural Technology}\end{aligned}$$

1 Epoch-making events in computer algeba

The following is only an incomplete and restricted list of the events. For more details, see the selected bibliography at the end of this paper.

1959: computer manipulation of polynomials in celestial mechanics [Appl].
1961: computer program calculating freshman-level integrals (Slagle [Systm]).
1965: **FORMAC** (Sammet [Systm]), first general-purpose system.
1965: Gröbner basis construction algorithm (Buchberger [Book]).
1966: subresultant algorithm for polynomial GCD (Collins [Book]).
1967: univariate factorization algorithm over $\mathbf{Z}_p$ (Berlekamp [Factri]).
1969: integration algorithm of elementary functions (Risch [IntDE]).
1969: univariate factorization algorithm over $\mathbf{Z}$ (Zassenhaus [Factri]).
1968 – 73 : large-scale application to quantum electro-dynamics [Appl].
1972 – : **REDUCE** (Hearn [Systm]), world-wide distributed system.
1973 – 90 : **MACSYMA** (Moses, et al. [Systm]), big mathematical system.
1973: modular multivariate GCD algorithm (Moses and Yun [Book]).
1973: multivariate factorization algorithm (Wang and Rothschild [Factri]).
1974 – : **CAYLEY/Magma** (Cannon [Systm]), a system for group theory.
1974 – : **SCRATCHPAD/Axiom** (Jenks et al. [Systm]), system with abstract data-types.
1975: sylindrical decomposition algorithm (Collins [Book]).
1978: geometry-theorem proving with characteristic set (Wu [Book]).
1979 – : **μ-MATH/Derive** (Stoutemyer and Rich [Systm]), system on micro-computer.
1980 – : **Maple** (Geddes et al. [Systm]), compact and portable general-purpose system.
~1975 – : application of Gröbner basis to commutative algebra [Book].
~1980 – : study of automated geometry-theorem proving [Book].
~1985 – : study of parallel execution of algebraic computation [Book].
1987 – : **Mathematica** (Wolfram et al. [Book]), symbolic-numeric-graphic system.
1989 – : study of approximate algebra [AppAlg].

2 Features and thema of algebraic systems

Modern and general-purpose algebra system is self-containing in that it is equipped with very low to high level mathematical operations, pretty complicated because it contains many data-types categorized variously, it is a long-term growing system, and is an integrated system with facility of symbolic, numeric and graphic computations.

2.1 Self-containing system

In numeric computation, the programming language system, such as FORTRAN or C, provides us with many numeric operations, such as evaluating elementary functions, solving linear and algebraic equations, computing eigen-values and eigen-vectors, calculating various kinds of definite integrals, and so on. In algebraic computation, the programming language system, such as Lisp or C, does not provide us with any algebraic operation, and each algebra system must be equipped with even very low level operations, such as polynomial arithmetic, polynomial GCD, etc. Therefore, general-purpose algebra system becomes of large-scaled. In fact, more than 100 man-year power has been devoted to develop **MACSYMA**.

2.2 Many data-types categorized variously

A mathematical object is usually categorized variously. For example, polynomial, the most common data-type in computer algebra system, may be categorized as either an element of a ring over $\mathbf{Z}_p$, $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{Q}(\alpha)$, $\mathbf{R}$, or $\mathbf{C}$ (may be represented approximately by floating-point numbers), an element of a field, of an integral domain or of a Euclidean domain, etc. In addition to polynomials, we must also handle algebraic numbers, algebraic functions, rational functions, and many other mathematical objects. Furthermore, these many data-types must be organized hierarchically. Thus, a general-purpose computer algebra system is quite complicated.

One idea to organize many various algebraic data-types hierarchically is that the system itself does not prepare every necessary data-type but it provides the user with a facility of defining new data-types hierarchically. **SCRATCHPAD/Axiom** has been constructed on this idea. However, providing such a facility itself is not easy. How to handle many various data-types is still a problem in designing algebra system.

It should be commented, however, that data-types may not be categorized so finely as mentioned above in the system for scientific and technological applications.

2.3 Long-term growing system

In each algebra system, there is a loose hierarchy on algorithms, too; at the lowest level we have algorithms on numbers, such as integer GCD and factorization, at the second lowest level, we have many algorithms on polynomials, such as polynomial GCD and factorization, and so on. In addition, there is a hierarchy among algorithms on each mathematical operation; for example, [Univariate factorization over $\mathbf{Z}_p$] $\prec$ [Univariate factorization over $\mathbf{Z}$] $\simeq$ [Univariate factorization over $\mathbf{Q}(\alpha)$] $\prec$ [Multivariate factorization over $\mathbf{Z}$] $\prec \cdots$. Considering operations in higher algebra, we can observe many levels of hierarchy.

The above-mentioned hierarchy forces us to construct corresponding procedures hierarchically. Furthermore, compared with procedures of numeric computation, procedures of algebraic computation require much more time to program. Hence, a long-term is necessary to develop general-purpose system, and the system will continue to grow several decades usually. In fact, **REDUCE** has been growing continuously these 25 years.

2.4 Integration of symbolic, numeric, graphic computations

Drawing graphs is absolutely necessary in school mathematics. Similarly, graphics is very useful in algebra system, too. **Mathematica** is the first system that is equipped with facility of integrating symbolic, numeric and graphic computations, and it has achieved a great fame by this facility. Today, every modern general-porpose system is equipped with this facility.

It should be commented that current symbolic-numeric integrated computation is very limited in ability. We will explain this in **4.3**.

3 Features and thema of algebraic algorithms

So far, quite many algebraic algorithms of various kinds have been invented, many of them are quite efficient. In this section, we do not explain individual algorithm separately but point out prominent features and thema of algebraic algorithms complehensively.

3.1 Canonical forms and reductions

Many algebraic algorithms can be viewed as procedures which transform given expressions into unique *canonical forms*. For example, the complete partial fraction representation is a canonical form of rational function, and rational function in $\sin(x)$ and $\cos(x)$ can be transformed into a canonical form by the replacements $\sin(x) \to (y - y^{-1})/2i$ and $\cos(x) \to (y + y^{-1})/2$, with $y = e^{ix}$, and by transforming the resulting rational function in y into a canonical form.

Practically very important case is a set of polynomials

$$\big\{P_1(x,y,\dots,z),\ \cdots,\ P_r(x,y,\dots,z)\big\} \subset K[x,y,\dots,z], \quad K = \text{a number field.}$$

In order to define a canonical form, we introduce two operators, a term order operator $\succ$ and and a term elimination operator $\mathcal{E}$. For example, we define $\succ$ by the degree w.r.t. main variable x and $\mathcal{E}$ by the following *leading term elimination.*

$$\mathcal{E}_{\mathrm{lt}}(P_i, P_j) \ = \ \frac{\mathrm{lcm}}{\mathrm{lt}(P_i)} \cdot P_i - \frac{\mathrm{lcm}}{\mathrm{lt}(P_j)} \cdot P_j \quad (i \neq j),$$

where $\mathrm{lt}(P)$ is the leading term of P (highest degree term w.r.t. x),

and lcm = Least-Common-Multiple of $\mathrm{lt}(P_i)$ and $\mathrm{lt}(P_j)$.

Note that the leading term in $\{P_i, P_j\}$ is eliminated by the operation $\mathcal{E}_{\mathrm{lt}}$, and we have

$$\mathcal{E}_{\mathrm{lt}}(P_i, P_j) \ \prec \ \text{higher order element of } \{P_i, P_j\}.$$

Therefore, the leading term elimination can be viewed as a *reduction* w.r.t. $\succ$.

Applying $\mathcal{E}_{\mathrm{lt}}$ to a set of two polynomials $\{P_1, P_2\}$ successively, we can generate a polynomial remainder sequence (*rem* denotes the remainder operation)

$$\big(R_1 = P_1,\ R_2 = P_2,\ R_3 = c_3\,\mathrm{rem}(R_1, R_2),\ \cdots, R_k = c_k\,\mathrm{rem}(R_{k-2}, R_{k-1})\big),$$
$$\text{where } \ c_i \in K(y,\dots,z),\ R_i \in K[x,y,\dots,z] \quad (i = 3,\dots,k).$$

If $R_{k+1} = \mathrm{rem}(R_{k-1}, R_k) = 0$ then $R_k = c\,\mathrm{gcd}(P_1, P_2)$, $c \in K(y,\dots,z)$, and (R_k) is a canonical form of the ideal (P_1, P_2) over $K(y,\dots,z)$. Applying $\mathcal{E}_{\mathrm{lt}}$ to a set of r polynomials $\{P_1,\cdots,P_r\}$ successively, we obtain Ritt-Wu's characteristic set $\{R_1, R_2,\cdots,R_s\}$ which plays an essential role in determining the common zeros of $\{P_1 = 0,\ \cdots,\ P_r = 0\}$ algorithmically [Book].

We can define $\succ$ by a monomial order $\succ_{\mathrm{mono}}$ which orders every monomial in $K[x,y,\dots,z]$ uniquely up to numeric coefficients, and define $\mathcal{E}$ by the following *head term elimination.*

$$\mathcal{E}_{\mathrm{ht}}(P_i, P_j) \ = \ \frac{\mathrm{lcm}}{\mathrm{ht}(P_i)} \cdot P_i - \frac{\mathrm{lcm}}{\mathrm{ht}(P_j)} \cdot P_j \quad (i \neq j),$$

where $\mathrm{ht}(P)$ is the head term of P (highest order monomial w.r.t. $\succ_{\mathrm{mono}}$),

and lcm = Least-Common-Multiple of $\mathrm{ht}(P_i)$ and $\mathrm{ht}(P_j)$.

In this case, if $\mathrm{ht}(P_j)\,|\,\mathrm{ht}(P_i)$ hence $\mathrm{lcm} = \mathrm{ht}(P_i)$ then $\mathcal{E}_{\mathrm{ht}}(P_i, P_j) \prec P_i$ and the above operation defines a reduction w.r.t. $\succ_{\mathrm{mono}}$. Applying $\mathcal{E}_{\mathrm{ht}}$ to every pair of elements of $\{P_1,\cdots,P_r\}$ successively, we obtain a Gröbner basis $\{Q_1, Q_2,\cdots,Q_s\}$ of the polynomial ideal $(P_1,\cdots,P_r)$. This is the essence of celebrated Buchberger's algorithm [Book].

3.2 Intermediate expression swell and modular method

In performing algebraic computation, we often encounter large-sized expressions during the computation, although the initial and final expressions are small-sized or medium-sized. This phenomenon is called *intermediate expression swell.* We show this phenomenon in the GCD calculation by Euclidean algorithm.

Example 1 Calculation of GCD by pseudo-remainder sequence, where *pseudo-remainder* of polynomials F and G, with $\deg(F) \geq \deg(G)$, is defined as follows.

$$\mathrm{prem}(F,G) \;=\; \mathrm{rem}\Big(\mathrm{lc}(G)^{\deg(F)-\deg(G)+1}F,\ G\Big),$$

here $\mathrm{lc}(P)$ is the leading coefficient (coefficient of the leading term) of P. Let P_1 and P_2 be polynomials given as $P_1 = x^8+x^6-3x^4+4x^3-5x^2+4x-3$ and $P_2 = 3x^6-2x^4-5x^2+7x-13$. The pseudo-remainders, with initial polynomials P_1 and P_2, are calculated as follows.

$$\begin{aligned}
P_3 = \mathrm{prem}(P_1,P_2) &= 3 \times (-2x^4+15x^3+19x^2+x+38),\\
P_4 = \mathrm{prem}(P_2,P_3) &= 243 \times (-1493x^3-1705x^2-473x-3286),\\
P_5 = \mathrm{prem}(P_3,P_4) &= 1417176 \times (-29152x^2-20590x-11421),\\
P_6 = \mathrm{prem}(P_4,P_5) &= 2175722297384036782464 \times (108918x-577823),\\
P_7 = \mathrm{prem}(P_5,P_6) &= (\text{a very long integer}) \times (1).
\end{aligned}$$

Finally, we find that $\gcd(P_1,P_2) = 1$ because $\deg(P_7) = 0$. How wastful computation we have done to get this simple result ! □

The intermediate expression swell occurs so frequently, and it makes the corresponding computation very slow. In order to perform the computation efficiently, we must suppress the expression swell almost completely. Fortunately, there is a general technique to avoid the swell, the so-called *modular method.*

We explain how the modular method improves the computation in Example 1, where the expression swell occurs only in the numeric coefficients. The idea is very simple: we perform the computation modulo some big prime p or primes $p_1,\ p_2,\ \ldots,\ p_k$. If $p/2$ or $p_1p_2\cdots p_k/2$ is larger than any |coefficient| of the final answer, and if the primes do not divide leading coefficients of $P_1,\ P_2,\ \cdots,\ P_7$, which we say that the primes are *lucky*, then we can calculate the final answer correctly. For details of modular methods, see the literature [Book].

3.3 Construction of target expressions directly

There are various kinds of modular methods. Currently, most efficient algorithms for polynomial GCD and factorization, multivariate as well as univariate, adopt the modular method. Furthermore, the method is quite effective for Gröbner basis construction, too. The following Example 2 explains how multivariate factorization is performed by a modular method.

Example 2 Let $F(x,y,z) = x^2 + (-yz - z^2 + y + z + 2)x + yz^3 - y^2z - z^3 - yz + 2z$. Let $S = (y,z)$ be the ideal generated by y and z. We first factorize $F(x,y,z)$ modulo S, which is nothing but factorization of univariate polynomial $F(x,0,0)$:

$$F(x,y,z) \equiv F(x,0,0) = x^2+2x = x(x+2) \pmod{(y,z)}.$$

With $G^{(0)}(x) = x$ and $H^{(0)}(x) = x + 2$ as factors of F modulo S, the generalized Hensel construction allows us to calculate modular factors $G^{(k)}(x, y, z)$ and $H^{(k)}(x, y, z)$ satisfying

$$F(x, y, z) \equiv G^{(k)}(x, y, z) H^{(k)}(x, y, z) \pmod{(y, z)^{k+1}}, \quad k = 1, 2, \cdots.$$

In our case, $G^{(1)}$, $H^{(1)}$, $G^{(2)}$, $H^{(2)}$ are calculated uniquely as follows.

$$G^{(1)}(x, y, z) = x + z, \qquad H^{(1)}(x, y, z) = x + 2 + y,$$
$$G^{(2)}(x, y, z) = x + z - yz, \quad H^{(2)}(x, y, z) = x + 2 + y - z^2.$$

We see that $G^{(2)}$ and $H^{(2)}$ are factors of F : $F(x, y, z) = G^{(2)}(x, y, z) H^{(2)}(x, y, z)$. □

In order to find factors from the modular factors correctly, we need several additional steps, but Example 2 explains the essence of a multivariate modular method.

Example 2 gives us an impression that the computation there is performed not to derive the answer through many procedual steps but to construct the target expression directly. In fact, $G^{(0)}$ and $G^{(1)}$ are parts of the answer $G^{(2)}$. This style of computation allows us to get the answer very quickly. We comment that Risch's algorithm for computing indefinite integrals of elementary functions also constructs the final answer directly.

3.4 Main fault - impracticality

Today, numeric computation is an infrastructural technology in scientific and industrial worlds. On the other hand, algebraic computation has not attained such a position yet, although everybody will agree that computer algebra systems are very useful and powerful tools for theoretical researches. Why does this difference arise ? The main reason, the author thinks, is lack of practicality in algebraic computation.

Current computer algebra system seems to be satisfactory from the viewpoint of mathematics; it gives us mathematically correct answers, although it often spends large computation time. It is, however, never satisfactory from the viewpoint of practical applications. Consider, for example, solving univariate algebraic equation $f(x) = 0$. Even if $\deg(f) = 100$, numeric computation gives us all the numeric roots to precision 10^{-10} within a second. On the other hand, even if $f(x)$ is as simple as $\deg(f) = 5$, algebraic computation may not give any answer or it will output algebraic numbers of large expression-size. From the viewpoint of practical users, numeric values of roots are much more useful, even if they are approximate, than exact roots in non-numeric expressions.

In applications, most users want to get the answers in simple, concise, and easily understandable forms within a moderate time. From the viewpoint of practical users, we can point out the followings as unsatisfactory points of conventional algebraic computation.

1. No answer is obtained sometimes: solving algebraic equations, etc.

2. Very difficult to get answer sometimes: computation on special functions, etc.

3. Large computational time is required usually $\Leftarrow$ exact computation.

4. Large-sized expression is output usually $\Leftarrow$ exact representation of expression.

Since these points are intrinsic properties of conventional algebraic computation, innovatory advancement of the computational method will be necessary to improve the points.

4 Where we go next ?

Summarizing the computer algebra so far, we may say

$$\begin{aligned}[\text{Computer algebra so far}] &= [\text{Constructive algebra}] + [\text{Efficiency}] \\ &= [\text{Computerization of classical algebra}].\end{aligned}$$

Then, what happens next ? Apparently, computerization of algebra will continue, to a wider range and to a higher level, i.e., to computerization of mathematics. Furthermore, computer algebra will go outside universities and launch into new worlds, in particular, into high/middle schools and industrial world. However, launching into new worlds will require innovatory advancement of algebraic computation, as we will explain below.

4.1 Towards "computer mathematics"

We may say that computer algebra so far is, roughly speaking, restricted to handling purely algebraic expressions: we usually did not handle mathematical expressions composed of logical symbols $\vee$, $\wedge$, $\forall$, $\exists$, $\neg$, etc., relational symbols $>$, $\geq$, $\neq$, etc., or set symbols $\cup$, $\cap$, $\subset$, $\subseteq$, $\in$, $\notin$, etc. (Of course, in several branches of computer algebra, e.g., automated geometry-theorem proving, some of such expressions have already been handled so far.)

Future computer algebra, or we had better say "computer mathematics", will surely handle logical, relational, and set-theoretic expressions. As for logical expressions, study of *computational logic* has been performed intensively so far, and recent study revealed that computer algebra and its techniques are useful for efficient processing of logical expressions. Therefore, research towards integration of algebraic and logical computations will begin.

4.2 When launching into high/middle schools

Today, teaching how to use computer algebra system is being included into university curriculums. However, teaching mathematics in high/middle schools by using computer algebra system is only in an early experimental stage. The existing computer algebra systems are designed to be used by highly educated users only, and they are not suited for teaching elementary mathematics. The author thinks that algebraic system for middle-level education will be almost completely different from current ones. Which kind of system is the best for such education ? This is a big question for us.

4.3 When launching into industrial world

In **3.4**, we pointed out that conventional algebraic computation lacks practicality. In industrial world, practicality is one of the most important properties of computation. Remember that the most useful computational method in industrial world is numerical simulation, and it is so practical that it can be applied to almost any problem in industrial world.

Most of the conventional algebraic algorithms are for deriving exact expressions or solutions which can be obtained in idealized cases only. On the other hand, most mathematical problems in industrial world cannot be idealized, because of complicated shapes of objects or boundaries, various kinds of contaminations, and so on. In order that algebraic computation is utilized widely in industrial world, the algorithms must be flexible enough to cope with these complications, while conventional algebraic algorithms are too rigid to cope with. Furthermore, algebraic systems must be able to handle expressions with floating-point number coefficients freely, because computation in industrial world is performed mostly with floating-point number arithmetic.

One may think that we can make algebraic computation practical by combining it with numeric computation. This idea is very old; it appeared in the middle of 1970's [SymNum]. Since then, many systems have been constructed to combine numeric programming language FORTRAN with symbolic programming language Lisp. Such systems allow us to perform only "patchwork-like" symbolic-numeric computation, that is, the computation is purely numeric in one stage and purely algebraic in another stage, and algebraic computation remains still not so practical. In order to make algebraic computation truely practical, it will be necessary to combine it with numeric computation in algorithmic level.

Combining symbolic computation with numeric one in algorithmic level is not so easy as one thinks. The reason is as follows. Conventional symbolic algorithms rely on discrete mathematics in which discreteness of integers and rational numbers plays an essential role, while numeric algorithms rely on analysis for which continuity of real and complex numbers is essential. In order to make the algebraic computation very useful industrially, the author thinks that we must develop flexible algebraic algorithms by utilizing continuity of real and complex numbers.

4.4 A new approach - approximate algebra

In order to make the algebraic algorithms flexible, the author has proposed "approximate algebra" recently and is developing various approximate algebraic algorithms in collaboration with his colleagues. The algorithms developed so far can be classified into two classes.

Class 1 : Algorithms handling polynomials and rational functions approximately.

Class 2 : Algorithms handling power-series (infinite theoretically, but truncated practically) with integral and/or fractional powers.

For algorithms in Class 1, we define *norm* $\|P\|$ of polynomial P as follows.

$$\|P\| \overset{\text{def}}{=} \text{maximum of the |numeric coefficient|'s of } P.$$

As a representative of algorithms in Class 1, we explain *approximate factorization.* Suppose a multivariate polynomial $F(x, y, \ldots, z)$ is decomposed as

$$F(x,y,\ldots,z) = G(x,y,\ldots,z)\,H(x,y,\ldots,z) + \delta F(x,y,\ldots,z), \quad \|\delta F\|/\|F\| = \varepsilon \ll 1.$$

Then, we say that F is factorized approximately into G and H with *accuracy* ε.

Example 3 $F(x,y) = x^3 + 3.0yx^2 + (2.2y^2 - 1.7)x - 2.9y$, hence $\|F\| = 3.0$.
Although F is absolutely irreducible, i.e., irreducible over $\mathbf{C}$, it can be decomposed as follows.

$$F(x,y) = (x^2 + 1.3yx - 1.7) \times (x + 1.7y) - 0.01(y^2x + y).$$

Therefore, F is approximately factorizable with accuracy $\sim 10^{-2}$. □

One may think that approximate factorization is very difficult to perform because of the existence of "extra term" δF. However, it can be done rather easily; in fact, we can perform approximate factorization efficiently by the generalized Hensel construction with floating-point number arithmetic and linear algebraic operations on numeric matrix [AppAlg].

In industrial world, we often handle polynomials determined experimentally or contaminated by various reasons. Conventional factorization algorithms are completely powerless for such polynomials; they are powerless even for polynomials with floating-point number coefficients. However, approximate factorization algorithm is applicable to such polynomials without any difficulty, telling us the norm of extra terms. It should be noted that the algorithm is useful for not only practical but also mathematical problems. For example, we can check absolute irreducibility of multivariate polynomials efficiently by using the algorithm. Furthermore, principal idea of approximate factorization leads us to a unified method of various kinds of multivariate polynomial factorizations [AppAlg].

The author thinks that "approximation" is a key to develop flexible algebraic algorithms. This may be understood by comparing numeric with algebraic algorithms of solving univariate algebraic equation; approximation changes the mathematical properties of solutions completely, enabling us to calculate approximate solutions unbelievably simply. Similarly, approximation is quite effective for solving multivariate algebraic equation. It is well-known that solutions of bivariate algebraic equation $F(x,y) = 0$, w.r.t. the main variable x, can be expressed by Puiseux series (fractional-power series) in y. Compared with conventional solutions expressed in radicals, solutions in Puiseux series can be calculated without any mathematical difficulty. Only one "fault" of Puiseux series solutions is that they are approximate ones because we can calculate only truncated series and not infinite series.

Conventionally, Newton-Puiseux's method is used to calculate the Puiseux series solutions, which is very inefficient. Recently, the author has found in collaboration with his colleagues an efficient algorithm of calculating Puiseux series solutions of multivariate as well as bivariate algebraic equations. The algorithm is based on the generalized Hensel construction and is being used to solve various algebraic problems practically [AppAlg].

Selected bibliography

[Book] Books on computer algebra.

E. R. Berlekamp, *Algebraic Coding Theory*, McGraw-Hill, New York, 1968. B. Buchberger, G. E. Collins and R. Loos (Ed.), *Computer Algebra - Symbolic and Algebraic Computation*, Springer-Verlag, 1982; B. Buchberger, Gröbner bases: An algorithmic method in polynomial ideal theory, *Recent Trends in Multidimensional Systems Theory*, Section 6, Reidel, Dordrecht, 1985; J. H. Davenport, Y. Siret and E. Tournier, *Computer Algebra - Systems and Algorithms for Algebraic Computation*, Academic Press, London, 1988; S. Wolfram, *Mathematica - A System for Doing Mathematics by Computer*, Addison-Wesley, 1988; S. C. Chou, *Mechanical Geometry Theorem Proving*, Reidel, Dordrecht, 1988; R. E. Zippel (Ed.), *Computer Algebra and Parallelism* / Lect. Notes Comp. Sci., **584**, Springer-Verlag, 1992; K. O. Geddes, S. R. Czapor and G. Labahn, *Algorithms for Computer Algebra*, Kluwer Academic Pub., Boston, 1992; T. Becker and V. Weispfenning, *Gröbner Bases - A Computational Approach to Commutative Algebra*, Springer-Verlag, 1993.

[Appl] Review papers on earlier applications of computer algebra.

W. H. Jefferys, Automated algebraic manipulation in celestial mechanics, C. ACM **14** (1971); D. Barton and J. P. Fitch, General relativity and the application of algebraic manipulation systems, *ibid*; A. C. Hearn, Application of symbolic manipulation in theoretical physics, *ibid*.

[Systm] Papers/Manuals on systems on algebraic computation.

J. R. Slagle, A heuristic program that solves symbolic integration problems in freshman calculus, J. ACM **14** (1963), 4; J. E. Sammet and E. Bond, Introduction to FORMAC, IEEE Trans.

Electr. Comp. **13** (1965), 4; A. C. Hearn, REDUCE-2 User's Manual, Dept. Comp. Sci., Univ. Utah, 1973; The MATHLAB Group, MACSYMA Reference Manual, Lab. Comp. Sci., M.I.T., 1974; J. J. Cannon, A general purpose group theory program, *Proc. 2nd Conf. on Theory of Groups* / Lect. Notes Math., **372** (1974), 204; A. Rich and D. R. Stoutemyer, Capabilities of μ-MATH 79 computer algebra system for the Intel-8080 microprocessor, *Proc. EUROSAM '79* / Lect. Notes Comp. Sci. **72** (1979), 241; R. D. Jenks and B. M. Trager, A language for computational algebra, *Proc. SYMSAC '81*, ACM (1981), 6; B. W. Char, K. O. Geddes, W. M. Gentleman and G. H. Gonnet, The design of Maple: a compact, portable and powerful computer algebra system, *Proc. EUROCAL '83* / Lect. Notes Comp. Sci. **162** (1983), 101.

[Factri] Papers on polynomial factorization.
E. R. Berlekamp, Factoring polynomials over finite fields, Bell System Tech. J. **46** (1967), 1853; H. Zassenhaus, On Hensel Factorization I, J. Number Theory **1** (1969), 291; P. S. Wang and L. P. Rothschild, Factoring multivariate polynomials over the integers, Math. Comp., **29** (1975), 935; A. K. Lenstra, H. W. Lenstra and L. Lovasz, Factoring polynomials with rational coefficients, Math. Ann., **261** (1982), 515.

[IntDE] Papers on integration and differential equation.
R. H. Risch, The problem of integration in finite terms, Trans. AMS **139** (1969), 167; J. Davenport, *On the integration of algebraic functions*, Lect. Notes Comp. Sci., **102**, Springer-Verlag, 1981. M. F. Singer, B. D. Saunders and B. F. Caviness, An extension of Liouville's theorem on integration in finite terms, *Proc. SYMSAC '81*, ACM (1981), 105; P. S. Wang, Symbolic evaluation of definite integrals by residue theory in MACSYMA, *Proc. IFIP '74*, North-Holland (1974); M. F. Singer, Liouvillian solution of n-th order homogeneous linear differential equations, AJM **103** (1980), 661; S. Watanabe, A technique for solving ordinary differential equation using Riemann's P-functions, *Proc. SYMSAC '81*, ACM (1981), 36.

[SymNum] Review papers on symbolic-numeric computation.
W. S. Brown and A. C. Hearn, Applications of symbolic algebraic computation, Comp. Phys. Commun. **17** (1979), 207; E. W. Ng, Symbolic-numeric interface : a review, *Proc. EUROSAM '79* / Lect. Notes Comp. Sci., **72** (1979), 330.

[AppAlg] Papers on approximate algebra.
T. Sasaki and M-T. Noda, Approximate square-free decomposition and root-finding of ill-conditioned algebraic equations, J. Inform. Proces., **12** (1989), 159; T. Sasaki and M. Sasaki, Analysis of accuracy decreasing in polynomial remainder sequence with floating-point number coefficients, J. Inform. Proces., **12** (1989), 394; M. Ochi, M-T. Noda and T. Sasaki, Approximate greatest common divisor of multivariate polynomials and its application to ill-conditioned systems of algebraic equations, J. Inform. Proces., **14** (1991), 292; T. Sasaki, M. Suzuki, M. Kolář and M. sasaki, Approximate factorization of multivariate polynomials and absolute irreducibility testing, Japan J. Indust. Appl. Math., **8** (1991), 357; T. Sasaki, T. Saito and T. Hilano, Analysis of approximate factorization algorithm I, Japan J. Indust. Appl. Math., **9** (1992), 351; T. Sasaki and M. Sasaki, A unified method for multivariate polynomial factorizations, Japan J. Indust. Appl. Math., **10** (1993), 21; T. Sasaki and F. Kako, Solving multivariate algebraic equation by Hensel construction, preprint of Univ. Tsukuba (Jan., 1993); T. Kitamoto, Approximate eigenvalues, eigenvectors and inverse of a matrix with polynomial entries, Japan J. Indus. Appl. Math., **11** (1994), 73; T. Sasaki, T. Kitamoto and F. Kako, Error analysis of power series roots of multivariate algebraic equation, preprint of Univ. Tsukuba (Feb., 1994); T. Sasaki, A study of approximate polynomials, I – representation and arithmetic –, Japan J. Indust. Appl. Math., **12** (1995), 137; K. Shiihara and T. Sasaki, Analytic continuation and Riemann surface determination of algebraic functions by computer, Japan J. Indus. Appl. Math. (to appear); T. Sasaki, A study of approximate polynomials, II – properties of approximate factors –, preprint of Univ. Tsukuba (Jan., 1995), etc.

Is the mathematics we teach the same as the mathematics we do?[1]

W. J. Davis, Horacio Porta, and J. J. Uhl

The Ohio State University and
University of Illinois at Urbana-Champaign

Abstract:

Writing in the Bulletin of the American Mathematical Society, William Thurston said: "Mathematicians have developed habits of communication that are often dysfunctional." Our paper describes how the new course Calculus&Mathematica sets up an new method of communication of mathematics in the form of interactive texts running on computers. The lecture hall is replaced by the computer lab. The lecturer is replaced by students actively learning for themselves. Classroom meetings are genuine discussions with the whole class participating. Students do mathematics in much the same way that active research mathematicians do mathematics.

1. Issues and questions

In a recent article in the Bulletin of the American Mathematical Society, William Thurston called attention to the sad state of the mathematics classroom:

"We go through the motions of saying for the record what the students 'ought' to learn while students grapple with the more fundamental issues of learning our language and guessing at our mental models. Books compensate by giving samples of how to solve every type of homework problem. Professors compensate by giving homework and tests that are much easier than the material 'covered' in the course, and then grading the homework and tests on a scale that requires little understanding. We assume the problem is with students rather than communication: that the students either don't have what it takes, or else just don't care. Outsiders are amazed at this phenomenon, but within the mathematical community, we dismiss it with shrugs."

This brings up the question: Does what is taught in the typical mathematics course even qualify as mathematics?

Also in the Bulletin of the American Mathematical Society, Saunders Mac Lane offered:

intuition-trial-error-speculation-conjecture-proof

as a sequence for understanding of mathematics.

In contrast, the sequence in place in most modern mathematics courses is:

---

[1] A preliminary version of this paper was presented at the UNESCO-NSF Symposium on Science Education in the 21st Century at the American University in Cairo, April,1995.

lecture - memorization - test.

Most working mathematicians agree with MacLane's description, thus leaving the inescapable conclusion that the mathematics we do is not the same as what is commonly offered in the classroom. The questions for the next century are:

-> Can the mathematics we offer in the classsroom be more like the mathematics we do?

-> Can we ignite students' mathematical interests?

-> What role do computers have in dealing with these questions?

Writing in the American Mathematical Monthly, Phillip J. Davis indicates how the answer to the computer question sets up answers to the others:

"The capabilities of all mathematicians are elevated by their association with computation. The transformation by the computer of triangle geometry and of many other areas has, paradoxically, reconfirmed and strengthened the the vital role of humans in the wonderful activity known as mathematics. Put it even more strongly: mathematics develops in such a way that the role of the mathematician is always manifest. . .

In connection with visual output, I have even argued for the recognition of "visual theorems" . . . where what the eye "sees" need not even be verbalized let alone formalized in traditional formal mathematical language. . . subtle feeling that that language cannot even name, let alone set forth. . .

As regards mathematical education, I think the message is clear. Classical proof must move over and share the educational stage and time with other means of arriving at mathematical evidence and knowledge. Mathematical textbooks must modify the often deadening rigidity of the Euclidean model of exposition:"

2. Calculus&Mathematica as a prototype reaction to the issues and questions

Calculus&Mathematica is a new computer laboratory calculus course developed at University of Illinois at Urbana-Champaign and the Ohio State University expressly to deal with the questions and the issues raised above. The course is freshly built from the ground up. The purpose of the course, the ways of implanting mathematical ideas into students minds, the ways of motivating students to learn, and the ways of making students retain the important ideas have all been rethought.

As a result, Calculus&Mathematica is the most thoroughly new calculus course available today, and it presents a new model for successful learning of calculationally heavy sciences. Not screened from the essence of calculus by labor-intensive calculations and plots, students in Calculus&Mathematica get right to the good stuff. From the very beginning, they see calculus emerge as the first course in scientific measurement, calculation, and modeling. Students also see calculus as a highly visual and often experimental scientific endeavor just as research mathematics is. The medium is an live electronic interactive text composed of lessons written in Mathematica Notebooks. Each interactive Calculus&Mathematica lesson consists of the following set of Mathematica Notebooks:

Basics Notebook, for the fundamental ideas,

Tutorials Notebook, for sample uses of the basic ideas,

Give It a Try Notebook, for actual student work, and a

Literacy Sheet: for what a student should be able to handle away from the machine.

The National Research Council report *Moving Beyond Myths* describes Calculus&Mathematica as follows:

"An innovative calculus course . . .[which] uses the full symbolic, numeric, graphic, and text capabilities of a powerful computer algebra system. Significantly, there is no textbook for this course-only a sequence of electronic notebooks.

Each notebook begins with basic problems introducing the new ideas, followed by tutorial problems on techniques and applications. Both problem sets have "electronically active" solutions to support student learning. The notebook closes with a section called "Give-it-a-try," where no solutions are given. Students use both the built-in word processor and the graphic and calculating software to build their own notebooks to solve these problems, which are submitted electronically for comments and grading.

Notebooks have the versatility to allow re-working of examples with different numbers and functions, to provide for the insertion of commentary to explain concepts, to incorporate graphs, and plots as desired by students, and to launch routines that extend the complexity of the problem. The instructional focus is on the computer laboratory and the electronic notebook, with less than one hour per week spent in the classroom. Students spend more time than in a traditional course and arrive at a better understanding, since they have the freedom to investigate, rethink, redo and adapt. Moreover, creating course notebooks strengthens students' sense of accomplishment."

Unlike point - and - click multimedia and print page turners, each example in Calculus&Mathematica can be modified as the student sees fit and rerun; so that each example in Calculus&Mathematica is as many active interactive examples as the student wants.

The whole premise behind Calculus&Mathematica is that students who have the opportunity to go about their work in a way similar to the way working research mathematicians go about their work have a good chance for success. Here are some of the principles on which Calculus&Mathematica is based:

Communicate new ideas visually and experimentally; get an idea across before putting language on.

Unique to Calculus&Mathematica is the attempt to get mathematical ideas into the students' minds visually before words are put on. To paraphrase Stephen Jay Gould: Scholars are trained to analyze words, but students are visual animals. Well-conceived visualizations are not frills, they are foci for modes of thought. The course is driven by well-chosen re-executable, interactive computer graphics and student-produced graphics inviting the students to experiment, to construct for themselves, to describe, and to explain what's happening in their own words.

Through interactive visualizations, Calculus&Mathematica tries to stick the basic calculus ideas into the students' unconscious minds before it transfers the ideas into English. For instance, students experiment with simultaneous plots of f[x] and f'[x] to acquire an understanding of the meaning of the derivative. Students experiment with plots of the exponential function and are imprinted with its awesome growth. Students who have never heard of convergence experiment with plots of functions and their Taylor series expansions, soon discover that the convergence is what advanced mathematicians call "uniform on certain compact intervals.'" And they invent the word "cohabitation" to describe what they see. Students experiment with running trajectories through vector fields and become comfortable with vector fields. They know that gradient fields drain at relative maximums. As a result of thier experience, most of them can tell you why solutions of Laplace's equation cannot have an interior maximum. Reason: The gradient field of a solution of Laplace's equation has no sinks.

Always give the students the opportunity for a creative response; give the students an active role in their own learning. Don't try to think for the students.

Calculus&Mathematica students take an active role in their own learning by selecting material from the electronically alive Basics and Tutorials to learn (and possibly rework) as they need it, at their own pace. If a point doesn't get through, then they are free to modify and rerun as they see fit. At all times, they have the opportunity to pursue their learning actively and creatively. This lone aspect of C&M puts C&M at a great distance from lecture-based calculus courses and the new passive point-and-click multimedia courses coming onto the market. In the final analysis,this aspect of C&M is totally natural because this is the way research scientists do their work.

Approach mathematics as a science, not as a language or as a liturgy.

Often mathematics is taught as a ritual or liturgy in which the professor functions as curator of the dogma and arbiter of truth. Sometimes mathematics is taught as a language, a language which, as Blaise Pascal pointed out, "must be fixed in [the student's] memory because it means nothing to [the student's] intelligence." All too rarely is mathematics taught as the science that it is. The Calculus&Mathematica course attempts to teach mathematics as a science in which the student is the active investigator. With wise use of the computer to help introduce the ideas through the eyes, Calculus&Mathematica replaces the usual sequence:

lecture - memorization - tests

with this variant of Mac Lane's sequence:

visualization - trial - error - speculation - explanation.

In this format, calculus becomes the same as the mathematical activity in which active mathematicians engage.

Ask students for explanations, not proofs.

The words "prove" and "show" are the most terrifying words inexperienced math students ever encounter. The word "explain" is not so terrifying because explanations are usually assumed to be not so formal as a proof. On the other hand, a good explanation usually contains the main ideas of a formal proof; so that concentrating on explanations instead of formal proofs does not degrade mathematical understanding. In fact, rigor and understanding are often separate: Rigor is in part of the brain, but understanding permeates the brain, the heart and the soul.

Rigor without understanding and understanding without rigor are both possible. In any case, the ability to recite a memorized proof of a theorem is not the same as understanding the theorem. The real goal is to understand. And that's what Phillip Davis is talking about when he says, "Classical proof must move over and share the educational stage and time with other means of arriving at mathematical evidence and knowledge."

Use a computer-based genuinely interactive text.

Conventional printed texts have a paralyzing effect on learning because they force the student into a passive, subservient role. Thomas S. Kuhn explains it best: "Science students accept theories on the authority of teacher and text, not because of evidence. What alternatives have they, or what competence?"

The Calculus&Mathematica electronically alive interactive text, in which every example is as many examples as the student needs, is an environment in which the student can accumulate as much evidence as the student requires. The result: The student actively learns, in part, on the basis of the student's own authority and not just on the authority of the teacher or of the text.

Eliminate introductory lectures.

In his Bulletin article, Thurston states: "Mathematicians have developed habits of communication that are often dysfunctional. . . .most of the audience at an average colloquium talk gets little of value from it." Just as mathematics colloquium talks are failures, introductory lectures in mathematics classes are failures. Reasons:

-> Introductory lectures are full of answers to questions that have not been asked.

-> By necessity, introductopry lectures are full of precise terms not yet understood by the students.

-> Introductory lectures provide the strong temptation for the teacher to try to do the thinking for the students.

-> Introductory lrctures tend to center the course on the lecturer instead of the students.

To paraphrase Schopenhauer: Attending introductory lectures is equivalent to thinking with someone else's head instead of with one's own. Instead of introductory lectures in Calculus&Mathematica, regular discussions are held, but not until the visual ideas have congealed in the students' minds as a result of their lab experience. These discussions emphasize answers to questions the students ask.

Motivate students to want to learn by serving up problems whose importance is recognized by the students.

"What's this stuff good for?" is a student question often heard from students in ordinary calculus courses but seldom heard from Calculus&Mathematica students. The reason is that the mix of student problems in C&M puts students in a position to try calculus out to see what calculus can do for them in terms of their own lives and in terms of their own planned professional futures in measurement, calculation and science. Students think carefully about how to apportion their efforts as part of their planned futures. Possessing an uncanny ability to recognize frivolous or artificial classroom problems, students usually tune out of ordinary calculus courses, but they rarely lose interest in Calculus&Mathematica.

Keep the language in the vernacular.

Students fail in writing about mathematics because their textbooks are written in language they cannot understand. As a result, they resort to rote memorization because much of what the read and hear means little to their intellects. Paul Halmos even went so far as to say the job of the mathematics teacher is to translate the textbook into the vernacular. It does not have to be this way. Calculus&Mathematica is written in the vernacular in words, phrases and sentences that the students can understand and adapt in their own writing.

Give the students a chance to organize their thoughts by explaining themselves in writing.

Calculus&Mathematica students visually absorb ideas uncorrupted by strange words, and they address the problem of communicating what they have learned only after they have a visual understanding of the idea under discussion. The first step is to visually determine what the truth is; the second step is to explain it. Students in ordinary calculus courses are deprived of the excitement of discovery and explanation. C&M students write a lot of mathematics and they are unexpectedly good at it. Two reasons for this:

-> The Mathematica Notebook front end gives the students a unified environment for graphics, calculations and write-ups.

-> The language used in Calculus&Mathematica is informal enough for the students to adapt it to their own writing.

Give the students the opportunity to learn the mathematics and the programming in context.

Ordinary attempts to bring applications into calculus tend to separate the mathematics from the applications. Similarly, ordinary attempts to bring technology into calculus tend to separate the mathematics from the technology. Calculus&Mathematica always puts the mathematics in the context of measurement and puts the programming in the context of mathematics. Most importantly, C&M exploits the technology in effort to introduce new ideas. As a result, the applications, the programming, and the mathematics all feed off each other. A C&M student put it best:

"I have started to notice aspect of one class carrying over to another. Similarities in fields I thought unrelated before. An interconnection between math and language and programming and everything just kind of fits together a little better now."

Give the students professional tools.

Students preparing for careers in a calculational science see computers or workstations running Mathematica as professional tools. Believing that the ability to use professional tools is part of their overall education, C&M students typically throw themselves using Mathematica-equipped computers. They understand, perhaps better than their teachers, what vistas these professional tools open up.

3. Does it work?

The study by Kyunmee Park and Kenneth Travers, which compares standard calculus and Calculus&Mathematica, states: "Generally the findings from an achievement test, concept maps, and interviews where all favorable to C&M students. The C&M group obtained a higher level of conceptual understanding than did the standard group without much loss of [hand] computational proficiency. . . .[Some believe] that a laboratory course in calculus is very time consuming, and that students can become overly dependent on Mathematica. But this research found that the C&M course allowed the students to spend less time on computations and better direct themselves to conceptual understanding. Accordingly there was an increase in the students' conceptual achievement without a serious decrease in computational achievement. . . .Furthermore, the C&M group's disposition toward mathemetics and the computer was far more positive than that of the standard group. . . . Generally, the C&M group seemed to more clearly understand the nature of the derivative and the integral than did the standard group. . . . A positive side effect of the [computer] lab was the rapport that was established among the students. When students gathered around the computer, worked together, and shared and developed ideas, a great deal of mathematics was learned. ...[Computer] capabilities helped students discover and test mathematical results in much the same way that a physics or chemistry student uses the laboratory to discover and test scientific laws. Those capablities provided the opportunities for the students to consider more open-ended questions and to encouter more realistic problems than often found in traditional calculus texts."

Calculus&Mathematica students seem to agree. Some of their words:

"I've been studying math for years now and doing pretty well at it. But I never knew what I was studying before you got me on this computer and I could see it."

"It always gave me great satisfaction to look at a finished product on the screen. One thing that really surprised me was the focus on communication. . . . [C&M] really encouraged me to explain my results using real-world English. . . ."

"I like the 'hands on" type learning . . . . Here if you don't get it right the first time, you sit here at the machine until you do get it right, therefore learning in the process."

"In this course, I was allowed the freedom to prove something to myself. . ."

"Most of all, [the course] is great because it applies EVERYTHING to real situations. I feel that traditional math courses that I have taken have fallen short in this area. . . . This course is showing me how math is everywhere in the real world. I can't wait to get out there and put it to the test that really counts!!"

"Once you realize how important the basic theories are, it is like an avalanche, you start concentrating on the basics more and the new formulas make so much sense you could've derived them yourself. I do recognize that the true meaning and significance of [Calculus&]Mathematica is making critical decisions - understanding which math algorithm is necessary next in solving the problem, yet always understanding the whys, hows, and relationships between formulas... "

"I love. . . the freedom it gives me with my time and my schedule. But it isn't just because I find it easy that I continue and will next semester choose these courses. It is because of the nature of how we learn the math."

"In no other class, . . . have I been provoked to such intense brainstorming and internal argument, nor have the results of that intellectual tempest been so rewarding. . . .[This] course has led me to resolve to continue my struggle with mathematics."

Is Calculus&Mathematica a complete answer to what ails calculus education today? Probably not. Calculus&*Mathematica* has not proved to be ideal for all students. Some students who don't see a clear need for calculus in their planned futures are unlikely to be willing to make the commitment necessary to learn in C&M. Students who believe mathematics is nothing more than hand rote procedures are wary about getting help from the computer. Students who believe that a teacher's job is to teach passive students are also unlikely to do well in C&M. There are large populations of these kinds of students out there, and they are probably better served by calculus courses other than C&M.

But students who enter calculus with high expectations and motivations resulting from their own professional plans in a calculational science are likely to blossom in C&M. This includes high percentages of engineering students and math students. It also includes motivated rural high school students in the C&M Distance Education Program at Illinois. Life science students at Illinois have done so well in C&M sections restricted to life science students that the School of Life Sciences at Illinois has financed C&M labs for all of their freshman students.

We have been personally overwhelmed by the way students have thrown themselves into Calculus&Mathematica. We hope that Calculus&Mathematica and better courses to follow will help to pave the way to a time at which mathematics becomes just as live for its students as it is for its practitioners.

The authors thank Paul Weichsel of the University of Illinois and John Ziebarth of the National Center for Supercomputer Applications for helpful comments.

References

Bill Davis, Horacio Porta, and Jerry Uhl, *Calculus&Mathematica*, Addison-Wesley, 1994.

Phillip J. Davis, "The Rise, Fall, and Possible Transfiguration of Triangle Geometry: A Mini-history" American Mathematical Monthly 102 (3), March 1995, pp. 204-214.

Thomas S. Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, 1970.

Saunders Mac Lane, "Responses to Theoretical Mathematics . . ." Bulletin (New Series) of the American Mathematical Society 30 (2), April 1994, pp.190 - 193.

Moving Beyond Myths, National Research Council Report, National Academy Press, 1991.

Kyungmee Park and Kenneth J. Travers, "A Comparative Study of a Standard and a Computer-Based College Freshman Calculus Course" (Preprint).

William P. Thurston, "On Proof and Progress in Mathematics" Bulletin (New Series) of the American Mathematical Society 30 (2), April 1994, pp.161 - 177.

Addendum: An annotated problem from Calculus&Mathematica

The following annotations of a student problem from a Give It a Try section of a Calculus&Mathematica lesson are attempts to communicate the flavor of the ideas behind the course.

Problem Title: The Heat seeker Part 1)

Problem statement:

When you go with the following function f[x,y], you can spot the maximizer at {3,5}.

At this point, the student activates the following live code:

```
In:
Clear[f,gradf,x,y]
f[x_,y_] = 4 - ((x - 3)^2 + (y - 5)^2);

gradf[x_,y_]  =  {D[f[x,y],x],D[f[x,y],y]};
N[Solve[gradf[x,y]  ==  {0,0},{x,y}]]
```

This is Mathematica's response:

```
Out:
{{x -> 3., y -> 5.}}
```

Look at this plot of gradf[x,y] at various points on the circle of radius 1 centered at the maximizer at {3,5}:

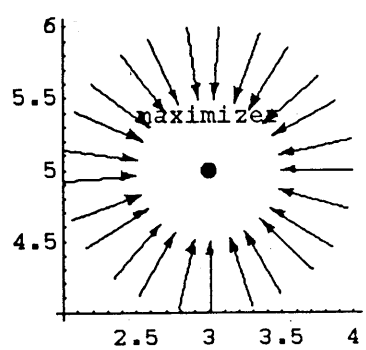
The student activates this live Mathematica code.

```
In:
maximizer  =  {3,5};
```

```
maxpoint  =
    {Graphics[{PointSize[0.05],Point[maximizer]}],
      Graphics[Text["maximizer",maximizer,{0,-4}]]};

radius  = 1;
Clear[x,y,t]
{x[t_],y[t_]}  =  maximizer  +  radius  {Cos[t],Sin[t]};

scalefactor  =  0.25;

gradients  =  Table[
        Arrow[scalefactor(gradf[x[t],y[t]]),
                 Tail->{x[t],y[t]},Red],{t,0,12,0.5}];

Show[maxpoint,gradients,Axes->True];
```

This is Mathematica's response:

Out:



Why do you think that those scaled gradient vectors are pointing the way they are?

(The student writes his or her response here on the live computer screen and turns it in as a computer file when the assignment is finished.)

Authors' comment on the pedagogy of this part:
This is to make sure that the students understand that the gradient field tries to drain at maximizers. The plot is sure to make a big visual impression and if the student is not able to explain why this happens, the student to digs into the Basics or Tutorials to learn why it

happens. Visual impressions wordlessly set up mathematical principles regularly in Calculus&Mathematica.

The learning process here is complete when the student explains in his or her own words why gradient vectors are pointing the way they are. One of the goals of mathematics is to explain why things come out the way they do. Calculus&Mathematica students do a lot of this.

Part 2

Problem statement:

The Calculus&Mathematica Missile Company is working on some primitive heat seeking devices and you are chief engineer of the TAD (Target Acquistion Division). The current problem under study is to program a device to go to the hottest point in a temperature distribution.
For instance, if

$$\text{temp}[x,y] = 100/(1 + (x - 2.5)^2 + 2\ (y - 3.5)^2)$$

measures the temperature at a point {x,y}, then the hottest point is {2.5,3.5} because the denominator is smallest at this point.

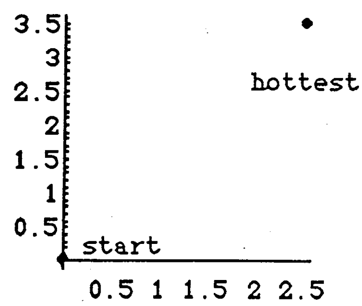
You can use the gradient to try to make a heat seeking device that starts at {0,0} and tries to seek out the hottest spot {2.5,3.5}.

Here's a look:

In:

```
Clear[temp,gradtemp,x,y]
temp[x_,y_] = 100/(1 + (x - 2.5)^2 + 2 (y - 3.5)^2);

gradtemp[x_,y_] = {D[temp[x,y],x],D[temp[x,y],y]};

hottestpoint = {2.5,3.5};

start = {0,0};

hotpt = Graphics[{{PointSize[0.04],Red,
          Point[hottestpoint]},
           Text["hottest",hottestpoint,{0,4}]}];

startpt = Graphics[{{PointSize[0.04],Blue,Point[start]},
           Text["start",start,{-1.5,-1}]}];

setup = Show[hotpt,startpt,PlotRange->All,
          Axes->True,AxesOrigin->{0,0},AspectRatio->1]
```

Out:



The heat seeker can't tell the exact location of the hot spot, but it can sense the gradient of the temp[x,y] merely by noting the hottest direction at a point {x,y}.
If the heat seeker is at a point {x,y}, why should you program the heat seeker so that it leaves {x,y} in the direction of gradtemp[x,y]?

Authors' comment on the pedagogy of this part:
One purpose of this part is to see whether the student can use the mathematical principle that the gradient points in the direction of greatest initial increase. Another purpose of this problem is to try to place the student in a professional setting to drive the idea that calculus is not just an idle classroom activity but is an essential part of their lives and of their planned futures as professional scientists. Sprinkled heavily through the course, problems that put students in real life or professional calculus situations are prime motivators for students to learn in Calculus&Mathematica.

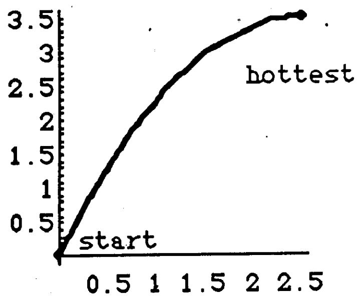
Part 3

Problem statement:

Given:
-> The heat seeker can update its direction every instant.
-> The heat seeker is programmed so that it leaves {x,y} in the direction of gradtemp[x,y].
Explain why the following plot displays a good approximation of the heat seeker's actual path when the seeker starts at {0,0}:

In:
```
Clear[x,y,t]
equationx  =  (x'[t]  ==  gradtemp[x[t],y[t]][[1]]);
equationy  =  (y'[t]  ==  gradtemp[x[t],y[t]][[2]]);
starterx  =  (x[0]  ==  0);
startery  =  (y[0]  ==  0);
endtime  =  10;

approxsolutions  =
         NDSolve[{equationx,equationy,starterx,startery},
           {x[t],y[t]},{t,0,endtime}];

Clear[seeker]
seeker[t_]  =  {x[t]/.approxsolutions[[1]],
                       y[t]/.approxsolutions[[1]]};

seekerplot  =
ParametricPlot[seeker[t],{t,0,endtime},
         PlotStyle->{{Thickness[0.02]}},
```

```
                                DisplayFunction->Identity];
Show[setup,   seekerplot,
        PlotRange->All,   DisplayFunction->$DisplayFunction];
```
Out:



Authors' comment on the pedagogy of this part:
When the student runs this plot on the machine, the student is left with the indelible impression that the theory that the student has announced in the last part actually works. This, in turn, builds enough student confidence to make the student write the explanation. Unlike students in ordinary calculus courses, experienced C&M students are known for their confidence and ability to write mathematics. Another purpose of problem is to introduce visually the idea of a trajectory in the gradient field and to relate it to systems of differential equations that the students have already studied.

Part 4
Problem statement:

Given:.
-> The heat seeker can update its direction every instant.
-> The heat seeker is programmed so that it leaves {x,y} in the direction of gradtemp[x,y].
Give a plot of the heat seeker's path when the seeker starts at {0, 2}.

Authors' comment on the pedagogy of this part:
After a lot of building of confidence and understanding, now it is the student's chance to show off what he or she has learned. This gives the student a feeling of success and accomplishment and whets their appetites for more.

Part 5
Problem statement:

The people over at the assembly division tell you that the heat seeker can't be built so as to update its direction at every instant. Instead, it will update its direction many times, but it will move on straight line segments between direction updates. Your group at TAD reacts to this information by programming the heat seeker as follows:
If the heat seeker is at

{x[k-1],y[k-1]},

then the heat seeker moves to a new point

{x[k],y[k]} = {x[k-1],y[k-1]} + jump gradtemp[x[k-1],y[k-1]]

where the jump is positive number selected by trial and error.
For appropriately small jump numbers why is this a good update?

Authors' comment on the pedagogy of this part:
This contributes to the realism of the problem. Students (and everybody else) know that things rarely go exactly the way they were planned. On the math side, students have not studied the method of steepest ascent at this point of the course. This is a way of sneaking in method of steepest ascent by introducing it in a real situation and putting the students in the position of explaining why it should work. It is also another reinforcement of the idea that the gradient points in the direction of greatest initial increase.

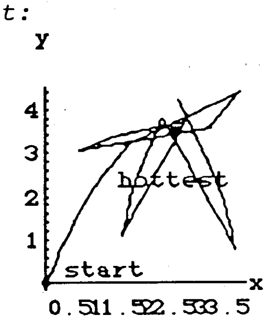
Part 6

Problem statement:

Now go from theory to practice.
Start at {0,0} and program the heat seeker with jump = 0.04 and 40 updates:

In:
```
jump  =  0.04;
updates  =  40;

Clear[next,point,k]
next[{x_,y_}]  =  {x,y}  +  jump  gradtemp[x,y];

point[0]  =  start;
point[k_]  :=  point[k]  =  N[next[point[k-1]]];

path  =  Graphics[
        {Thickness[0.01],
          Line[Table[point[k],{k,0,updates}]]}];


Show[hotpt,startpt,path,
        Axes->Automatic,AxesLabel->{"x","y"},
        AspectRatio->Automatic];
```

Out:



The heat seeker got close, but it blew its cool just as it was about to accomplish its mission. Use trial and error to program in a jump size and an update number that send the heat seeker steadily to the hot spot so it can ignite its warhead and blow that hot spot to smithereens.

Authors' comment on the pedagogy of this part:

This is a good example of using graphics that the student can interact with to further student learning. When the student succeeds in guiding the heat seeker to the target, the student gets a genuine feeling of excitement, confidence and satisfaction. Numerical issues seldom touched on in ordinary math courses are vivid here. Interactive problems of this type are nearly impossible in printed media or in point-and-click multimedia.

Part 7

Problem statement:

One way to increase the efficiency of the heat seeker is to use one of the larger jump sizes at first and run it until it goes bats. Then use the last good point generated as a new starting point with a new, reduced jump size and run again.
Try this out starting at {0,0} on the same function as above, incorporating any additional ideas that come to you.

Authors' comment on the pedagogy of this part:
Now the student is on his of her own and has the opportunity to be creative. Opportunities for the students to be creative are present throughout C&M and are quite rare in ordinary calculus courses. In addition, the computers give the students the chance to try out their ideas, learning in the process. In ordinary courses, students turn in answers and wait for them to be graded to find out whether they are right or wrong. In C&M, students use Mathematica to get instant feedback as they test their ideas.

Wrap up:

Each problem in Calculus&Mathematica is new and, like the problem above, each problem has been subjected to a thorough pedagogical and scientific shakedown through five years of field testing. The rigidly formal language, usually associated with ordinary mathematics courses, has been softened so that students can learn, talk and write in the same style in which the course is written.

Because Mathematica code is knitted into the electronic course from beginning to end, students learn Mathematica on an as-needed basis throughout the course. This eliminates the need for any prerequisite course in Mathematica programming. At the beginning, students simply copy, paste, and edit code and interact whenever they please. Later on, they write a lot of their own code. C&M students who have taken calculator-based courses often report that because of the interactive text, C&M is more user-friendly than the calculator.

GEOMETRY PROBLEM-SOLVING AND ITS CONTEMPORARY SIGNIFICANCE

Wu Wen-tsun
Institute of Systems Science, CAS
Beijing 100080, P.R.China
(wtwu@mmrc.iss.ac.cn)

Abstract.

1. Geometry Problems to be Solved.

The geometry theorem-proving in ancient Greece is well-known. As for in ancient China, one considered usually geometry problem-solving rather than geometry theorem-proving. Most of geometry problems to be solved arose from practical requirements. We may cite in particular:

(I) Problems arising from measurements, e.g. the measurement of sun height over the ground level or the measurement of hill height over the sea from the shadows of gnomens.

(II) Problems about area and volume determination. In particular a formula appeared in a classic due to Qin Jiu-shao of Song Dynasty (+1249) which is equivalent to Heron's formula about area of a triangle in terms of the three sides but in a different form.

(III) Problems involving Gou-Gu Form (i.e. a right-angled triangle), e.g. the determination of such forms with 3 sides in *rational* ratios.

(IV) Problems arising from actual lifes, e.g. the determination of the extensions of a city or an agricultural field in walking around in some way.

A lot of such problems with their solutions appeared already in the two earliest classics fortunately preserved up to the present day, viz. ≪*Nine Chapters of Arithmetic*≫ completed in 1c B.C., and ≪*Zhou Bi Mathematics Manuel*≫, believed to be no later than 100 B.C. These two classics will be cited below as ≪*Nine Chapters*≫ and ≪*Zhou Bi*≫ respectively. To these two classics we may add the all the more important classic ≪*Annotations to Nine Chapters*≫ due to Liu Hui in 263A.D., to be cited below simply as ≪*Annotations*≫.

The results of the geometry problems solved were usually expressed in the form of *Shui* which may sometimes be interpreted as *theorems* in the form of explicit formulas.

2. Methods of Proving Geometry Theorems in the Form of Formulas.

Instead of set of axioms the ancient Chinese formulated rather some simple and plausible principles and then proved the geometry theorems in the form of formulas or solved the geometry problems by logical reasoning.

The simplest one of such principles is perhaps the *Out-In Complementary Principle*: When a planar (or a solid) figure be cut into pieces, moved to somewhere else, and combined together, the area (or volume) of the figure will remain the same.

Diverse applications including geometry formula-proving and geometry problem-solving have been made by means of this seemingly trivial principle in an elegant and sometimes unbelievably striking manner. Most of formulas stated in the preceding section were proved in basing on this Principle. We may cite also the proof of the general Gou-Gu formula (i.e. the Pythagorous Theorem) based on the above Principle. The proof appeared in ≪*Zhou Bi*≫ and was attributed to some scholar Shang Gao in year around 1122 B.C.

3. Area and Volume Theory in Ancient China.

The Out-In Complementary Principle permits to establish a satisfactory theory of planar polygonal areas. However, it is far from being sufficient for the establishment of a volume theory of polyhedral figures. To overcome the difficulty Liu Hui in his ≪*Annotations*≫ introduced some Principle as follows. Let a rectangular parallelopiped be cut slantwise into two parts, and then cut one of the two parts again slantwise into two parts of which one is a pyramid with square base and the other a tetrahedron with 3 successive edges in mutually perpendicular directions. Liu showed rigorously by some limit considerations that in the case of a cube the volumes of these two parts are always in the ratio of 2:1. Let us call the Principle in the general case as *Liu Hui's 2:1 Principle*, then, together with the *Out-In Complementary Principle*, a satisfactory theory of volumes of polyhedral solids will be easily established.

For curved solid figures (also for curved planar figures) the above Principles are clearly insufficient to determine the volumes (or curved areas). However, Zhu Geng, son of the great scholar Zhu Cong-zhi (+429,+500), formulated precisely a Principle which furnishes a general method of determining volumes of curved solids in principle. In paricular, in following some suggestions of Liu Hui in his ≪*Annotations*≫, and in combining with the other two Principles, Zhu Geng determined the volume of a sphere in terms of its diameter. Remark that this Zhu-Geng Principle, or more exactly Liu-Zhu Principle, was already applied without precise formulation, to the determination of simple curved solids by Liu Hui. Remark also that this very Principle was later rediscovered in 17c under the name of Cavalieri Principle, which was applied to the determination of various kinds of curved solids, and became one of the cornerstones leading to the discovery of infinitesimal calculus in 17c.

4. Volume Theory in Western World.

A comparison with the area and volume theory as developed in Hilbert's classic ≪*Grundlagen der Geometrie*≫ in 1899 may be of some interest. Hilbert, based on the notion of *Zerlegungs-aequivalenz*, (alike to the Out-In Complementary Principle), had shown how the theory of polygonal areas may be founded. However, the method works no more in the case of polyhedral volumes and its solution forms the content of Problem 3 of Hilbert's well-known 23 mathematical problems in 1900. The problem was solved soon by Dehn in 1902, and was complemented by Sydler much later in 1965, in some intrincate manner. Besides, a lot of papers were published in Mathematische Annalen shortly after Hilbert's classic in 1899 which showed that the involvement of some sort of limit concept was inavoidable for the establishment of a satisfactory theory of polyhedral volumes. The problem of polyhedral volumes is thus a highly non-trivial one. In comparing the way of Liu Hui and that of Hilbert-Dehn-Sydler for the settlement of the problem, the author has the impression that the work of Liu Hui on volume theory should remain one of the most fundamental, the most important, and the most beautiful chef-d'oeuvres throughout the whole history of mathematics.

5. Creation of Heaven Element Method.

A turning point about geometry problem-solving occured in Song Dynasty (+960, +1274) and Yuan Dynasty (+1271, +1368). During that period the notions of *Heaven Element, Earth Element*, etc., (equivalent to the present day *unknowns* x, y, etc.), together with allied notions of polynomials and methods of elimination, were introduced. This furnishes a systematic way of solving geometry problems in the following way. Designate the unknowns of the geometry problems to be determined by Heaven Element, Earth Element, etc. The geometry conditions about the geometry entities involved will be turned into some algebraic relations between them which are usually equivalent to polynomial equations in modern form. By eliminating the unknown elements in succession we get finally equations each time in one alone of the unknown elements. When the known datas are in numerical values, what is usually the case, we may solve these equations in succession by means of methods of numerical solving developed in Song Dynasty. The results will give then the required solutions of the geometry problems in question. We remark in passing that polynomial equations-solving instead of geometry theorem-proving was the main theme of studies throughout the whole history of mathematics in ancient China.

The method created in that period was called the *Heaven Element Method.* It was in essence a method of algebrization of geometry which permits to reduce in relatively trivial manner geometry problems to the algebraic problems of solving of systems of polynomial equations. Elimination method, which appeared already in ≪*Nine Chapters*≫ for the solving of systems of *linear* equations (equivalent to present day Gaussian elimination) was also developed to the solving of general systems of polynomial equations in the period of Song-Yuan Dynasties. Numerous examples may be found in some still existant classics. Unfortunately the development of

Chinese ancient mathematics stopped at this crucial moment during the period of Yuan and Ming Dynasties.

6. Analytic Geometry of Descartes and Later Developments.

The algebrization of geometry, created in ancient China and stopped in 14th century, was revived in Europe in 17th century. It was even systemaized to the form of analytic geometry owing to the creation of R.Descartes. Descartes emphasized in his 1637 classic on geometry the geometry problem-solving rather than geometry theorem-proving. This was just in the same spirit as in ancient China. The analytic geometry also permits to prove geometry theorems by mere computations in contrast to the proving by purely geometrical reasoning of Euclid. The classic of Descartes showed also the way of reducing geometry problems-solving to polynomial equations-solving, which was just what the Chinese scholars in Song and Yuan Dynasties had tried to do.

In a posthumous work Descartes had founded some doctrine in saying that all problems can be reduced to the solving of problems in mathematics, then to problems in algebra, then to problems of polynomial equations-solving, and finally to problems of solving algebraic equations in single unknowns. In comparing with works of our ancestors, we see that our ancestors and Descartes were on the same lines of thought in the development of mathematics.

The computational method of solving geometry problems (including proving geometry theorems) in reducing them to the solving of polynomial equations lacked however a general way of dealing with the associated system of equations usually in embarassing confused form.

In recent years the Chinese mathematicians, mainly in MMRC, (Math. Mech. Res. Center, Institute of Systems Science, Chinese Academy of Sciences) have developed a general *mechanization method* for solving of arbitrary systems of polynomial equations which permits to be applied to geometry problem-solving, in particular geometry theorem-proving. Our achievements may be considered as a continuation of what our ancestors had founded as well as partial accomplishments of Descartes doctrine.

7. Diverse Applications.

Among the applications of our general method of geometry problem-solving, mainly done by members in MMRC and their collaborators in recent years, we may cite, leaving aside the geometry theorem-proving, the following ones for instances:

(a). Automated deduction of unknown relations.

Ex. Qin-Heron's formula for area of a triangle.

(b). Automated deduction of geometric loci equations.

Ex. Coupler point equation of a 4-bar linkage.

(c). Geometry Constructions.

(d). Inequalities and optimization problems.

Ex. Non-linear programming.

(e). Mechanism studies.

Ex. Design of 4-bar linkage in knowing 5 crank planar-positions.

(f). CAGD (= Computer Aided Geometric Design).

Ex. Surface-fitting, or more precisely smooth join by algebraic surfaces of given algebraic surfaces along given algebraic curves.

8. Contemporary Significance.

For geometry below we mean elementary geometry in the usual sense, though not necessarily so restricted.

A. Technological Significance.

The geometry-theorem proving in Euclidean fashion will furnish us source of inexhaustible high-rank amusements full of beauty and elegance. It is also of importance to learn for mathematical training. In contrast to this the ancient Chinese emphasized, just like Descartes, geometry problem-solving rather than geometry theorem-proving, which is in the main applications-oriented. This does not mean that Chinese ancient geometry lack beauty and elegance or unimportant for mathematical training, as seen clearly from the derivations of various formulas and also the volume theory of Liu Hui described in the previous sections. On the other hand the modern technology, being closely related to geometry, is a source of innumerable geometry problems to be solved. The ancient Chinese paved the way of dealing with such modern technological geometry problems in reducing them to the problems of polynomial equations-solving which have been tackled in particular by members of MMRC. That this way is quite hopeful may already be seen from examples in the last section.

B. Educational Significance.

Geometry occupies an important position in high school education. Owing to ignorance of current situation of present day geometry education, we shall make only some premature suggestions for possible consideration.

1. Geometry intuition connected with logical reasoning is one of the basic training and should not be neglected. However, it should not be too much restricted to lines and circles alone. Space intuition should be emphasized and strenghtened.

2. The original Euclidean system of axioms is full of flaws. The modified Hilbertian system of axioms, though rigorous, is too cumbersome to be of pedagogical value. In personal opinion it seems better to reject axioms at all in replacing it by simple and plausible Principles and then prove geometry theorems or solve geometry problems in basing on these Principles. This is somewhat similar to the learning of Newtonian mechanics in basing on 3 Newton's Laws, just like 3 Newton's Principles.

Ex. Triangle-Congruence Principle: Two triangles are congruent if and only if s.a.s., or a.s.a., or a.a.s., or s.s.s., but not a.s.s..

3. Geometry teaching should be applications-oriented. Thus, geometry problem-solving instead of geometry theorem-proving should be emphasized. Only few important and indispensable well-chosen theorems should be proved to serve as training examples by logical reasonings. Geometry constructions and loci determination should be paid more attention.

4. Algebra and geometry should be kept in pace in the teaching. For example, oriented segments, oriented areas, and oriented angles should be introduced to be dealt with in an algebraic way.

5. Analytic or coordinate geometry (planar as well as space) should be emphasized. Geometry theorems proved should be so chosen to be oriented toward the establishment of coordinate geometry.

Ex. Theorem. A parallel to the bases of a trapezoid will divide the two slant sides in equal ratios.

6. Elementary differential calculus with geometrical applications should be taught in high schools. Algorithmic-reasoning should be made familiar to students. Use of computer to mathematics and geometry should be encouraged, emphasized and strenghened.

SOME ASPECTS OF TECHNOMATHEMATICS IN EDUCATION

Rajendra Akerkar
Research Institute for Symbolic Computation
J. Kepler University
4040 Linz, Austria
e-mail: rakerkar@risc.uni-linz.ac.at

Rupali Wagh-Akerkar
Institute for Mathematics
J. Kepler University
4040 Linz, Austria.
e-mail: wagh@bruckner.stoch.uni-linz.ac.at

Abstract

Mathematics is without doubt a key to the key technologies of our time. The staggering advances in computer capabilities have radically changed the methods of industrial research. Many complex technical problems, which were earlier handled experimentally, are now solved numerically. Mathematical models are now taking place of real models in technological research. This accounts for the growing demand in industry for qualified employees who are able to *mathematicize* technical problems. We discuss WHAT and WHY technomathematics in university research and teaching. We also discuss how to build a bridge between universities and industries in asia, which will be a challenging task as well as a valuable goal for asian countries. Finally, we summarize the discussion with suggestions on HOW to get in.

1 Introduction

Mathematics is without doubt a key to the key technologies of our time. The staggering advances in computer capabilities have radically changed the methods of industrial research. Many complex technical problems, which were earlier handled experimentally, are now solved numerically because of the growing complexity of industrial processes and this requires advance knowledge in the form of mathematical modelling and computation. This shows for the growing demand in industry for qualified employees who are able to *mathematicize* technical problems. It is a job of the technomathematician to assure that the models are correct and that the methods used actually produce accurate solutions.

Nowadays, the usual procedure in an industry is based on the interaction and teamwork which is becomming more important as to assure quality and to retain competitiveness. People trained in quantitative methods of all areas of mathematical sciences are increasingly being requested to contribute as active and equal partners in technology and system developement.

At this time, a new kind of thinking in mathematical science education is emerging, especially in the industrialised countries. In this thinking, one would like to exploit the diversity of the mathematical science community to develop a variety of programs based on talent and opportunity.

The focus of this paper is on technomathematics in academic departments of asian universities. In the first section we dicuss What is technomathematics and Why it is in academic departments. second section describes how to build a bridge between universities and industries in asia, which will be a challenging task as well as a valuable goal for asian countries. In final section we summarize the dicussion with comments on how to get started.

2 Technomathematics

2.1 What it is

In the modern world, mathematical models are taking place of real models in technological research. Industry must continually improve the quality of products to remain competitive. Also, it must shorten the time it takes to develop a concept into a product. With the increase in complexity of the process for manufacturing the product as well as with lengthy and costly physical experiments, the demand for mathematical modelling and simulation is increases.

In short, the developement of mathematical model with algorithms for computer simulation to obtain solutions for industrial problems is what we call a *technomathematics.*

First of all, the framing of the problem must be appropriate with least theoretical assumptions so as to maintain the original problem. With necessity there must be some scope of flexibility to use the advance knowledge of industrial processes. This will benefit mathematical sciences in the sense for developing theories. It is better to have communicative partners with other relative areas in all stages of the work. The mathematical modelling, analysis, and solution must be meaningful to industry and in real life.

The scope of technomathematics is very broad and can be described by borrowing the following lines of Albert Einstein,

> Its realm is accordingly defined as
> that part of the sum total of our
> knowledge which is capable of being
> expressed in mathematical terms.

These words were originally used to define physics. Taken literally the statement clearly includes the mathematical theories of biology, communication, economics etc. and is perhaps a more adequate description of technomathematics.

Technomathematics itself is a collection of different fields of mathematical sciences used by industries in some or the other way. In manufacturing industries alone, nearly every area of mathematical sciences is being needed. There were many areas of applicable mathematics which contributes to industry. Some of them are *ordinary and partial differential equations, control theory, fuzzy logic, dynamical systems, visualization, robotics, discrete equations, Monte Carlo methods, stochastics models, Markov chains etc.*

The different areas of industrial manufacturing and management where the above quoted fields of mathematical sciences are basically used for *manufacturing processes, process control, statistical quality improvment, cost based performance measure, solid modeling, computer based information management, capital budgeting for flexibility etc.*

In industry, the object is to control and design the phenomena to maximize the output and to minimize the cost. There were problems which were formulated in a way that requires techniques of optimization and operations research. Currently, there is a great demand for devising parallel versions of sequential algorithms.

Sometimes technomathematics alone donot provide complete answers to industrial problems, which can be provided in collabration with the natural sciences or social sciences. The solution of industrial problems requires an interdisciplinary team approach.It is a major factor in determining what skills are needed in technomahtematics and what the focus of an academic technomathematic program should be.

2.2 Why it in the Education

In the last few years, most of the industrialized countries are practicing and training mathematicians for industrial jobs. It has shown great success and this success is based on applying available analytical and computational techniques in addition to discovering new techniques.

As we observe nowadays, the graduates of mathematics are unable to find jobs and due to this the mathematics departments in asia are reducing in size. To change this trade the asian mathematical science community should educate significant numbers of students in ways that give them strong qualification for better job market and thereby lessen the risk of unemployement.

Taking this into consideration, mathematics departments of asian universities should change their way of teaching and research so as to increase the market value of their students or researchers. Since the culture and values of mathematics in industry differ significantly from those in academia, we must train mathematics undergraduates and graduates in such a way that they can accept the challange to work with natural and social scientists, engineers, managers and bussiness people to develop complete and more accurate models for manufacturing processes, marketing distributions and product performance etc.

The goal of teamwork in an industry is to solve problems either by existing or new mathematics. It is not necessary to have the most perfect solution but it is sufficient to achieve some solution in real time. For this, technomathematician must have a sufficient background so as to recognize the model for better solution. offcourse, there is much work in developing new models and solution procedures or new variants of existing models and solution procedures for the infinite variety of problems. In this regard, we will suggest the following qualifications for technomathematicians so that they can have better chance for getting job in industries :

1. Sufficient knowledge in relative area of mathematical sciences.
2. Computational skills.
3. Better communication skills.
4. The ability to learn new concepts.
5. The ability to work with people from other disciplines.

With all the above qualifications, a technomathematician may be able to solve problem using different aspects of modelling, analysis, and computation. Modelling is the basic step in the technomathematics problem. In the next subsection, we will give an idea of modelling with some simple examples and before concluding this section we will give an algorithm for solving technomathematics problem.

2.3 Mathematical Modelling

Mathematical modelling has been successfully used by almost all scientists and engineers through-out the ages, but its importance as a discipline to be studied and cultivated has been realised only during last few decades. At this point, we recall our definition of technomathematics, which is the development of mathematical model. Mathematical modelling essentially consists of translating real world problems into mathematical problems, solving the mathematical problems and interpreting these solutions in the language of the real world. As an easy examples of mathematical modelling, we can consider five problems :

1. Find the mass of the Earth (without using a balance!)
2. Find the effect on the economy of 25 per cent reduction in income-tax (without actually reducing the rate!)
3. Estimate the population of India in the year 2010 A.D. (without waiting till then!)
4. Estimate the average life span of a bulb manufactured in a bulb industry (without lighting each bulb till it gets fused!)
5. Estimate the total amount of insurance claims a company has to pay next year (without waiting till the end of that year).

One idea of solving these probles is same as that of solving *word problem* in algebra. Suppose the age of the father is four times the age of the son and we are told that after five years, the age of the father will be only three times the age of the son. We have to find their ages. Let x be the age of the father and y be the age of the son, then the data of the problem gives

$$x = 4y,\ x + 5 = 3(y + 5),$$

giving $x = 40$ and $y = 10$. here above two equations give a mathematical model of the biological situation, so that the biological problem of ages is reduced to mathematical problem of solution of a system of two algebraic equations. The solution of the equations is finally interpreted biologically to give the ages of the father and the son.

The mathematical relations we get during such development may be in terms of algebraic, trancendental, differential, difference, integral or even in terms of inequalities. Thus for above-mentioned problems

1. we try to express the mass of the Earth in terms of some known masses and distances.
2. we examine the effects of similar cuts in the past or develop a mathematical model giving relation between income-tax cuts, purchasing power in the hands of individuals and its effects on productivity and inflation etc.
3. we extrapolate from data from previous censuses or develop a model expressing the population as a function of time.
4. we take a random sample of bulbs, find their life-span and use statistical inference models to estimate the life span for the population of bulbs.
5. we use probabilistic models for life expectancy of individuals.

It is much easier to solve the mathematical equations provided we know how to formulate them and to solve them.

Better analysis contributes to better understanding of the problem. As experiments become more expensive and time consuming the pressure for replacing more experiments by mathematical modeling and computation grows.

Here we give a simple algorithm for solving technomathematics problem :

Algorithm

Input : Technomathematics problem.

Output : Computational result.

1. Identify the problem and **goto** step 2.
2. **Do** mathematical modeling of the problem and **goto** step 3.
3. **Do** mathematical analysis of the model and **goto** step 4.
4. Develop computational method and computer codes and **goto** step 5.
5. Compare the computational results with observations and **goto** step 6.
6. **If** satisfied **then** stop **else goto** step 1.

End.

Finally, implement and integrate the solution into process.

In the section 2.2, we have suggested the qualifications required for technomathematicians to have better chances for job in industry. From these, it is expected that the student will develop, through his personal experience in creative activity, a wisdom which enables him to act as a technomathematician. In the latter capacity, he must be able to present a perspective of the total activity so that the younger generation will be able to decide how and where to devote their efforts, and to carry on their work with confidence. From the above discussion, one can easily know that the need for technomathematicians is expected to grow.

3 How to Build Bridge Between University and Industry

The best way of having relation with an industry is to have direct intellectual contacts with respective person from industry. For this, there must be successive meetings with industrial people and the best way is to have an active scientific association with the involvement of industries. In first few meetings the aim of association should be clearly stated. Then, interested people must visit the industry to see their way of working or handling problems. After this, if they are able to relate the mathematical issue then they must present some seminars and try to convince people from industry. It is not essential to give complete answers to their problem but partial results will make a nice impact. Timely approximations are much better than precise solutions that are late. After some positive interations, there is a scope of having some compact courses for students and research worker in industry to work on problems discussed in seminars. With this approach, there is a chance to have better contacts with industry. It is not necessary that every industry came in contact will prefer interations but this is not the problem, may be some other industry will feel to take a chance.

In setting up a relation between university and industry, we should first find out the need, preferably in local industry, for technomathematics programs on various levels and with the stress on various areas. If a need for a certain area is established and the department has advisor or teacher capable of managing a program in that area, then without losing time a program could be set up in that area.

In Asia, there had not been a tradition of technomathematician. People now in this profession come from a variety of backgrounds. Thus there is a definite need and for this practicing technomathematicians must do something important for asian community. This is to educate others in our subject (special line of research) in a well-motivated manner. This would create mutual understanding, establish mutual confidence, and build up the community spirit.

In the concluding section, we present a discussion of the education program with some suggestions.

4 How to get in

Now question is, how can we arrange an educational program in technomathematics to achieve the goals described above?

First of all, the education must be started in the master's level, during the formative period of the youth. University departments willing to create new mathematics programs should consider technomathematics programs as one option among the many options that are open to them. And, to guarantee that technomathematics will be an important option, one or more faculty should start establishing relationships with industry.

We may thus suggest a basic educational program in technomathematics as follows

1. an education in the attitude of technomathematician,
2. an education in the usual (methods of technomathematics) working ability,
3. a survey in technomathematics with the stress on the real world or industrial problems,
4. a basic education in pure and applicable mathematics,
5. an education in atleast one branch of science in depth.

The courses (1), (2), and (3) should be taught by technomathematician or capable faculty member who can manage a program in related area of technomathematics;

course (4) by pure and applied mathematician; and course (5) by professional scientist in various universities. The post-master's education should be an extension and a continuation of the basic program suggested above. It must include the development of the ability of the student to do research.

There is a need to have highly qualified technomathematicians specifically devoted to such activities. We wish to make following remarks towards accomplishing this goal.

Remark 1 : Technomathematics programs should be offered in the universities, at both the master's and research levels, aiming at the intellectual caliber.

Remark 2 : The university administration and the scientific community should be made aware of the need of technomathematicians.

Remark 3 : The universities should create faculty positions specifically in technomathematics and appoint to these positions only people of the high caliber.

Remark 4 : Government should expand its support to academic activities in technomathematics.

While excitement is often created by new developments in any scientific subject, traditional subject matter, sifted through many generations of critical thinking and brought up to its modern form, can often serve as a core of knowledge on which future developments can be based. It is indeed these thoughts, ideas and reasoning, which must be taught in the master's level as a source of stimulation and as a basis for future research efforts. The solid foundation will always help us to reach new horizons if we adopt the progressive point of view.

References

[A] R.D. Anderson, *Employment of New Ph.D.'s*, Notices of the American Mathematical Society, 18, 1971.

[BB] D.N. Burghes and M.S. Borris, *Modelling with Differential Equations*, Ellis Horwood, 1981.

[B] S. Benkoski, *Preparing for a Job Outside Academia*, Notices of the American Mathematical Society, 1994.

[D] P. Davis, *Some Glimpses of Mathematics in Industry*, Notices of the American Mathematical Society, 40, 1993.

[G] H.P. Greenspan, *Applied Mathematics as a Science*, Amer. Math. Monthly, 68, 1961.

SOME THOUGHTS ABOUT THE USE OF COMPUTER ALGEBRA SYSTEMS IN UNIVERSITY TEACHING

Helmer Aslaksen
Department of Mathematics
National University of Singapore
Singapore 0511
Republic of Singapore
aslaksen@math.nus.sg

1 Introduction

At the moment, the use of computers in mathematical research and teaching is getting a lot of attention. Many mathematicians seem to have caught a "computer virus". Computers can be as addictive as alcohol to mathematicians, and a lot of us seem to think that computers are a panacea for both research and teaching. The bureaucrats prefer spending money on buying equipment rather than hiring people, and the vendors of the computer algebra systems bombard us with evangelism. The result is that a lot of people, especially outsiders, seem to envision that mathematicians will soon be made redundant by computers.

But computer algebra systems have their fair share of problems, too. First of all, there are intrinsic limitations to what these programs can do. For a discussion of such problems see [A] and [S]. In this paper I will instead focus on some of the problems related to implementing them in teaching mathematics at the university level.

At the Department of Mathematics at the National University of Singapore, we are currently in the process of introducing computer algebra systems (CAS) into our pure mathematics courses. There are many reasons for doing this.

1. **Exposure.** It is important for the students to be aware of and familiar with the tools that are available.

2. **Graphics.** The graphical components of CAS can be wonderful tools for visualizing complicated concepts.

3. **Numerics.** The numerical component of CAS allows for interesting combinations of exact and numerical methods.

4. **Applications.** Using CAS enables us to discuss more realistic examples, rather than the often artificial applications that abound in calculus textbooks.

5. **Concepts.** Using CAS allows us to focus on the concepts rather than the computations.

6. **Labs.** Using CAS, one can radically change the way mathematics is thought, by introducing a more experimental, lab-based approach.

One aspect that I feel is significant about CAS is that it helps bring pure and applied mathematics closer together. It becomes natural to use more numerical methods in calculus for instance, and one can introduce non-trivial applications. But that doesn't mean that applied mathematics will take over the whole field. In a different paper ([A]), I considered some problems related to elementary complex functions, and showed that it is somewhat of an Achilles' heel for CAS. It is interesting to see that what is needed is "esoteric" pure mathematics like branch cuts, demonstrating once again the centrality of pure mathematics.

But it is important to realize that introducing CAS into the classroom raises a lot of problems and questions.

1. **Cost.** These programs are quite expensive, and they put serious demands on the hardware, especially if one wants to take advantage of the graphical capabilities.

2. **Computer-illiterate staff.** There are still a lot of computer-illiterate staff in most mathematics department. Senior staff may feel threatened by plans to introduce technology and teaching methods that they are not familiar with.

3. **Demand on the students.** A lot of students are having a hard time with basic calculus. Do we really want to also teach them how to use a complex CAS?

4. **Time consuming.** If we want the students to use CAS, we will need to set aside time to help them with the programs. We will also need to write suitable teaching material and to maintain the software and hardware.

5. Which program to use? What if some people in the department want to use Maple, and some want Mathematica?

6. **Basic skills.** The students need certain basic skills in calculus, and some people are concerned that they will not master it if they have access to CAS.

7. **Exam.** Will the students be tested on the use of the CAS? Will they be allowed to use it during the exam?

In the US, many departments have introduced calculus labs with quite good results. But these programs are usually conducted at big schools where there are many graduate students who can supervise students during lab hours, and with adequate technical staff. If the professor has to install and maintain software and hardware and teach computer-phobic students how to turn on the computer, that might quickly become a huge burden.

I have come to the conclusion that a suitable compromise is to not go for the lab approach, but just use CAS as a demonstration tool during lectures. I have experimented with both bringing a notebook computer to class and printing the output on transparencies. The CAS was available on the campus network, and the students had access to my input files. In this way the students get some exposure, without having to worry about the idiosyncrasies of CAS input syntax.

This approach also allows different staff to use different programs. It is confusing for the students if they have to different programs in different classes, but if they are not required to learn the syntax, then that problem disappears.

I'm sure that experts in mathematical education will object to this approach. It is in fact a fairly conservative approach, that is quite compatible with traditional teaching. But it is a convenient way of getting started, and based on some limited experience, I feel it can be a good solution.

References

[A] Helmer Aslaksen, Multiple-valued complex functions and computer algebra, preprint no. 631, Dept. of Math., National Univ. of Singapore, 1994.

[S] David R. Stoutemyer, Crimes and misdemeanors in the computer algebra trade, *Notices Amer. Math. Soc.* **38** (1991), 778–785.

INCORPORATING COMPUTER ALGEBRA INTO COLLEGE LEVEL MATHEMATICS

Chris Barling
Swinburne University of Technology
John Street, Hawthorn 3122, Australia
crb@stan.xx.swin.edu.au

At the Mooroolbark campus of Swinburne University of Technology we have a small group of students studying Mathematics and Computer Science. They take a program of three or four years post-secondary study, commencing at about the equivalent of U.S. College level. Their major mathematical studies are in Applied Statistics and Operations Research, with sufficient mathematical techniques to get by in these areas.

For the last two years these students have been studying in what we call multi-modal format, which is intended to embrace a variety of innovations: instruction both on and off campus, at home and in local study centres, electronic networking, and the full use of best available technology. One particular consequence is that all the students in this group have full use of their own laptop computers at all times.

This has created a very special opportunity in the teaching of their basic mathematics subjects. For the last twenty years or so we have expected students to have their own electronic calculators. Recently we have moved seriously into the use of graphical calculators, which has led to a number of modifications in our syllabus and methods of teaching (see [Barling 1991 and 1993, and other references therein]).

It seemed obvious to me that it would be rather bizarre to have students with their own laptop computers also required to buy another piece of much simpler equipment, and thus we reached the decision that instead of using calculators they would work with a suitable package on their computers. The package required would at least satisfy their calculating needs and match the graphing capabilities of a graphical calculator, but presumably go much further as well. This opened an exciting range of possibilities! After due consideration of cost and educational factors we settled on the computer algebra package Maple as the one to adopt, at least on a trial basis.

In our conventional course (which has continued at our Hawthorn campus) students study basic one-variable calculus in the first semester of first year and then extend in second semester into linear algebra, multivariable calculus and so on. Our intention was that the students would become self-sufficient in the use of Maple: rather than working through prepared exercises they would learn the basic syntax and operation of the package and thus be able to generate their own work freely, in all areas of the syllabus where it was relevant (which is most). This necessitated an initial period of getting used to the rather strict syntax used in the package, and learning the basic structures.

For the first group, in 1993, because of initial difficulties in setting up the program, the students studied in the conventional way for one semester (with borrowed calculators) and then switched to Maple mid-year. This was not very satisfactory for a number of reasons. In the first semester there was a lack of interest in learning the finer points of the use of the calculators, as students knew their use would be temporary. Then, when we were finally able to switch to Maple in the second semester, there was an enormous rush to get through the quite difficult initial stage before the students could continue with their new work. Some topics studied in semester one would have profited enormously from the use of Maple, but by the time the package was available to them the students had moved on and on the whole had lost interest in relearning old material. There were also difficulties caused by running two quite different programs on two campuses. Although we tried (for administrative reasons) to keep the two campuses running in parallel, studying the same topics each week, the pressure of learning Maple as well meant that we quickly got out of step. Overall, the students were highly enthusiastic about using Maple but invariably said that they would have preferred to have met it much earlier, if possible at the start of the year.

In 1994 with the second group of multi-modal students we followed that advice and introduced Maple at the beginning of the course. Once again there was an initial rush to get through the first stages of using the package at the same time as learning new mathematics. This time we managed to keep the two campuses together but with some difficulty. Once again the students were quite enthusiastic about using the package, but were worried by the initial rush: many suggested introducing it some weeks into the semester (but had no suggestions as to how we should operate in the meantime!).

For the third intake of students in 1995 we have tried another approach, by postponing some of the initial content until later in the year and concentrating largely on Maple for the first two weeks. From then onwards we have kept the two campuses together, and on the whole this seems to have been the best solution. The new students seem to be comfortable with their work and are using the package competently.

I should now outline what parts of our curriculum are particularly affected by the use of this package. Naturally, being designed as a computer algebra package, it is strongest when dealing with algebra, which at the level of our students means the traditional (real) algebra of polynomials, rational functions and the like: evaluating, simplifying and so on. It also deals naturally with the routine processes of differentiation and integration, which at this level operate only as algebraic processes following simple rules. Maple is designed to work exactly with integers, rational numbers and (less elegantly) surds, to factorize and solve polynomials, and to do more interesting things that can be reduced to mechanical processes. It is brilliant at traditional algebra, but less adequate with transcendental functions. It prefers to work exactly, but will work numerically when asked, although its numerical procedures need careful use to maintain accuracy.

For beginners, after an initial introduction, I have found it best to start at the very beginning and look carefully at how Maple does calculations. All of our students are familiar with conventional electronic calculators, and we begin by seeing how Maple deals with the sorts of problem for which they would normally turn to a calculator. Some of the differences are surprising to them. It is not easy to explain even to beginning tertiary students why Maple does some of the things it does with surds, powers etc. We then run through the standard functions: trigonometric functions and their inverses, exponential functions and logarithms, looking at both calculations and simplification processes.

Up to this point we have not done much more than using Maple as an elaborate and difficult scientific calculator. But once we move into classical algebra and begin to work with polynomials and the like, Maple comes into its own. It can expand an expression, or factorize it. it can substitute one expression into another. It can apply that multipurpose word "simplify" in a multitude of situations, and usually does what you would expect.

Maple can solve algebraic equations with great skill, and will find numerical roots of equations when asked. It can solve sets of simultaneous equations, and inequalities. It can plot and manipulate graphs (a little more clumsily than a typical graphical calculator) for one or several functions, and for many relations.

In calculus itself Maple does all of the routine calculations: it will find limits and derivatives, and any definite or indefinite integral that secondary or first-year tertiary students are likely to come across. This is the point at which you inevitably have to question the curriculum intensely: if machines can differentiate and integrate with such facility, why do we spend so much time on the arcana of it? Here, I am sure, is where the next battle of the war of technology against traditional methods will be waged.

Maple will calculate an approximating polynomial and sum an infinite series — again, of the type that we traditionally expect students at this level to be able to handle. It can solve differential or difference equations and work easily with matrices. It works happily with real or complex numbers. It can produce three-dimensional graphs. All of these clearly open enormous possibilities for the competent student.

In the last five years we have gone through a process of rethinking all of our curriculum in the light of the ready availability of graphical calculators (see [Barling, 1993]). How you use a piece of technology is largely determined by the degree to which teachers and students can have access to it. Now we have repeated the exercise with the even more advanced capabilities of the computer algebra package. We have been in the unique position of working with a group of students who have full access to the package at all times, both on campus and at home, and we have had to redesign their program accordingly. The students use their computers as freely as most students use their electronic calculators, so we are naturally aiming to maximize the effective use of the package.

This means rethinking the whole curriculum as to what topics are appropriate and how they are to be taught. Some notable effects are described in the following paragraphs. Most teachers and students will have rather less access to any computer package, and their opportunities will be restricted accordingly.

With Maple or a similar package freely available, examples can be drawn from almost as wide a field as you can imagine, without having to worry about the difficulties of the manipulation, at least in algebra and basic calculus. Once the basic commands are mastered, computer algebra lets you roam almost where you will.

Many traditional topics can be omitted or just described, with the "hack" work being done by the machine. Certain topics in systematic integration spring to mind.

You quickly become aware of the limitations of what can and cannot be done exactly. Maple users, for instance, are far more aware of the unsolvability of general polynomials of degree five or more than most of their contemporaries.

As a general rule, it is more important than ever that students gain a basic understanding of what the topic is about as well as being able to do the manipulations expected of them. In some topics I have found it wisest to teach almost all of the traditional material before bringing in Maple to illustrate it. On other occasions Maple is useful as part of the teaching process right from the start. Judicious use is advised.

If you are planning to integrate a piece of technology into your students' normal ways of doing things, you cannot treat it as a last-minute add-on. Notes and other course materials have to incorporate the relevant ideas and instructions at the logical place, not in an Appendix. All of our materials now exist in two forms, one for the Maple users and one for the conventional students.

When we first introduced graphical calculators a significant number of students had trouble integrating the graphical information they were obtaining with more conventional algebraic or symbolic information (see [Boers and Jones, 1992 & 1994]). The students using Maple do not seem to be having this sort of difficulty, perhaps because they are not often tempted to try problems in alternative ways. On the other hand their abilities do seem to have been significantly enhanced: compared to their predecessors and contemporaries not using Maple we have established a clear improvement in performance both in manipulative tasks (where Maple obviously helps) and in questions of comprehension that involve no technology. Using Maple seems to have helped these students to learn their mathematics better.

We all need constantly to be re-thinking our curriculum — both the content and the way we teach and assess it — and keeping a clear eye on what are our real aims for our students. In vocational education the aim is not academic perfection but a functional, competent professional. Beyond their basic skills and vocabulary, what our students need is confidence, the ability to learn a new topic, to solve problems, to know what to do in an unfamiliar situation. Giving them maximum access to professional-standard technology surely has an important role in achieving this.

References:

Barling, C. R. (1991). Calculus with a graphical calculator. In J. O'Reilly & S. Wettenhall (Eds), *Mathematics: Inclusive Dynamic Exciting Active Stimulating* (pp 380-384). Melbourne: Mathematical Association of Victoria.

Barling, C. R. (1993). Graphical Calculators: Potential vs Practice. In J. Mousley & M. Rice (Eds), *Mathematics: Of Primary Importance* (pp 57–60). Melbourne: Mathematical Association of Victoria.

Boers, M.A.M. & Jones, P.L. (1992). Is calculus made easier with the graphics calculator? In M. Horne & M. Supple (Eds), *Mathematics: Meeting the Challenge* (pp 392-398). Melbourne: Mathematical Association of Victoria.

Boers, M.A.M. & Jones, P.L. (1994). Students' use of graphics calculators under examination conditions. Int. J. Math. Educ. Sci. Technol., 1994, vol. 25, no 4, 491–516.

The first three of the above papers have been reprinted in:
T. Andrews & B. Kissane (Eds), *Graphics Calculators in the Classroom* (pp 59–63, 123–126 and 65–71). Adelaide: Australian Association of Mathematics Teachers (1994).

TI82 GRAPHING CALCULATOR, MATRICES, AND TABLES

Carole A. Bauer
Triton College
Home: 5724 Sunnyside
Chicago, IL 60630 USA
CBauer@eworld.com

I. History of Graphing Calculator Use at Triton College

Graphing calculators were first required in a few experimental sections of College Algebra in the Spring Semester 1992. Two professors (Carole A. Bauer and Jon D. Weerts) taught the experimental sections (TI81 graphing calculators required) as well as traditional (non-graphing calculator) sections. The results of a study of student learning and student satisfaction encouraged the professors to increase the number of sections (Fall 1992) of College Algebra that required the students to purchase graphing calculators. This semester also produced positive student satisfaction with using a graphing calculator in mathematics.

In the Spring Semester 1993, other full-time faculty joined the move into the use of technology in College Algebra. In addition, C. A. Bauer and J. D. Weerts expanded the use of technology by requiring students in a combined College Algebra-Trigonometry course and in Trigonometry to purchase the TI81 graphing calculator. By the Fall Semester 1993, most full-time faculty of these pre-calculus courses were requiring TI81 graphing calculators. During this same semester all part-time faculty of these courses were requiring TI81 graphing calculators. In this semester it was noticed that a number of students commented that they had used graphing technology in their secondary school mathematics classes.

The Mathematics Department voted to require graphing calculators in all sections of the precalculus level courses starting in Spring 1994. Although some use of technology had been present in other courses such as calculus, finite mathematics, and differential equations; during this semester C. A. Bauer taught Calculus I and required students to purchase a graphing calculator if they did not currently own one. In this same semester, J.D.Weerts taught some experimental sections of a developmental course (Intermediate Algebra) with students required to purchase the TI82 graphing calculator. Students in all courses were advised to purchase the TI82 as the calculator of choice, however the TI81 and the TI85 were also acceptable choices. The ease of using the Texas Instrument graphing calculators and the availability of overhead units for instruction make these calculators the choice at Triton College.

During the Fall Semester 1994, the majority of students are opting to purchase the TI82 or the TI81 for class. Almost all sections of the courses from pre-calculus through differential equations use technology in the form of graphing calculators. Some of the developmental courses are also using technology.

II. Capabilities of Graphing Calculators

Students who use graphing calculators have the power to investigate problems that would be difficult to manage with only paper and pencil. With each new version of the calculator, additional features appear and expanded memory is apparent. The TI81 can store three matrices of maximum size 6 by 6. The TI82 can store five matrices of maximum size up to 99 by 99, depending on available memory. The TI82 can be linked calculator to calculator or calculator to computer to facilitate the transfer of data or computer programs.

A solid background in mathematics is still necessary to interpret the results determined with a graphing calculator. For example, a square matrix does not have an inverse if its related determinant is zero. Using the TI81, it is possible for such a singular matrix to appear to have an inverse. The technology in the TI82 has improved so that apparently this does not happen.

The TI82 has a built-in table generator that allows students to investigate many types of problems and a split-screen capability that allows students to view a table and a graph at the same time.

III. Student Responsibilities

Students must learn not only how to operate their calculators, but also the limitations of the technology. New and faster ways to investigate problems and their solutions have been opened with the use of technology. Solving a ten by ten system of linear equations by hand would be tedious and time consuming. Using a TI82, keying in the problem would take the most time, the solution is found with a few key strokes.

Students must consider the nature of approximate answers and when errors can produce results that are not acceptable. It is not enough to correctly key in a problem. Students must be aware of the dangers of approximation and estimation.

IV. Teaching With an Overhead Graphing Calculator

Instructors can use an overhead unit to display an example or to lead a discussion of a concept. Multiplying two matrices during the solution of a problem can be quickly accomplished and a question posed "What if the conditions where changed to ...?"

Incorrect procedures that students frequently use with the graphing calculator can be demonstrated followed by the correct procedures. It is essential that students are aware of the dangers of the incorrect use of any type of technology, including graphing calculators.

V. Using Technology to Enhance Mathematical Learning.

If a student only uses a calculator to multiple two time five or to graph the straight line y = 3x, then the student is not properly using the power of the graphing calculator. Instructors should structure lessons where an appropriate use of technology will enhance a mathematical concept. Using matrices or tables on a graphing calculator to investigate a significant problem in mathematics is worthwhile.

Students should be encouraged to investigate and explore questions at all levels. For example, a student in calculus posed the following question during a discussion of initial value differential equations: "Can the function be found by graphing the derivative and investigating the initial conditions?" A student in college algebra posed the following question after a discussion of horizontal asymptotes: "If I go far enough out, would the curve ever be the same as the asymptote?" Technology may or may not be able to demonstrate the answers to questions like these but the students are being led into thinking about mathematics.

IMPROVING LEARNING OUTCOMES WITH THE HAND-HELD GRAPHING UTILITY

Martin K. De Beer
University of Guam
UOG Station
Mangilao, Guam 96923

In this paper, I share my experience teaching pre-calculus to students with the appropriate use of the graphing calculator as a learning device in a manner designed to improve learning outcomes.

The students in my pre-calculus classes come from culturally diverse backgrounds with varying degrees of preparedness in mathematics. The students are reminded frequently that they need to acquire graphical, numerical, and analytical competencies in studying the behavior of functions by mastering the traditional pencil-and-paper techniques of free-hand sketching of graphs of algebraic, transcendental and circular functions. The graphing calculator can be a useful tool to re-inforce learning.

Technology is pervasive in our society. Students will acquire the learning devices such as the graphing calculator. It is preferable for the Instructor to allow for optimal use of such learning devices in the curriculum. The graphing calculator is a versatile learning device. It is portable, user-friendly, and relatively inexpensive.

The dynamic imaging capability of the graphing calculator is exploited with suitable parameter and variable changes in studying the graph of a function. These exercises provide for superior reinforcement in the learning process. The students soon realize that simply keying in a function on their graphing calculators, while yielding a wonderful display on the screen, does not help them learn effectively.

The students completed a feedback instrument dealing with graphing calculator use, at the end of the course. Fifty percent of the students in the class had graphing calculators. Classroom activities with the graphing calculator in a cooperative learning environment exposed all the students in the class to the use of the graphing calculator.

An example of a class exercise starts with me explaining to the students the necessary skills needed to draw the graph of the tangent function on the chalkboard, together with horizontal and vertical shifts. Then, I ask one of the students with a graphing calculator to key in the function on his/her graphing calculator and display the graph. This screen display is shared with all the students in a cooperative learning effort. Parameters of the function are then varied and the respective graphs on the screen are shown to the students who appreciate the ease with which one can do this.

The students, especially those with graphing calculators, were convinced that these learning devices vastly improved their understanding of the course material and thus contributed to their success in the course. Ironically, the two top students in the class did not have graphing calculators, taking my caution rather literally, and preferring to use free-hand techniques. These two students were, nevertheless, very positive on the value of the graphing calculator as a learning device.

Samples of written student responses regarding the use of the graphing calculator are listed:

> "It seems to help because after you've done all the work and you get the same graph on the calculator, it makes you realize that you know what you're doing."
>
> "Graphing calculators should be used, but only to help students learn. Students should not take advantage of the calculators. They should learn the steps taken to reach the answer."
>
> "It allowed me to change and redraw function graphs quickly."
>
> "Looking at the graph on the calculator screen makes the graph of that certain function visible. You can check if your answers are right."
>
> "The graphing calculator is important to double check. Sometimes my graphs are very different because of a simple mistake."
>
> "I was able to manipulate function parameters and variables in many ways which would allow me to see the effects immediately. I feel naked without a calculator."
>
> "I found the graphing calculator activities in class to be helpful because I personally use my graphing calculator to re-inforce all sketched graphs on the chalkboard."
>
> "There was alot of things that I didn't know about functions, but with the help of my calculator, I discovered exactly how sin, cos, and tan functions behave."

> "When in doubt of your own work, the calculator (when keyed correctly) will provide the correct graphs. You are able to analyse or compare the two graphs. Thereby discovering the behavior of functions."

I am encouraged to continue the appropriate use of learning devices such as the graphing calculator in my classroom to improve the learning outcomes. I suggest Mathematics Instructors recognize the pervasiveness of these learning devices in our midst and that they consider using these devices in a manner that they find appropriate. The students will acquire these devices whether we like it or not.

References:

[Sc] A. H. Schoenfeld, A Brief Biography of Calculus Reform, UME Trends, Joint Policy Board for Mathematics, January 1995.

[Ku] A. Kugushev, Teaching, Learning, and Publishing in the Age of the Inconceivable, Focus MAA, December 1993.

[Vé] Y. Vélez, Some Thoughts on the Funding of Mathematics, Notices Amer. Math. Soc., February 1994.

Moving the Gatekeeper: Changes in the Mathematics Classroom when Computer Access is not the Issue

John Bullock, Massey High School, Auckland, New Zealand

Mike Thomas, Mathematics Education Unit, University of Auckland

Jackie Tyrrell, Mathematics Education Unit, University of Auckland

> The first use of a technology always consists of striving to do better what had been done before.......Time and the growth of ideas are usually needed before the idea of using a new technology to do something that had never been done before can even be conceived....Although the ultimate goal [may be] the same, the means....were epistemologically different in that they used a different ways of thinking (Papert, p.185)

Overview

This paper is a report on some of the teacher changes observed within the computer cluster of a one year Exploratory Study, Technology in Mathematics Education (TIME), undertaken by the Mathematics Education Unit, University of Auckland and teachers from eighteen schools in the greater Auckland area, on behalf of the New Zealand Ministry of Education[1]. During the process of writing the report all those involved came to the realisation that in spite of encouraging results, more questions than answers had emerged in the relatively brief course of the research in which they had been immersed.

For the purposes of the project the schools were divided into three groups, one for each of the technologies, calculators, graphic calculators and computers. The graphics calculator cluster teachers and the computer cluster teachers both let their students use personal calculators exactly as they might have in the past.

The division of schools into separate technology types was not done to compare them in any way, but so that the teachers and students could give their full attention to the implementation of whichever technology they had. The schools were provided with two computers of their choice, to be available to the research teachers at all times. This recognised that access is a key issue in the implementation of computing technology into any curriculum area. The choice of platform was the teachers' also, so that any personal blocks over unfamiliar equipment could be minimised. Some schools wished to add the two machines to their computer room, however they changed their minds when it was made clear that if that were to happen then the computer laboratory would need to be available to the mathematics teachers for all their periods with the research classes. Thus the teachers had continuous unchallenged access to two computers for mathematics. Each machine had Microsoft Excel 4 for Windows loaded, or Eureka for the Acorns. Other software, such as Logo, Grid Algebra and the Itmo series, was put on the Acorns.

Selection of the schools for each cluster was based on location, type of school (co-educational or single sex) and socio-economic and ethnic mix. It was not a random selection but did attempt to cover the cross-section of the New Zealand population based on our perception of the population. For the computer schools in the project the ethnic proportion was

54% European 21% Maori 16% Polynesian 8% Asian 2% Other

The principals described the socio-economic position of the students in their schools as ranging from severe low income to upper-middle income, as determined by the Ministry Elley scale.

The ratio of female students to male students was 49% to 51%

[1]The project staff acknowledge the funding and support of the Ministry of Education of New Zealand in this study.

Introduction

In this study there were glimpses of a mathematics completely turned around with style of teaching being the key to passing the power of the technology to the students. All in the classroom became learners in this situation.

When computers first became available to teachers in schools there was much optimism as to the benefits that would accrue from their use in mathematics teaching. Later a U.K. Mathematical Association review, which looked at the potential of the computer, stated:

> Styles of teaching which involve the use of microcomputers can aid greatly the acquisition of... important high-level skills which are exactly those which will be required in the future.
>
> (Ball et al, 1987, p.8)

However the reality is often not the same as that perceived or hoped for by an enlightened minority. As Freudenthal (1981) predicted, the transition to the computer paradigm has proved to be one of the major problems of mathematics education throughout the eighties and into the nineties. Cockcroft (1982) also realised that, for the computer era:

> ...there are two fundamental matters which need to be considered. The first concerns the ways in which calculators and microcomputers can be used to assist and improve the teaching of mathematics in the classroom. (Cockcroft, 1982, para. 374)

A difficulty, perhaps, is that "assisting and improving" is viewed in the way of "striving to better what was done before. The question which remains is why so many mathematics teachers seem unable or unwilling to make use of technology widely acknowledged to have great potential for mathematics learning. Hoyles (1992) indicates that research has only recently acknowledged the complexity of the role of the teacher in the implementation of technology, and the crucial role of teacher beliefs and assumptions.

Several studies which have looked at teachers using computers in the mathematics classroom have provided some insight into the criteria for successful implementation. One such study was the Queensland Sunrise Centre Project in Australia which began in 1990 and was a major initiative charting the direction in which technology use in education could proceed. Finger & Grimmett (1993, p.86) discuss 12 major issues which need to be addressed when managing and supporting technology initiatives in schools. They note that it is vital: to recognise the importance of people; to build ownership of the rationale of the initiative with the key participants; and to appreciate both that technology has implications for classroom organisation and management, and that teachers require both training and professional development in the use of the new technology. Logan & Sachs (1987) found that in teacher development programs that produced successful learning, participants believed in the importance of the task on which they were working. Finger & Grimmett (1993) also list strategies which they believe promote success in projects involving teachers in implementing technology, including:

> Regular meetings of project, school and regional personnel who have responsibility for the innovation.
>
> ...it is imperative that teachers be provided with the opportunity to vary their professional behaviour in such a way that they can facilitate the adoption of changes in educational practice afforded through the use of technology for learning. (Finger & Grimmett, 1993, p.88)

Acceptance of the ownership of the project by those involved was also identified by Treagust & Rennie (1993, p.51) as one of the factors necessary for success.
The Primary Laptop Project at Warwick University in the U.K., where students were given high levels of personal access to computers in mathematics, commenting on why they thought that they had seen so much success with the technology:

> ...in order to achieve these results, much more than access to personal technology was needed. It seemed to us that the support given to teachers at the planning stage helped them in a fundamental way to appreciate how activities could be set up which would have the potential for this learning to take place...
>
> (Ainley & Pratt, 1993, p.8)

The two aspects described above, namely the need for teacher support in order to raise confidence levels and the significance of the participating teachers' attitude to the project, appear as a common thread in much research. Teachers particularly need to feel involved in the ownership of the research and have a positive attitude towards it if it is to be successful. In the New Zealand project reported on here these were major initiatives right from the start. We hoped to build up the teachers' belief in the importance and value of the research, and to provide encouragement through regular joint meetings, discussions and classroom visits.

We have observed that when computers are introduced into the mathematics classroom they appear to change the social structure of the learning environment. There appears to be a shift in students' focus from the central authority of the teacher, to the computer. This led us to consider questions such as:

- How is the inter-change of mathematical ideas different in the computer classroom?
- What is the role of the computer in these inter-changes?
- Could the computer be acting as a catalyst in the promotion of student learning and teacher change?

To attempt to synthesise best practice for the introduction and promotion of the computer in the mathematics classroom it appeared necessary to address the role of the computer in the permeation of the understanding and ideas of mathematics. That is how does the computer alter the construction of each student's mathematical understanding and the view of mathematics held, and so presented, by the teacher?

In view of the exploratory nature of the study which made parallel observations with teachers using calculators and graphic calculators, and the relatively short time frame under which it was designed to operate it was not expected that major findings would emerge. As (Papert, 1993) observed viewpoints on technology can be epistemologically different and in this project we found that each of the groups had a differing reaction to the task ahead. The calculator teachers felt that they had been using calculators for years so that there was nothing to research. The graphic calculator teachers were issued with a piece of technology that few have, and it draws pictures while those using the computers thought that they had been "too hard" for fifteen years or so, already, so here we go again. The effect of these starting positions on the courses that each of the three groups followed is outlined in the full report (Bullock et al., 1994).

Briefly, the calculator group members were free of equipment problems so that their focus was on new ways of effectively incorporating calculators.

The graphic calculator teachers saw their technology as a piece of equipment, not unlike a calculator, that will do what teachers have always done but more accurately and reliably, so that this group was able to pursue the path of

> ...striving to do better what had been done before (Papert, 1993).

They then surged forward initially before reaching a block and having to re-evaluate the mathematics and the wider implications of what they were undertaking.

The computer group was issued with a single piece of software, for several reasons. The mathematics teachers in New Zealand were dealing with a complete new National Curriculum so that the imposition of more personal and professional change had to be made as realistic as possible given that the teachers were also required to fulfil the requirements of their school department schemes and testing programmes. By offering a single piece of open ended software, the research could concentrate on finding the limits of the software without skipping off to something else when the application was not immediately apparent. It also meant that the teachers would need to look for innovative uses rather than picking up a ready made solution to someone else's problem. Even so, for much of the first term a recurring comment was

> when are we going to get some software we can use with the kids.

It would take two and a half terms for most of the teachers involved to shift and look to do what they had not done before. When it happened, it was like a flash of insight, changing their viewpoint.

> I've just realised something, about twenty minutes ago, it's not that I've got a computer, how can I do some maths. It's I've got some maths to do how can the computer help me to do it
>
> I'm very proud of myself, I just designed a spreadsheet to do some of my work
>
> I reckon I could do just about anything now. I didn't feel like that at the beginning.

> I was struck by the sudden shift from totally hopeless to completely convinced through the course of one good activity - one day.

We found that this change is not a minor shift which can be easily accommodated or go unnoticed if the successful integration of computers is to take place.

> ...a total rethink on lesson and unit plans has been required

> There is a great deal of effort required for teachers to move outside their 'comfort zone', a huge supporting beam of the teacher barrier to the implementation of technology

The results of the TIME project seem to mirror the work of Ainley & Pratt (1993) at Warwick University in that both research support staff and the teachers themselves felt much as the Warwick team did.

The students have to go through the same process of change as their teachers, since often their perspective of computers is that they are for games or 'other things'. Our study suggests that the change for them will take place over a much longer period of time if they are only subjected to intermittent exposure through highly planned trips to the computer room for a particular unit of work.

On the other hand:

> If you start the year with it they accept it as normal. It's no big deal to them they just get on with it.

If the shift of mindset from the traditional approach to a technology enriched approach to the teaching and learning of mathematics can be successfully negotiated, then there is the opportunity to teach students to begin to think mathematically. To move away from the detail of the internal calculations associated with the solution of the particular current problem towards the general principles involved with more importantly, other, similar problems. Number and estimation begin to take on a pervasive importance underlying the nature of the technological method. What becomes important is not the detail of number work, in the sense that the student needs to be able to actually make a calculation using a clay tablet and a stick, but the form and feel of answers which may be being generated and the relation to the underlying principles. For instance a group of 11-12 year olds

> realised while they were using the number chart that the times tables do not stop at 10 or 12.

The students seem compelled to talk about the mathematics that is happening on the screen.

> [I'm] not a group person myself, but putting students in front of a screen, [it] seems to be natural for them to discuss things.

> I now allow the students more time to experiment and test different ideas, and time to discuss and share their ideas and findings.

Students were empowered to explore an array of possible solutions without vast expenditure of energy and effort calculating or graphing data. Thus in a single lesson they were able to try a variety of approaches to solving a problem. Unsatisfactory results were seen as providing learning experiences, rather than the stigma of a "wrong" answer.

> [It seems that we] can ask a student to be critical of what's written on the screen in a way that we can't criticise what they have in their books.

> I think that the potential that maths offers is for them to become conscious critical thinkers in the widest sense. There are opportunities for this with [the spreadsheet]

Students were able to generate results for a large number of examples then use the information to make decisions or generalisations.

> [The time saved] is better spent on drawing conclusions.

From the TIME study we note that the notion of basic requirements for the study of mathematics seems to need reviewing. For example, no longer is it necessary to know that the product of 19.87 and 12.32 is 244.7984 but it is important to be able to estimate that the answer will be about 20 by 12, which, for students who know their times tables, is 240, or 20 by 10, and hence the answer is of the order of two hundred and something. Not only have the basics changed but also there needs to be time spent finding ways of assessing just what is being learned and whether it is what is important for mathematical thinkers of the future. A local report in New Zealand headlined that students in Scottish primary schools, where constant calculator use had been encouraged for some time, were outperformed in a basic skills test by Japanese children, who had been brought up on a regimen of rote learning. Presumably calculators were banned for the test. It is our belief that these types of reports serve only to fuel teacher resistance and inhibit the diffident from making that start for yet another year and community attitudes to harden about the morality of doing a job "the easy way" all the time so being left high and dry when your batteries go down.

One of the most common points the teachers commented on was the *improved student discussion of mathematics* as a direct result of the introduction of the computer into their classroom. Typical comments they made were:

> Putting the students in front of a screen it seems to be natural for them to discuss things. It encourages stopping to look at what you've been doing. *[T12]*
>
> Students mostly show a willingness to help each other and to share their knowledge and discoveries *[T3]*
>
> An interesting phenomenon is that often it is not the person operating the keyboard who is doing the most mathematics, rather, those seated around watching the screen and making suggestions. *[T8]*
>
> The computer becomes a talking point, which creates more discussion. *[T16]*
>
> I have noticed an increase in the volume of noise in the class when introducing spreadsheets. The discussion is generally on task but because of different ways of solving or setting up a workable sheet discussion is necessary. I have found that the student discussions are valuable *[T29]*
>
> The biggest change that I have noticed in student behaviour would be the increased levels of discussion, based on the computer activity. Nearly all the students freely volunteer their ideas and are able to articulate them *[T11]*

Further, the teachers found that not only did students discuss more but they noticed *improved student co-operation in mathematics* with students seen to be helping each other with the work to a greater extent than they would normally do. In this area the type of comments made by the teachers included:

> When it's pen and paper, they tend to be looking at each other, to be copying off each other, whereas when it's on the computer they're helping each other out. *[T5]*
>
> I have witnessed numerous instances where students, who would not normally co-operate, do so with the full realisation that...they would not have the privilege of using the equipment *[T11]*
>
> ...led to a greater group co-operation, talking things through, discussion *[T2]*
>
> Group dynamics tend to be better than usual on the computer than off it. *[T8]*

One example of such co-operation related by a teacher was that of a girl who wanted to take her friend in for another session on the computer because "she isn't very good at it. Could we do that?" The teacher replied that it was OK with him! There was also some evidence of inter-group relationships, with different groups of students being keen to discuss with other groups the work they had been doing on the computer.

> Groups were also discovering different aspects of the programme at different times and were keen to find out what other groups were doing. *[T2]*
>
> If one group gets stuck they'll look over the shoulder of another group to see what they're doing, to get their ideas. *[T5]*

The encouragement of this *inter-group mathematical discussion* is another example of a factor produced by the computer altering the social structure of the classroom. The likelihood of this being a stimulus to better learning is evidenced by the students' answers to one of the questions in a survey given to them at the end of the study. When they were asked how they like to learn:

- Σ 40% said by working with a friend
- Σ 30% said by working it out themselves
- Σ 12% said by asking a teacher

Thus working it out for themselves, which the technology encourages, or with the help of a friend is clearly preferred to the traditional method of asking a teacher. The computer empowers the students by giving them more control over their learning. This is a clear shift in authority over learning in the classroom, and one which we believe has beneficial consequences. One with which teachers themselves need to come to terms.

One possible reason for the better co-operation which the computer engenders is that the computer seems to act as a catalyst in the exploration of the meaning of mathematical language. It seems that many students fail to comprehend much of the sophisticated language used by the teacher in the mathematics classroom, a point often overlooked by teachers. The computer actively promotes discussion on this:

> When I walk around I am quite surprised that most of the talk is on task, often one student will be describing the work to the others in their own language *[T3]*

> [Using the computer for mathematics] generates mathematical discussion. If you believe mathematics is a language, then it enriches this language. *[T3]*

> The language causes problems for our kids. They won't ask what a word is, even if they're unsure - they just keep the first picture they get in their minds, and don't change it...Looking back I should have got their word for the key commands. *[T38]*

It appears that the computer's presence in the classroom alters the social interaction between the students themselves and between the students and the teacher. Without the computer students often seem to feel unable to ask questions of the teacher about the language and symbolisms of mathematics and are often denied the opportunity to engage in little among themselves. When the computer is introduced into the classroom it enables discussion on these issues between the students. This may be because the neutral nature of the machine provides a less threatening environment, or possibly the insecurity of the teacher not having "the answer" then is more likely to allow discussion, but the full reasons for this need further investigation.

Reduced Control of Students through Increased Motivation.

In spite of its lack of novelty value in today's technological society, the computer still has tremendous power to motivate and influence the students' learning of mathematics. *Improved student motivation in mathematics.* was often commented on by the teachers. Typical remarks from the teachers were:

> Students seem to be more driven to get results when using the computer. They want to see things happen and get work onto the screen. *[T5]*

> Students do seem to be more motivated when using the computer and their level of observable 'on task' behaviour certainly is increased when using the technology *[T11]*

> I have a normal teacher's gut feeling that they would rather do maths incorporating the computer *[T6]*

> In my third form classes the students are completely enthusiastic when it comes to using the computers. *[T8]*

This improvement in motivation in mathematics was confirmed by the students, who in response to "When I use the computer my mathematics is...", 26% said it was more interesting and a different 27% said it was easier.

A teacher noticed that, after the bell went for lunch time and the class had made a quick exit as usual, the bags of those who had been using the computer were still on the desk. On walking next door where they had been, he was surprised to be told that they were aware that it was lunchtime but "could we get our spreadsheet finished, especially the code?" He commented how they loved making up the code, and said that

> This was the first time in my 20 years experience of teaching that the lunch bell was of no consequence. *[T6]*

One reason for this increased motivation is clearly the perceived novelty or fun value of the computer, or as one teacher expressed it

> In both of the classes that I have worked with, the students seem more motivated by the computer lessons. Computers are generally associated with fun (Teacher: How was that lesson Student 1: Aw that was <u>fun</u>, sir. Student 2: Yeah that was really neat, can we do that again?) *[T3]*

Commenting on a different possible reason why the computer might have such an effect, one teacher said:

> One wonders if they see the computer as less critical than homo sapiens, or if the computer has a more polite way of saying 'try that again', or is it that the computer refrains from using a red pen? *[T11]*

This teacher also related the comments of the students who said, after one of the support team had left the classroom, that

> the computer is nicer than [the team member]. It only makes a noise when we make a mistake, it doesn't say words. *[T11]*

These are interesting observations on the way in which the computer seems to lessen the pressure on the individual student because the relationship with the computer appears de-personalised and, apart from the less obvious screen feedback, consists primarily of one-way communication that is more acceptable to them. This situation is one where the child often feels more in control, in contrast with the teacher-student relationship, where they perceive the teacher to be firmly in control of their learning situation.
Being placed more in control they also appear to feel that they have greater privacy in their learning

> Because the computer is a non-critical tool many students felt more comfortable attempting to solve problems where the errors would be kept private. Also many liked the idea that there was no permanent written record of what they had done and therefore could not be used to refer to past mistakes. *[T3]*

Coming to terms with the reduction in written work is another area that requires adjustment by the teacher. A major point of discussion early in the project was directed at the observation and handling of student work. Should they have a disk each and save their efforts, should they save them on the hard drive or should they print everything out for the teacher to mark. The internalisation of some of the learning process is recognised by the teachers and while some did attempt to track all student work initially they came to see that the most effective learning took place if the students could make their way without the spectre of always having to be right hanging over them. This need to keep one's learning, and particularly failings, private was observed in a classroom. On one occasion, a boy was working on his own on a computer and there was a girl on the computer next to him. He was observed turning the screen of his computer away from her so that she could not see what he was entering. This was at a point where he was having difficulty with the problems he was attempting.

Another feature of tackling real world problems is handling realistic data. In the project the students were able to tackle problems where the amount of data required would mean that they could not be investigated otherwise. One example of this was when the students were involved in discovering the features of a frequency distribution for the number of 1's, 2's etc thrown for 10,000 throws of a normal, unbiased die. Using the random number generator of the spreadsheet, filling down through 1000 cells and using the dynamic graph drawing facility, the students were able in a short time to construct the bar graph for the die rolls. Noting their results and then repeating this a further 9 times and combining the distributions they were able to get a feel for the distribution in a time so short it would not be possible to emulate it by other means. This greatly reduced timescale in investigative work, leading to the students looking at problems they could not be introduced to by other means is another powerful benefit of the computer in the mathematics classroom that we have observed.

Summary

On the basis of our year's study we have seen that when the computer is integrated into the mathematics classroom it changes the social structure of the classroom in ways which enhance student learning. Using the computer seems to:

∑ improve cooperative learning between students
∑ increase mathematical discussion and
∑ improve the motivation of students towards their mathematics thereby reducing the strain of teacher control.

To fully benefit there is a learning time which must be worked through with students and teachers needing to review their assessment procedures and to monitor the learning and mathematical progress of students as they work on the computer.

However, attaining this level of integration requires a shift in the thinking of the teacher with regard to their teaching style and their view of the role of the computer in student learning.

The teachers' behaviour with and view of the computer act as a model for the students. Positive behaviour evokes a positive response and vice versa.

The role of the teacher changes from provider to guide.

Those teachers working in schools where their mathematics department had a common positive focus on the use of computers become confident users in a shorter period of time.

When issues arising from computer use are not addressed by the mathematics department then those wishing to make change expend enormous effort trying to fit the computer into existing structures.

We also found, as we would expect, that there were issues of storage, security and movement of the equipment that needed to be addressed before progress with the mathematics is possible.

New ways of managing the class are needed and issues of equity of access for student use have to be dealt with.

Fundamental changes in teaching and learning are characterised by a move towards mathematical processes and active learning experiences. Changes that only occur if the computer is not seen as an added extra which can be used for certain topics only. It must form part of the teacher's normal mathematics lesson.

Teachers take about 12 months of constant use and support before a lasting transition takes place.

Acknowledgments

The quotations used are from the teaching staff in the project.
The help of the mathematics departments of Helensville Primary School, Howick College, Howick Intermediate, Kaipara College, Tamaki College and Tamaki Intermediate is appreciated.

References:

Ainley, J & Pratt, D. (1993). Portable Mathematics and Integrated Technology, in B. Jaworski (Ed.), *Technology in Mathematics Teaching: A Bridge Between Teaching and Learning* (pp. 101 - 107). Birmingham, U.K.

Papert, Seymour. (1993). *The Children's Machine: Rethinking School in the Age of the Computer*, BasicBooks.

Bullock J, Barton B, Buzeika A, Ellis J, Regan M, Thomas M & Tyrrell J., (1994). *TIME report*, Mathematics Education Unit, University of Auckland.

Ball et al., (Eds), (1987). *Will mathematics count?: Computers in Mathematics Education*, The Mathematical Association, Hatfield, England: AUCBE.

Cockcroft, W.H. (1982). *Mathematics Counts*, London: H.M.S.O.

Finger, G. & Grimmett, G. (1993). Twelve Issues to Consider for Managing and Supporting Future Technology Initiatives in Schools, *Australian Educational Computing*, July 1993, 85-92.

Hoyles, C. (1992). Mathematics Teaching and Mathematics Teachers: A Meta-Case Study, *For the Learning of Mathematics I2-3.* (pp. 32-44).

THE USE OF THE MICROCOMPUTER IN OBSERVATION OF STUDENT DIFFICULTIES WITH THE ACCESSING OF MATHEMATICAL KNOWLEDGE

Mohan Chinnappan
Centre for Mathematics, Science and Technology Education
Queensland University of Technology
Brisbane, Australia

and

Michael J. Lawson and Dane Gardner
School of Education
Flinders University
Adelaide, South Australia

This research was supported by grants from the Australian Research Council and the Queensland University of Technology. Address correspondence to: Mohan Chinnappan, Centre for Mathematics and Science Education, Queensland University of Technology, Locked Bag 2, Red Hill, Qld 4059, Australia.
Email address: m.chinnappan@qut.edu.au

Introduction

Mathematics educators and researchers alike have shown an increasing interest in understanding how students acquire and use mathematical knowledge in the performance of a great variety of tasks. This focus on mathematical knowledge acquisition and utilisation is also reflected in the goals of major mathematics curriculum documents (National Statement, 1990; National Council of Teachers of Mathematics, 1989) which emphasise the constructive or generative aspects of mathematical learning and understanding. Thus, an important aim of classroom teaching is to help students not only learn mathematics concepts, principles and procedures, but more importantly, to examine ways of facilitating the activation of this prior knowledge during activities such as problem solving, conjecturing and construction of explanations. In order to achieve this aim, among other things, we need to understand the nature of students' mathematical knowledge.

Recent developments in cognitive science have generated data about how the nature of the organisation of mathematical knowledge drives high levels of generative processes typically observed in expert-like performances. There is an emerging consensus that the way mathematical knowledge is represented in memory has a significant effect on how that knowledge will be deployed during task attempts (Prawat, 1989). An important question about the organisational aspects of mathematical knowledge is how does the structure of the knowledge influence the accessing and subsequent use of that knowledge in mathematical tasks. Data relevant to this question has important theoretical and practical implications for mathematics instruction and teaching.

The challenge for researchers interested in investigating the use of knowledge in mathematical tasks is to establish what it is that students know about the task. In order to describe how knowledge is used, or not used, during performance it is necessary to be able to develop reasonable estimates of what knowledge students have available for use. It would also be desirable to gain some information about the ease of accessibility of that available information. In a recent study of knowledge activation and use during problem solving by high-school mathematics students we have addressed these problems of knowledge availability and accessibility (Lawson and Chinnappan, 1994). However, we are far from describing the nature of student's mathematical knowledge and how the quality of that knowledge constraints accessibility.

In this paper we report the design and trialing of a Hypercard testing procedure known as the Recognition and Testing System (RATS) that will enable mathematics teachers and researchers generate data relating to geometry knowledge representation and its access.

RATS is designed to provide three types of information about students' geometric knowledge. Firstly, it provides an estimate of what a student knows about geometry and trigonometry. The program identifies and records the geometric and trigonometric knowledge that can be recognised and labelled by a student. For purposes of this investigation we have conceived geometric and trigonometric knowledge as being composed of a set of units, or knowledge components. These components could be seen as nodes in a network structure or as units in a connectionist system.

A second type of information recorded by the RATS program is the level of assistance required by the student to access these knowledge components. As will be apparent to teachers and researchers interested in problem solving and reasoning, many students can access some relevant knowledge by themselves but require prompting or assistance to retrieve other knowledge components important for the development of a solution or argument. RATS is designed to identify how much assistance is required by a student in order to recognise and label specific knowledge components. Those students who require no assistance will move through the first level of the program and will finish quickly. Students who require some assistance will be taken through other levels of the program. A description of these levels is provided later in this paper. Knowledge components which a student cannot recognise and label when given the highest level of assistance are regarded as being not functionally available.

The third type of information gathered in the program is data on speed with which a student could retrieve the required name. The program traps the period of time required by the student to recognise a knowledge component and records both that time and the level of assistance at which the successful recognition was made. Speed of recognition data does not include time taken to type the name of the component.

The program records all the above student information in data files. Additionally, questionnaire and student categorisation data is recorded. At present, this additional information comprises demographic data and students' ratings of their general academic achievement, their mathematics achievement, and their attitudes to mathematics.

The logic of the design of the RATS program rests upon a set of assumptions made by cognitive researchers. The first of these is that knowledge is organised in schemas or clusters (so that, given an appropriate problem context, the solver is able to retrieve relevant information in developing a solution (Alba & Hasher, 1983; Neisser, 1976; Anderson, 1990). Sweller (1990) argued that the problem schema allows the classification of the problem and also the generation of moves that are appropriate for the development of a solution for the problem. The RATS program is designed to identify the knowledge components in a problem schema that can be recognised and labelled by a student.

A second assumption is that the prompting procedure used in RATS provides information about the state of organisation of the knowledge schema. Within a specific domain, it is generally agreed that a better organised or structured knowledge base facilitates the accessing of information contained within that domain. Organisation or structure implies the existence of connectedness between units of information. Prawat (1989, p.13) observed that 'organisation is largely a function of connectedness.' It is assumed that knowledge components that are only accessed with assistance are less strongly connected and so are less readily activated (Anderson, 1990). A final and related assumption is that the speed of recognition also provides information about the quality of knowledge organisation. Again it is assumed that longer recognition times indicate knowledge that is less well connected within the problem schema or between schemas (Mayer, 1975), or a knowledge structure that has a lower degree of coherence (Chi, Hutchinson and Robin, 1989) .

Recognition and Timing System - RATS

RATS is a Hypercard stack (or program) designed to offer the researcher or teacher a flexible system for testing subjects' knowledge of visually presented information or images (termed diagrams), and of the time required for recognition of whole diagrams or their components. The diagram constitutes the basic unit of structure for the program. Within each diagram the researcher defines sections (termed components) to be identified and labelled by the subject during testing. Buttons are defined for each component and are then positioned on the diagram where required. At the same time, the fields used for the subject's response are located near the component.

The contents of a diagram can be geometric shapes, maps, pictures or anything that can be drawn on or scanned into a Macintosh computer. The number of components that may be on a diagram is from one to as many as will practically fit on the diagram. The researcher or teacher may include as many diagrams as demanded by the investigation.

Levels of assistance

Each diagram can be presented at three levels of assistance. These are defined as Levels 1, 2, and 3.

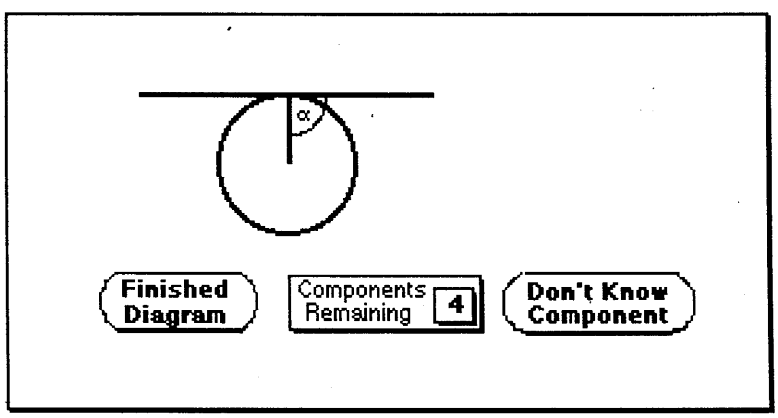


Figure 1. Sample screen layout for Level 1.

When a diagram is presented at Level 1, as shown in Figure 1, the subject is required to identify one component. Selection is done using a mouse and the subject's label for the component is typed into the field that is subsequently displayed, as shown in Figure 2. If a subject mistakenly selects a component for which a label is not known, the "Don't Know Component" button can be selected. Then the next diagram is displayed and the recognition and labelling procedure is then repeated. The remaining diagrams in the test are then presented.

Presentation of all diagrams is then repeated until all components have been labelled or until subjects indicate, by clicking on the "Finished Diagram" button shown in Figure 1, that no further components can be identified. If the subject correctly labelled all components of all diagrams the testing procedure is then completed. However, if any component remains to be correctly identified, the appropriate diagrams are redisplayed in Level 2. The determination of whether a component was correctly labelled is done using a verification algorithm that can be varied for each diagram.

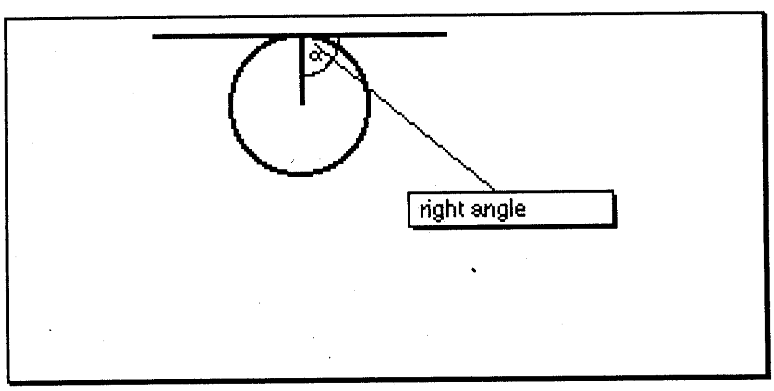


Figure 2. Level 1 screen

Diagrams that had all components correctly identified in Level 1 do not appear in Level 2. When a diagram is shown on the screen during Level 2, any components correctly named during Level 1 will be displayed and labelled, as shown in Figure 3.

Only those components that were incorrectly identified or not identified in Level 1 will need to be identified in Level 2.

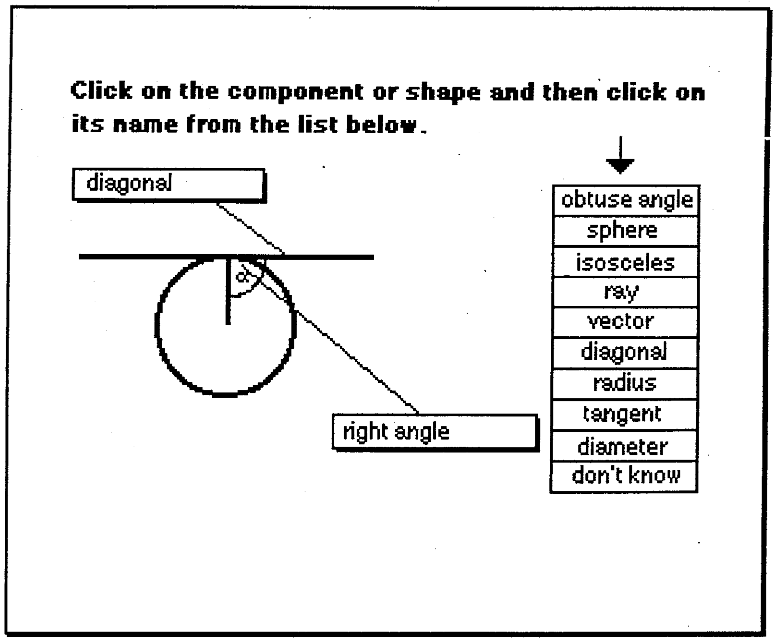


Figure 3. *Level 2 screen*

Identification by the student is again done by clicking on the component with the mouse, but this time, instead of typing the label subjects select one from a list of possible labels provided on the screen. A sample list is shown on the right of Figure 3.

At this level, the diagrams are presented using the same cyclic method used in Level 1. Again, there is the option for subjects to finish with this diagram if they can not identify any further components.

Any diagram in which all components have not been correctly labelled will be presented again in Level 3. Each diagram is displayed in the same way as in Level 2 using the same cyclic procedure. The difference in operation is that when a diagram is presented at Level 3, a component remaining to be correctly identified is highlighted on the screen and the subject must select a label from the list, as shown in Figure 4. As with Level 2, a "don't know" option is available. Once all diagrams are finished, the basic cycle of a test is complete.

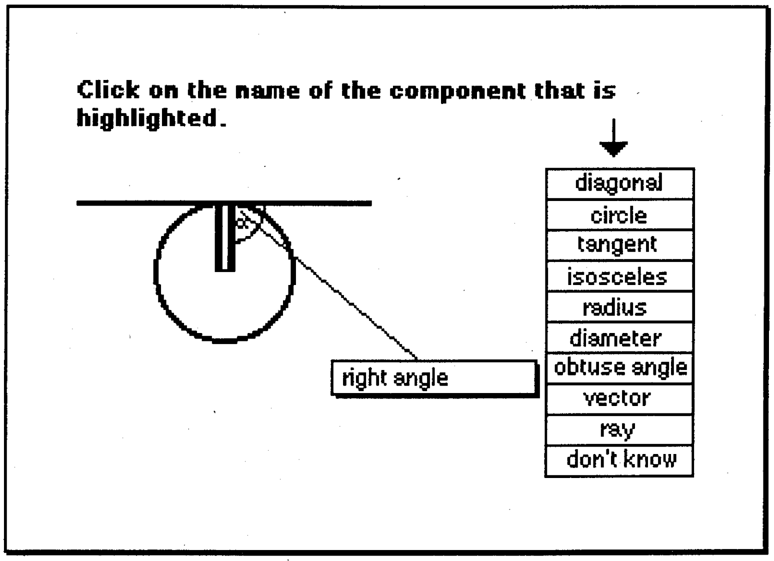


Figure 4. Level 3 screen

Minimising Impact of Differing Subject Backgrounds

The timing for speed of recognition is from presentation of the diagram to the selection of the component. The program includes a training stack in which all procedures used within the testing are presented and practised. Use of this demonstration stack is designed to reduce differences between students in familiarity with computers and with the procedures specific to the RATS program. The researcher then need not be concerned with a student's lack of familiarity with the keyboard. In research settings where individual testing is involved, or even in group settings where students work on a bank of computers, we have advised students to check spelling with the researcher or teacher, so that verification problems are minimised.

Computer Environment

RATS is written in Hypercard and will run on Macintosh systems with versions of Hypercard from Version 2.0. Although it is advantageous to have operational knowledge of Hypercard this is not necessary for operation of the system. It is also preferable that the Macintosh operating system used is at least System 7.0.

The amount of disk space required is a function of the size of the tests and the number of subjects being tested. The minimum disk space required is 600k. To accommodate Hypercard and System 7 a machine must have at least 4mb of memory.

Features and Applications

The nature of data that can be collected by RATS enable mathematics teachers and researchers to seek answers for a range of questions. As was indicated earlier in this paper, it is designed for the generation of data about issues relating to knowledge organisation and access within the domain of geometry and trigonometry. The levels of assistance built into the system provides an indirect index about the quality of students' geometry and trigonometry knowledge network. Knowledge components that the students could access at level 1 could be argued to be more tightly connected than those accessed at level 3; those not accessed at all despite provision of assistance can be regarded as not being functionally available.

RATS also provides an important avenue for classroom teachers who are looking for an alternative method for the diagnosis of conceptual difficulties. Such information can be used by teachers as part of a formative evaluation system. Our own investigations with RATS show that students find the program simple and interesting in operation. Although we have confined our use of the program to the field of high school mathematics it can be used with material in any area of the curriculum.

References

Alba, J.W. & Hasher, L., Is memory schematic? *Psychological Bulletin*, *93*, 203-231, 1983.

Anderson, J. R., *Cognitive psychology and its implications*. Third edition. New York: W. H. Freeman & Company, 1990.

Australian Education Council, *National Statement on Mathematics for Australian Schools* . Carlton, Vic: Curriculum Corporation, 1990 .

Chi, M. T. H., Hutchinson, J. E., & Robin, A. F., How inferences about novel domain related concepts can be constrained by structured knowledge. *Merrill Palmer Quarterly*, *35*, 27-62, 1989.

Mayer, R. E., Information processing variables in learning to solve problems. *Review of Educational Research*, *45*, 525-541, 1975.

National Council of Teachers of Mathematics, *Curriculum and Evaluation Standards for School Mathematics*. Reston, Va.:The Council, 1989.

Lawson, M.J. & Chinnappan, M., Generative activity during geometry problem solving: Comparison of the performance of high-achieving and low-achieving students. *Cognition and Instruction*, *12* (1), 61-93, 1994.

Neisser, U., *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: W. H. Freeman, 1976.

Prawat, R., Promoting access to knowledge, strategy and disposition in students. *Review of Educational Research*, *59*, 1-42, 1989.

Sweller, J., On the limited evidence for the effectiveness of teaching general problem solving strategies. *Journal for Research in Mathematics Education*, *21*, 411-415, 1990.

Teacher Conceptions and Approaches to the Design of Computational Learning Environments.

Kathryn Crawford, The University of Sydney, Australia.

Introduction

Teachers at all levels have been slow to incorporate new technologies into their teaching practice. Nowhere is the reluctance to change from paper and pencil techniques more evident than in the case of mathematics. Despite changes in the ways that mathematics is used in recreation and work (Willis, 1990, Fujita, 1993), in many Australian schools mathematics is taught in much the same way as it was twenty years ago (Speedy, 1989).

A systemic approach to human activity.

New systemic approaches in psychology (eg. Crawford, in press, Rogoff, 1994, Valsiner, 1994) suggest that people's beliefs and conceptions strongly shape their thinking, learning and actions. In particular, beliefs and conceptions influence the ways in which people react to new situations and challenges. In schools the activity system that constitutes an educational community can be thought of as consisting of teachers, their students, administrators and the cultural spaces, buildings and artefacts that are used by the community. The diagram below illustrates the inter-relationships and interactions in an activity system that involves a teacher students and a computer.

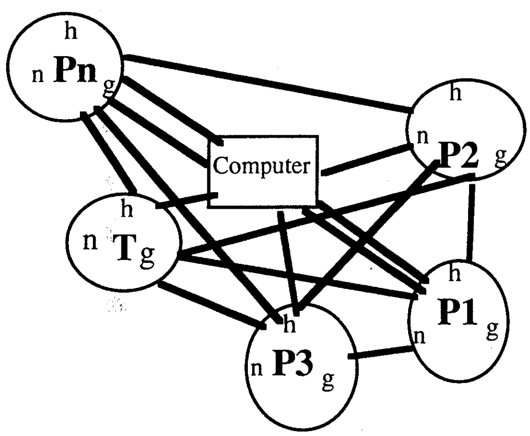


Figure 1.
Interaction of teacher and pupil system with and through a computer.

Each of the people present has a personal history of learning and experience, a personal need and a goal to which the activity at hand is subordinated. These factors influence the ways in which they think, feel and act. The computer, as a cultural artefact has a **meaning**, for each person in the group, that is culturally derived through past actions that have involved computers and explicitly related to the activity at hand. Software that is created for use on computers in educational settings is also a cultural artefact. Its form reflects the subjective perceptions of need and goals of the creators of the hardware and software. At present, computers are a highly significant tool outside schools but have not realised their full educational potential in schools. Research (Bornbholt, Crawford & Summers, 1992; Crawford, Gordon, Nicholas & Prosser, 1993) indicates that conceptions of the task at hand and personal needs shape the ways in which students approach tasks. It seems likely that similar relationships will be evident for teachers as they creative a context for learning which involves computer use by students.

In this paper the focus will be upon the interaction between a teachers and computers - the ways in which a teacher's personal history of experience of teaching, learning and computer use - their conceptions of the meaning of these terms - influence the ways in which they approach the computers as a component in a learning context.

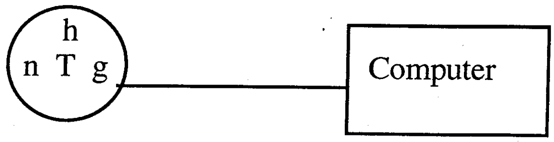


Figure 2 Teachers and Computers

Early educational software, in particular tutoring systems, were not designed and constructed by teachers. They reflected the programmers stereotyped conceptions of teaching, learning and the power relations between teachers and learners as they had been experienced in school systems. More recent software (eg, Laborde- Cabri Geometry, Swartz, Algebra -supposer), designed to facilitate learning through exploration, has not been popular in schools. Most early software, was based on a transmission model of teaching and learning, promoted rote learning of skills and procedures and provided evaluative feedback to users. The "computer as teacher" stereotype is still implicit in many computer interfaces. Thus, for educational users, experience with computers tends to reinforce traditional relationships between learners and teachers and to place users in a passive or dependent position. With this in mind, the following aspects of cultural experience were hypothesised to be key influences in teacher use of computers as part of an educational setting in schools.

- conceptions of teaching and a teacher's role
- conceptions of learning
- conceptions of computers as artefacts in educational settings.

Investigating teachers' conceptions.
A survey of school teachers was carried out using a phenomenographic (Marton, 1988) approach to gathering of data on their conceptions of **teaching**, **learning** and **computer** use in educational settings. The study is the first part of a larger study in which a larger sample of teachers will be surveyed. Seventeen teachers in a representative range of state and private schools at both primary and secondary levels participated in the small study. The response data indicated a wide range of beliefs and conceptions among teachers. There were also indications of inconsistencies and lack of integration between ideals and habits of practice.

Table 1 below indicates the five categories of description, and the number of responses in each category, that emerged when eighteen teachers wrote answers to the open ended question: *What is teaching?*

Category	Description	Number of responses
A	To present information or knowledge to the students	2
B	To present information based on students' previous learning and interests	3
C	To present information based on students' previous learning and interests and to establish a suitable classroom atmosphere.	5
D	To create an environment in which both teachers' and students' needs are considered and which includes experiences of group work, exploring and collaborating	4
E	As above but also includes the interests of both teacher and students - students work in areas meaningful to them and have ownership of their learning- interaction, negotiation and spontaneity are valued.	3

Table 1. Range and distribution of teachers' conceptions of teaching.

The five categories above are inclusive and represent a range of awareness on the part of the teachers. That is all teachers in the sample accepted the need to provide students with information. The categories represent a continuum from a focus on what the teacher does and an implicit transmission model of learning in which information is passed from teachers to students, towards a description in which teaching is increasingly described with reference to the task of creating and managing a setting in which particular kinds of learning processes take place. The categories display an increased attention to the students' experience. There is a substantial break between categories C and D. In categories A, B and C teaching is described in terms of what the teacher does and the factors which influence teacher preparation. In categories D and E there is a shift in awareness toward a consideration of student activity as a goal of teaching.

Table 2 below show the range of categories of description that emerged from teachers written responses to the question : *What is learning?*

Category	Description	Number of responses
0	No answer	7
A	Exposure to stimuli..guided introduction to experience	3
B	Teacher provides a variety of controlled situation and the students practice	4
C	Children learn in different ways through active and confident participation in activities which can be related to previous experiences and about which they feel positive.	3

Table 2. Range and distribution of teachers' conceptions of learning.

More than one third of the teachers who were surveyed did not respond to this question. This result suggests that the term learning is not well defined by teachers and is consistent with the results shown in Table 1 above. Of those who responded, more than two thirds provide a behaviouristic description of learning as a **response** to teaching. Only in category C does the description of learning focus on the students' thinking, feeling and acting in an educational setting.

Table 3 below shows the categories of description that emerged from teachers' written responses to the question: *How can computers be best used in a classroom?*

Category	Description	Number of responses
A	The computer as something else for teachers to teach or an "electronic teacher" use to do what teachers do.	12
B	Can be used to enhance information available and so that students learn to handle information/data.	2
C	Students have control of computers and they offer a way to cater for individual differences.	2
D	Computers provide a context for students' communication, discovery and construction.	1

Table 3. Range and description of teachers conceptions of computer use in schools.

This qualitative study was carried out using a small sample and should not be generalised. However, it is consistent with other larger studies (eg Crawford, Groundwater-Smith and Milan, 1989) which found that most teachers saw computing as a separate subject and did not use computers in mainstream subjects especially mathematics as they had difficulty seeing the relevance of a computer to their teaching role. The results also indicate that :

- Teachers have much more definite views of teaching than of learning.
- There were many inconsistencies between values espoused by teachers and their descriptions of their approaches to teaching.
- Their conceptions of computers as an artefact in an educational setting were strongly related to their conceptions of teaching and learning.

Designing and using a computer-based environment.
Using a computer as a part of an educational setting involves a creative design or novel educational problem solving task for teachers because computer technology is not yet usual in classrooms. Research (Semenov, 1978, Crawford, 19886) suggests that the processes of defining a novel problem, selecting a strategy to achieve a resolution and evaluating the results are particularly influenced by a persons conceptions of the factors involved. Thus it is expected that as teachers approach the new task of using computers as part of an educational setting their conceptions of their role as a teacher, how students learn and of the computer hardware and software will strongly influence their thinking.

A small purposeful sub-sample of teachers with differing beliefs were invited to engage in development and design of computational materials for classroom use in mathematics in which visual aspects of the computing medium were emphasised. The initial approaches to the task of two of them are briefly outlined below. Both teachers chose to explore the use of computers as they taught about polynomial equations.

Carol's design
When Carol responded to the questionnaire, discussed above, her responses were of the kind described in category A in each of tables 1 to 3 above. She described herself as using "traditional methods" to ensure that students "acquired" mathematical knowledge. When asked to design a computer-based educational setting she conceptualised the computer as an "electronic teacher" that would "teach" students mathematics.

Carol described the goals of her project almost exclusively in terms of mathematical content as:

"To study the relationships between polynomials and graphs."

She conceptualised the computer setting as "allowing" students to do certain things such as visualising relationships, changing variables and changing the graphical representation. She designed a series of carefully planned tasks for students. For her the screen was conceptualised as an "electronic chalk board".

In response to questions about how she expected the students to respond to the computer-based educational setting she provided the following behavioural objective that might have come from a traditional curriculum document:

"Given a polynomial, students are able to figure out its graph and sketch the graph."

As an afterthought, she hoped that students might find the computer-based setting motivating.

Helen's design

Helen held very different views about teaching learning and computers. Her responses to the initial questionnaire were classified as D, C and D respectively. She described her project, in very student centred terms, as:

"An investigation of student's use of mathematics in an open-ended investigation of polynomial graphs and the roots of related polynomial equations."

Her key goals were focussed on the learning experiences of the students. She was interested in the students' opinions about the open ended task that she was planning, what they were able to find out and the extent that their responses were as she had expected. All her planning rested on an assumption that the students would be active and self directed. She was concerned to create a learning environment in which they were able to act independently and reflect upon what they were doing.

Her plan was to evaluate her use of a computer-based medium in terms of the ways in which the students learned, the aspects of the interface that they attended to and the extent that they were able to offer conjectures.

Conclusions:

Teachers have a range of conceptions about teaching, learning and computers. These conceptions appear to be derived from their experience and shape the ways in which they approach novel tasks such as developing a computer-based learning context. In particular, their conceptions of teaching range on a continuum between a focus on what the teacher does and the "content" to be taught towards a view that pays increasing attention to the students' intentions about the activity and the quality of interaction. Teachers in the sample were less clear about the nature of learning. Many felt unable to provide a response. More than two thirds of respondents described learning in behaviourist terms as a response by students to teacher actions. Only a small minority viewed learning in terms of the thinking, feeling, acting of students in an interactive social context.

Preliminary results support the hypothesis that beliefs and conceptions about key aspects of an activity shape thinking, learning and approaches to the design of computer based learning environments. To the extent that teachers' conceptions of teaching define their role as a transmitter of information and knowledge, they approach the design of computer-based environments with provision of information as a major goal.

Assessment of students responses is primarily viewed as ensuring they have encoded and are able to reproduce the "correct" ideas and techniques. To the extent that they view the role of a teacher as a manager of an environment which fosters active and meaningful participation by students in the process of making meaning, they focus on issues associated which the quality of the environment and the probable responses of the student-users.

References.

Crawford, K. P., (1986) Cognitive and Social Factors in Problem Solving Behaviour". In the Proceedings of the *Tenth Conference of the International Group for the Psychology of Mathematics Education,* London, July 1986, 415-412.

Crawford, K., Bornholt L & Summers, F.(1993) Learning in a Logo writer Environment: A hierarchy of students' conceptions of tasks. a paper presented at the *Annual Eurologo Conference,* Athens, Greece, August, pp 1-21

Crawford, K., Gordon, S., Nicholas, J and Prosser, M.(1993) Learning mathematics at university level, in W. Atweh (Ed.) *Contexts in Mathematics, The Proceedings of the Mathematics Education Research group of Australasia,* Annual Conference, Brisbane, July 1993, 209-214

Crawford, K., Groundwater-Smith, S., and Milan, M. (1990) *Gender and the Evolution of Computer Literacy.* revised research report to the N.S.W Ministry of Education published by the Government Printing Office.

Fugita, H. (1993) Principles in Organising University and Pre-University Mathematics Curricula For Scientists and Engineers. A paper presented at the South East Asian Congress on mathematics Education, June 7-11, Surabaya, Indonesia.

Marton, F. (1988). Describing and Improving Learning, in Schmeck, R. (Ed.), *Learning Strategies and Learning Styles,* pp 53–82. New York: Plenum Press.

Rogoff, B. Developing an Understanding of the Idea of Communities of Learners. *Mind, Culture and Activity,* vol. 1(4), 209-229.

Semenov, I. N. (1978) An empirical psychological study of thought processes in creative problem solving from the perspective of the theory of activity., Soviet Psychology, 16 (1), pp 3-46.

Speedy, G. (1989) Chair, *The Discipline Review of Teacher Education in Mathematics and Science,* Canberra, Government Printing Office.

Valsiner, J. (1994) Irreversibility of Time and the Construction of Historical Developmental Psychology. *Mind Culture and Activity,* Vol. 1(1), 25-42

Willis, S. (1990) Numeracy and society: The shifting ground. in S. Willis (Ed.) *Being Numerate: What Counts?,* ACER, Hawthorne, Victoria, Australia. pp 1-23

TECHNIQUE OF CHINESE LOGO USED IN MATH TEACHING IN SCHOOL

Fu Meiyu, Centre of educational developing and researching, Zunyi Guizhou

1. INTRODUCTION

Today, informational technology making rapid progress, more and more people have recognized the effect on the instruction made by computers. As basic instruction, the school math teaching, how to use computers properly according to the specialities of itself, to improve the effect of math teaching more effectively, and to train "creative" math wits suiting for the time's needs, have become problems concened extremely by whole math teachers. During ten years' math practical teaching, I have been devoting to CAI research on math. I have tried to carry on CAI of school math under circumstances of computer language such as BASIC, LOGO and Chinese LOGO. Also, I have used ready-made math CAI software of others. Finally, I found Chinese LOGO is more effective than the others.

2. CHINESE LOGO'S CHARACTERS

Chinese LOGO was developed by Zhang Wanzen, a senior engineer of Beijin University, in 1987. It's based on the MS LOGO created by Simon. Parbert, a mathematician of the AI laborary of MIT ofUSA. LOGO not only has the features of several kinds of excellent programming languages, but also is closer to nature languge, so it's easy to be accepted by nonspecialist on computer. LOGO's tortoise can drawlively picture and the modularity of program structure as well as the function of flexible numerical calculation, marvellous recursion, aboundant word and list processing and easy process planning and regulationmakes programming extremely convenient. Chinese LOGO remains all the function of MIT LOGO, which realized man-machine interaction with chinese in program, and it becomes very easy to add characters and chinese words freely. It's not only suit for the low age children's intellectual development but also for middle school student's studing psychology. It's not difficult for general school teacher of math and student to grasp the method of programming with Chinese LOGO.

3. APPLICATION OF CHINESE LOGO IN SCHOOL MATH TEACHING

The coordinate on LOGO screen is a plane system of rectangular coordinate which take the center of screen as grid origin. Under Chinese LOGO, we can set up various plane system of rectangular in different unit length , and mark corresponding letter and number in time.Through the shift of tortoise in the screen coordinate, the usual abstract and withered coordinate teaching is made imaginal, lively and concrete in the computer. This provide the teaching which emphasis and difficulty of the content combining data and form with advantages.

Computer technic is the symbol of moden informational scocial development,and math is the mother of computer. Using math to command computer is undoubtedly worth prouding. Chinese LOGO can make students in different degree enjoy commanding computer with math. First, let them learn some basic command and operating method of Chinese LOGO to link up man and computer.Then teach them to use math knowledge they learned to command computers' every action such as tortoise's drawing straight lines,rays,line segments,corners,number axises and marking letters according to the needs of math.Furthermore,teach them to learn the basic idea of LOGO programming, and using math to "creat" more LOGO procedure . In the benign cycle "applied mathematics conception→found math model → edit math program → run program → watch math changing rule → find new problem → deepen understanding the math conception",the students may find that as their learned math knowledge increases,their ability to program is continually improving and the skill of commanding computer is daily strengthening and the learning and grasping of new math knowledge is easier.All these are good for the cultivation of students' quality.

Since 1987,I have been engaging in the explore of CAI of school math teaching with Chinese LOGO.I successively made CAI experiment on the relevant content of such curriculum astrigonometric functions and analytic geometry and obtain encouraging results.

From 1991 to 1993,I made experiment on a common junior class (64 person) in NO.5 school of Zunyi for a math CAI period ,from Junior one to Junior three.This experiment proceeded on the CEC-I Teaching Network by the means that teacher guide students' studying to plan demonstration course of what is learned in class and make "experiment on math" in computer.The students of experimental class who learned basic drawing command and planing idea of LOGOat the stage of Junior 1 could program in LOGO and draw some basic geometry figures under the guidance of a teacher in junior 2.Drawing in computer needs such learned math knowledge relevant to the figures as geometry theorem,computing method and so on.For example,drawing aparallelogram needs the dicision theorem or nature theorem of collateral lines to determine the tortoise's corner degree; and its advance steps is determined to the rule "parallelogram's opposite side is equal". Finally,marking letters on the drawn figures made the students psychologically feel that the drawn geometry figures was perfect. Practice proved that when students draw geometry figures in Chinese LOGO in computer,they are in good mood and their impress is deep and the conception is grasped very well.Through the Teaching Network, teachers can teach various students in accordance with their aptitude conveniently.When I conducted the brief summary class on parallelogram,I demanded they used the procedure drawing parallelogram with parametersto draw all the special parallelogram on the same screen and mark responding names .Unexpectly,I saw through the network a student who was in bad base finished drawing the provided figures independently.Instantly , I gave praise to him and transmitted his figures through the network to the screens of all the students who congratulated him with warm applause.From his excited expression I found he had gained great encouragement,and it also urged other students laging behind. In the experimental class of comprehensive exercises about circle at the stage of Junior 3, I demanded students draw the common tangent of two circles in computer.According to previous experience, computing the length of it involve such knowledge on circle as nàture,tangent,collateral lines, Pythagoras's theorem and so on, so a few students feared to do this kind of exercises.Drawing the common tangent of two circles in LOGO in computer not only involved the above knowledge, but also related to such knowledge as coordinate setting and trigonometric functions for the location,run rout, direction and magnitude of corner and computing of advance step of tortoise should be considered. But it seemed that these students took great pains and were in good mood to do such work. A ordinaryly nauty boy was wholly absorbed in talking,thinking,operating and computing with his partner in computer and finished drawing according to requirement.He was glad to say: "well! well ! It's better than video games!"

Function teaching is the emphasis as well as difficulty of school math teaching. Under the guidance of teacher,the students set up the plane system of rectangular coordinates in unit length of :I through the procedure planned by themselves.And then ,they executed a procedure with parameters," TO D :X :Y :I", that drew a dot (:X,:Y) in the coordinates in unit length :I which was defined by the LOGO locating command SETXY :X :Y,and watched the one-one relation of correspondence of plane dot and ordered couple number on the screen.The lively demonstration left a deep impression on them and stengthened the understanding of the basic conceptions of this chapter.Under the guidance of the teachers,the students found that revising slightly the procedure D :X :Y :I made the procedure drawing a series of dots gotten.For instance the procedure TO D1 :X :A :Y :I is to draw the straight line Y=:A (:A is a constant).In contrast with the procedure drawing a dot , in the procedure D1, D :X :Y :I is called firstly to draw a dot (:X ,:A), and then the next dot (:X+1,:A) is drawn by the means of calling again D1 by itself through the recursive function of LOGO.From D1,we can find clearly that the position of variable :X is replaced by :X+1 but the constant :A is unchanged.The executing procedure D1 :X :A :I ,e.g ?D1 -100 35 1, is that the tortoise draw a series ofdots on the straight line Y=35 which is parallel to axis X in the plane system of rectangular coordinates in unit length 1, starting from X=-100 on the screen.The distributed density of dots depends on the data of changed :X+1.Through planing and running above procedure ,the students understood the conceptions of constant and variable easily.The teacher led the students to think what would be the result if the parameter :Y of procedure D1 :X :Y :I werenot costant :A but the expression on :X ? The students revised D1 immediately to watch and think.When they found three different kinds of interesting figures appearing separately as :Y=:X,:Y=1/:X,:Y=:X*:X, they were exercited and curious and eager to look up answer in the teaching material of math.It got twice the result with half the effort for the teaching of function conception. The followed procedure TO Y3X :X :I is the procedure drawing the dots of the figures of function Y=3X,in which :X inputs the horizontal ordinate of the figure's starting point. To run ?Y3X -30 1, e.g,is to draw the straight line Y=3X from X=-30 in the coordinates in unit length 1 and finally mark analytic expression on the drawn figure.Through such practice as programming,running and watching, the students got a deep impression of constant, variable,independent variable,dependent variable,relation of function and drawing the figuresof function according to its analytic expression.The following procedures refered to aboveare surely simple and clear.

```
TO D :X :Y :I                TO D1 :X :A :I           TO Y3X :X :I
  PU SETXY :X*:I :Y*:I         D  :X  :A :I             D :X 3*:X :I
  PD FD 1  BK 1                D1  :X+1 :A :I           Y3X :X+0.1 :I
END                          END                      END
```

After the students were able to use the recursive function of LOGO to program drawing function figures.the teacher guide them to use the function of word and list processing toinput the analytic expression of function as prototype.Then,it's convenient to watch the state that the figures varied with the parameters in its analytic expression.Under Chinese LOGO , the students marked corresponding expressions on the created figures in different parameters on the screen and sum up the changing law and the nature . In the quadric function teaching at the stage of Junior 3 the students watched the relation of the figures (paracurve) and the changing of coefficient a ,b,c of expression $y=ax^2+bx+c$ thoroughly in limited time.As the students participated in the programming , they knew the mothod by which computers drew figures according to their expressions is that they ordinaryly traced point drawing . So, they could understand the demonstration results more penetratingly, and the impression was deeper. Then,the teacher used the teaching material of demonstration to give questions by the means of man-machine interaction, and let the students seek expression rapidly or give expression to let them draw figures on the paper as soon as possible.The students competed with each other and tested through the computer's demonstration.To do so not only brought their enough studing initiative into play,but also strenghened their understanding the conceptions of math, developed the ability to answer questions,and improved the utilization ratio of limited teaching time.The students generally felt that it's not difficult to study the function.Those entering senior school among them even actively planned such demonstration programs as power function teaching in senior school for math teachers.

The practical problems of navigation, which relate to orientation in the math of Junior 3 usually made the students have headache.In computer, the students felt interesting touse the procedure planned by themselves that solutes triangle andsets up orientation to simulate navigation in Chinese LOGO and felt very interesting.Firstly,marked locations,submerged reefand so on according to the requierment on the screen . Then,took the tortoise as ship and the students as capitals to command it to advance obediently. It's clear at a glance whether the command was success or not.In the happy and relaxed practice, the students understood and grasped the solutions of this kind of problems very soon.During the course of Math CAI experiment in junior school, the students found while their knowledge of math was increasing, their abilities of programming in LOGO were enhanced day by day. For example,draw triangles in computer. In Junior 1,only could the regular triangle be drawn randomly,but after learned the Pythagoras's theorem and how to solute right triangles,right triangles and is osceles triangles could be drawn randomly.And then the procedure of drawing random triangles could be planned in computer to "solute triangles" freely after learned the law of sines and cosines.From programing practice, the students really understood the importance of studying math.

April 1993,the specialist of Guizhou teaching and research institute and math teachingcircles in the districtiion and the city of Zunyi made an appraisal of it and think "this experiment is scientific,practical and operatable.It's recommendable to spread." In the testof entering senior school in 1993, the experimental class got good results in math with the first grade on mean scores in the whole city of Zunyi.From the questionnaire to these students ,we can find such math CAI are generally welcomed by them.

THE STATISTICS OF THE RESULTS OF QUESTIONNAIRE ON MATH CAIEXPERIMENTAL CYCLING 1991-1993 OF NO.5 SCHOOL IN ZUNYI

The contents of questions	The distribution of those beginning in J1 on attitudes					The distribution of those beginning in J3 on attitudes				
	FA	BA	NA	FO	F	FA	BA	NA	FO	F
CAI deepen my understanding the conception of math.	.67	.33	0	0	0.835	.34	.50	.11	.05	0.485
I am intersted in studing programming in Chinese LOGO.	.60	.27	.13	0	0.67	.39	.50	.055	.055	0.5575
I think CAI enhanced my ability to analyse and solute math problems.	.50	.37	.13	0	0.62	.28	.33	.28	.11	0.195
I'm intersted in the teachers' demonstration teaching with teaching software.	.633	.233	.10	.033	0.6665	.39	.50	.055	.055	0.5575
I think that planning math program by myself make my understanding the conception of math deeper.	.53	.27	.17	.03	0.55	.28	.33	.22	.17	0.165
It's very intersting to do math exercises in computer.	.57	.27	.06	,10	0.575	.33	.615	.055	0	0.61
CAI lost studing time.	0	.13	.27	.60	−0.67	.05	0	.39	.56	−0.705
The parents rather support CAI.	.73	.17	.10	0	0.765	.61	.39	0	0	0.805
It's very pity that there is not computer room in the school and the operating time can't be assured.	.47	.23	.20	.10	0.385	.45	.39	.11	.05	0.54
There should be computer room in the school and operation at least once a week.	.633	.30	.067	0	0.7495	.67	.33	0	0	0.835
Hope that there will be CAI in the math studying in Senior school also.	.57	.43	0	0	0.785	.56	.33	.05	.05	0.65
Studing computer enhanced my ability to teach myself to some degree.	.37	.57	.067	0	0.6215	.22	.56	.17	.05	0.365

FA: full agree
BA: basically agree
NA: not agree very much
FO: full oppose
F : coefficient of attitude

FA: full aaree

| | | |

Of course,in every new contents of CAI ,the teacher's guidance is very important.The teachers should program the necessary demonstration of teaching in advance and design the procedure of teaching meticulously. They should lead the students to use learned math knowledge to plan and run programs , watch running results,discover new math problems and grasp new math conceptions in the computer teaching network . They should also monitor every student's operation and discover problems and feedback imformation in time.The teaching software of demonstration had better be excellent in the picture and its accompanying essay and have better interface ofman-machine interaction to observe and think. It's convenient to plan such demonstrating software in Chinese LOGO. Provided the teachers have good quality of teaching and the common skills of programing in Chinese LOGO, they can design the teaching materials of math CAI which are suited to current teaching practice according to their own teaching characters without more profound knowledge of computer.In the teaching practice, I have planned such materials for demonstration which form a complete set with the math materials in operation of Senior and Junior successfully in Chinese LOGO.The teaching materials of demonstration which contents are such as Senior's power function,exponential function,logarithmic function , trigonometric function,pictorial diagram of geometry of three dimensions,definition of conical section,coordinate translation, polar coordinates and parametric equation and so on ,Junior's basic conception of plane geometry, collateral lines, proving of Pythagoras's theorem, definition of circle,position relation of straight line and circle, position relation of circle and circle, rational numbers and number axis,plane systemof restangular coordinates, function and its figure, solving triangle, preliminary statistics and so on notably affected improving the quality of math teaching in practical class teaching.

4.CONCLUTION

The application of the technic of Chinese LOGO in math teaching of middle school can enables the students to participate in the pracice of math programming in person. It conforms to the congnitive law on the students' studying math. So it touches the teachers and students of middle school very much; it changes computers into good enviroment in which students go in for creative math intellectualism;it's conductive to developing student's math consciousness and enhancing their ability to solve problems with math.

Reference:

1."Retalk the math curriculum facing new century", Ersen Ding, Mathematical Journal,Beijin Normal University,Vol.2,1994.
2."The contents and the structure of computer curriculum", Cuanxin Nu & Nanyin Zu,Complication of Computer Teaching Materials of National School,1990.
3.Math Education,Dianzhou Zang,Renfen Tang & Hongkun Liu, Jiangxi Education press, 11, 1991.
4."The effect of the computer and imformation science on math or its teaching",Ersen Ding,Math Teaching,Normal University of East China,vol.5 & .6, 1991.
5."Inverstigation and research on the status quo of school computer application in Guangzhou,Department of Telelectures , Normal University of South China,vol.6,1988.
6.Math Education Research Guide, dianzhou zhang, Renfen Tang, Zaipin Dai & Siqi Li, Jiangxi Education press, 10,1994.

1st Asian Technology Conference in Mathematics
Nanyang Technological University, Singapore
18-21 Dec 1995

Teaching Linear Algebra using Matlab

Murli M. Gupta
Department of Mathematics
The George Washington University
Washington, DC 20052, USA
mmg@math.gwu.edu

Abstract

Two years ago, we introduced the Matlab software package in the teaching of our elementary course in linear algebra. Prior to this date our course was the traditional mathematics course with the chalk and talk being the dominant format. In this paper, we describe our experiment and the success in making Matlab an integral part of the course. We describe the organization of the course and present some of the non-traditional features of the course.

Introduction

The teaching of linear algebra at the introductory level has always been a traditional exercise in mathematics instruction. The usual method has been to consider linear systems and matrices of order 3x3 or 4x4 because that is the size which is conveniently handled in classroom settings with all of the work being done using pencil and paper.

In the last several years there have been attempts, worldwide, to introduce technology in the teaching of many of the mathematics courses including precalculus, calculus, linear algebra, differential equations. Technologies being introduced into the teaching of these courses range from the graphic calculators such as HP48 and TI85, computer algebra systems (CAS) such as Maple, Mathematica, Derive and Macsyma, spreadsheet packages such as Lotus 1-2-3, Quattro Pro and Microsoft Excel, and the Matrix Algebra package Matlab. Many reports of success with the introduction of technology have been reported in the recent years.

We describe how we introduced Matlab in the teaching of linear algebra, the challenges we faced, and how the course was modified through the student feedback.

Planning and Operation of the Computer Lab

In the spring of 1993, the author and a colleague initiated this experiment and during the summer of 1993, we attended a week long ATLAST workshop which is an NSF-ILAS project to Augment the Teaching of Linear Algebra through the use of Software Tools. At the ATLAST workshop we were exposed to the various ways Matlab can be introduced in the teaching of linear algebra; we also obtained a variety of source material for use in our classroom. Through group discussions, we were able to learn about the logistic, academic and other difficulties that were faced by others.

We proposed to the administration of our University that we wanted to introduce Matlab into the teaching of linear algebra. Our linear algebra course is a traditional 3 credit hour course that usually meets twice weekly (75 minute classes) for 14 weeks. We proposed to acquire a limited site license for the use of Matlab in one of our computer classrooms that hold 18-30 computers; the computers are networked and are connected to internet. We were granted funds to acquire Matlab to install on the University computers and also were allowed to work as a team to teach the particular course in Fall 1993 semester. (Normally, only one faculty member teaches a course.) The two of us took turns in teaching this course and attended all classes, even while the other was teaching. In addition, we instituted a lab portion of the course that met for one hour a week for alternate weeks in the computer classroom. Both of us (the faculty members) attended the lab sessions where we introduced the students to the use of Matlab and guided them through the nuances of the DOS machines as well. We also made arrangements for the students to be able to use the machines at other times when the lab was not being used for any other purposes.

At the mid-semester point we asked our students to provide feedback on the introduction of technology to this course. Typical responses were:

> Matlab helped to reinforce many concepts presented in class. It had the advantage of allowing us to test functions very quickly and without error. The time Matlab saved in doing calculations greatly helped me focus on learning the concepts rather than getting bogged down in the basic math.
>
> I think Matlab was very helpful tool in understanding the material. I found it to be a worthwhile exercise.
>
> I enjoyed the computer component because it took the full advantage of the technology at hand. Furthermore, I was able to immerse myself in matrices.
>
> Matlab assignments were probably most effective when like problem number 1 of this assignment. We're given something to figure out that takes a few steps, and the computer is used as a tool to get the answer, yet the problem still requires thought to set up.
>
> Matlab was a big aid in understanding the more complex aspects of matrix algebra because it enabled me to concentrate on the theories and their

> proofs, rather than worrying about the computations. The lab furnished an easy, accessible calculator to matrices that saved an abundance of time that could go to grasping the ideas rather than being used in tedious computations.

The criticisms about the course related to the unavailability of the computer lab on the weekends, the lack of a teaching assistant, the lack of manuals for the DOS machines and Matlab, and the general unfamiliarity of some of our students with the use of personal computers. Many of these concerns were resolved in the subsequent semesters when we modified the course to make Matlab laboratory an integral part of the course. Starting with Spring 1994 semester, we instituted weekly labs for one hour each week for the whole semester and also hired a graduate student to help in the lab as well as to provide consultations to the students. With these modifications in place, most of the problems we experienced in our first semester went away. During the last semester just completed (Spring 1995), we heard nothing but praise for the Matlab component of the course. The students seem to be liking the lab component of the course and did not have many suggestions to improve the operation of the course. The typical comments were:

> I like Matlab. It appears to be powerful software and makes the problems much more easy and interesting.

> I like working on the computer. The lab helps illustrate what we have gone over in the class.

The end of semester evaluations (anonymous) continue to indicate that most of the problems with the operation of the lab have been resolved and the students are satisfied with the operation of the course.

Course Contents

In the previous years, we used the texts such as *Elementary Linear Algebra* by Roland Larson and Bruce Edwards (Heath, 1988) and covered the standard topics such as systems of linear equations, matrix algebra and determinants, vector spaces and subspaces, eigenvalues and eigenvectors. While we pointed out the variety of applications, listed in the textbook or otherwise, there usually was no time available to devote to the applications of substance. With the permanent introduction of computer laboratory, we used the text *Linear Algebra and its Applications* by David Lay (Addison Wesley, 1994) with supplementary material from *Linear Algebra Labs with Matlab* by David Hill and David Zitarelli (Macmillan, 1994). We now cover all of the topics previously covered, often in more depth, and are also able to introduce a variety of applications such as least squares curve fitting (curves of order 1, 2 and 3), power method for computing the dominant eigenvalue of a matrix, stochastic matrices and Markov chains, dynamical systems, and iterative solution of linear systems. Some of the mid-semester feedback in

Spring 1995 semester related to the availability of applications:

> I like the fact that there are many real world applications of the material we are learning. The computer lab in particular drives this fact home.
>
> I like seeing where the material applies to the real world. Some of the problems in the book have given some sense of what the material can be used for.

Evaluation and Assessment

Any curriculum innovation is accompanied with the usual skepticism from the colleagues and administrators. We are happy to note that our colleagues were generally supportive of our efforts to introduce Matlab in the teaching of linear algebra. The questions still arise as to whether the students being taught with the new system are better prepared compared to the students going through with the traditional system. Such questions can be answered in several ways.

In our course, we give a number of take home assignments where the students are supposed to work on non-trivial exercises and are required to explain and interpret various steps involved in problem solving. (An example of a take home assignment from Spring 1995 is enclosed as Appendix A. The curve fitting to the data on the number of doctorates awarded to U.S. citizens provided many interesting comments from the students, especially when the predicted numbers for the future years were way out of line of an educated guess. The average score on this assignment was 87%.) In addition, we often ask the students to submit their lab reports, just as is done in chemistry or biology labs. This allows the teaching assistant and the professor to provide appropriate feedback to the students and also provides opportunities for individual counselling sessions with the individual students. The difficulties with the hardware/software can also be identified at an early stage and corrective measures taken in time.

We give traditional closed-book mid-term and final examinations where the students are not allowed to consult with any books or notes. In these examinations we are able to ask questions which we would not have asked in the previous years, before the introduction of Matlab to our classroom. For example, our final exams for the past two semesters included a question on finding the basis for null, row and column spaces for a 4x5 matrix whereas the exams for the previous years included similar questions for 3x3 or 3x4 matrices. The difference consists in the fact that we are now able to give questions of increased complexity and the students are able to do well on such questions because they are learning in depth and have a better retention. (At the Spring 1995 final examination, 10 out of 18 students scored 90% or better on this question.)

Some of the student feedback during the mid-semester indicated a strong desire to be tested in the computer lab. Typical comments were:

> Since we do most of our homework in matlab, why don't we take our test there?
>
> If you are going to continue to use the computer-give exams using the computer.
>
> I think an exam on the laboratory part of the class would be a good idea.

Partly in response to these comments, we have started giving our students a one hour test in the computer room. The Matlab test is an open-book test where the students are allowed to bring any material to the test; they are not allowed to consult with one another. The problems on the Matlab test are non-trivial and test the students' knowledge of the subject matter as well their facility with the computer lab. (A sample test from Spring 1995 is enclosed as Appendix B.) This test has become very successful and the average score in Spring 1995 was 74.6%, with 10 out of 18 students obtaining 80% or better score.

The students' overall performance also seems to have improved since the introduction of computer laboratory to our elementary linear algebra course. The average class score for the last two semesters (Fall 1994 and Spring 1995) was close to 83% whereas in the previous semesters the average class score hovered around 72%. In Spring 1995, 12 students out of a class of 18 obtained grades of A or B; in Fall 1994, 4 students out of a class of 9 obtained the grades of A or B. In contrast, in Fall 1993 which was our introductory semester with the new technology, only 9 students out of a class of 23 earned grades of A or B. This is an indication of our students' deeper and more thorough learning of the subject matter.

Future Plans

We have made Matlab laboratory an integral part of our introductory course in linear algebra and need no longer justify the use of the computer laboratory. We are considering acquiring the new releases of Matlab that have a Maple interface and allows symbolic manipulations of matrices. We are also introducing the symbolic software package Maple to various courses in calculus and differential equations.

Conclusions

We have successfully introduced Matlab computer laboratory to our elementary course in linear algebra and have discovered that we are able to give broader and deeper coverage to the material previously taught in this course. We are also able to introduce a variety of applications that could not be introduced in previous years. The student response has been very receptive and the student performance indicates that the introduction of technology has allowed our students to get a more thorough, even more rigorous, understanding of the subject matter.

Math 124-10 (Spring 1995)
Assignment #4

April 26, 1995
Due Date: Wednesday May 3, 1995

Problem 1: Use Gram Schmidt process (as described on page 362 of your text) to obtain an orthogonal basis for the space spanned by the following set of 3 vectors. Show all details of the process.

$$\mathbf{v}_1 = \begin{pmatrix} -1 \\ 2 \\ 0 \\ 1 \end{pmatrix} \qquad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} \qquad \mathbf{v}_3 = \begin{pmatrix} 2 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

Problem 2: The following table lists the number of doctorates awarded annually in mathematics to U.S. citizens during the years 1983 to 1992.

Year	No of doctorates
83	455
84	433
85	396
86	386
87	362
88	363
89	411
90	401
91	461
92	430

(a) Fit a least squares line to the given data and use this line to estimate (predict) the number of doctorates to be granted in 1994, 1995, 1996 and 2000.

(b) Fit a least squares polynomial of degree 2 to the given data and use this polynomial to estimate (predict) the number of doctorates to be granted in 1994, 1995, 1996 and 2000.

(c) What can you say about the quality of the above predictions?

Math 124-10 (Spring 1995)
Matlab Test (May 1, 1995)

Please submit the printout with detailed comments and explanations as appropriate.

Problem 1: Consider the following matrices

(I) $A = \begin{bmatrix} 3 & -1 & 0 & 2 & 1 \\ 2 & 5 & 2 & 1 & -1 \\ 1 & 2 & 3 & 0 & 1 \\ 1 & 2 & -1 & 4 & 3 \end{bmatrix}$ (II) $A = \begin{bmatrix} 1 & 2 & 3 & -1 & 1 \\ 2 & 1 & 1 & 1 & 3 \\ 1 & 1 & 3 & 2 & 1 \end{bmatrix}$

(a) Are the columns of A linearly independent?

(b) Are the rows of A linearly independent?

(c) Find a basis for the column space of A.

(d) Find a basis for the row space of A.

(e) Find a basis for the null space of A.

Problem 2: Consider the following matrices

(I) $A = \begin{bmatrix} 3 & -1 & 0 & 2 \\ 2 & 5 & 2 & 1 \\ 1 & 2 & 3 & 0 \\ 1 & 2 & -1 & 4 \end{bmatrix}$ (II) $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 1 \\ 1 & 1 & 3 \end{bmatrix}$

(a) Find the characteristic polynomial and eigenvalues of A.

(b) Show that the set of eigenvectors of A is linearly independent.

(c) Show that A is diagonalizable?

(d) Use the diagonalized form of A to compute A^3, A^5.

Problem 3: $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 1 \\ 1 & 1 & 3 \end{bmatrix}$

Use the **power method** to estimate the dominant eigenvalue of A.

Problem 4: Consider the following sets

$U = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 3 & 1 \\ 2 & 2 & 2 \\ 0 & -2 & 1 \end{bmatrix}$ $V = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & -2 \\ 2 & 2 & 0 \\ 1 & -1 & 3 \end{bmatrix}$

Show that the sets U and V span the same subspace of R^4.

EVALUATION OF A COMPUTER-BASED CALCULUS COURSE

Ruth Hubbard
Queensland University of Technology
GPO Box 2434
Brisbane 4001
AUSTRALIA
r.hubbard@qut.edu.au

A large cohort of students studied an introductory calculus course using DERIVE tutorials for one semester. The impact of the trial on the department, on the members of the department who acted as tutors and graders and most importantly on the students themselves was monitored and evaluated. The course was loosely based on some of the reformed calculus courses in the United States but the students were not volunteers, there was no alternative course. As expected, some of the people involved had some reservations about this innovation. However we learned a great deal, not only about this particular course, but also about the effects of implementing change in a university environment. Our experiences should be valuable for anyone contemplating similar changes.

Implementing change

There are many obstacles to implementing change in a discipline such as mathematics with a long established teaching tradition at university. Some of the obstacles are institutional such as design of classrooms, scheduling of classes, prescribed contact hours and methods of assessment. Other problems arise from the beliefs and attitudes towards teaching and learning mathematics of the teachers and students. Teachers tend to teach the way they themselves were taught and student coming to university already have preconceptions about mathematics teaching and learning gained from their experiences during 12 years of schooling. Any substantial teaching innovation is certain to attract criticism because it challenges established practice. The criticisms of academic staff of projects to foster independent learning were neatly classified by Cornwall (see [C], page 198) and apply equally well to this project. Therefore an important purpose of our project was to change some beliefs about mathematics teaching and to evaluate that change.

The project followed a trial with a small group of student volunteers in the previous year* which succeeded in some respects but also alerted us to some problems about implementing change. It is very difficult to evaluate a project which relies on student volunteers and to extrapolate the results to the whole population. In addition a small trial with volunteer students often has little impact on a mathematics department.

* This project was funded by a QUT Teaching and Learning Development Large Grant.

In view of these limitations we carried out this project* in a large heterogeneous class (230) and invited the Head of School to nominate ten academic staff members who could contribute to the project and might learn from it, to act as tutors. We realized there was a risk in this approach, because volunteer tutors would have been more sympathetic to the aims of the project but on the other hand, it is standard practice for the Head to appoint staff as tutors.

Major objectives of the project

Traditionally many students in first year calculus courses concentrate on memorizing processes and imitating examples without gaining very much understanding. This may be sufficient for passing standard examinations but does not prepare students very well for applying calculus to problems in the future. In addition, for some years, computers have been able to answer standard examination questions (much better than the students) (see [H], page 46) so some mathematicians (see [R], page 27) have begun to question the value of the traditional approach to teaching mathematics.

In this project we wanted to show that it is possible to develop a course which gives priority to the understanding of concepts, and uses the computer as an investigative tool in the concept development process. Some courses with similar objectives exist in the US but in teaching environments very different from those in Australia. In Australian universities, first year calculus students often attend lectures in groups of 200 or more. In the US the students are usually divided into sections of 30 or so and taught by a large number of teaching assistants. We therefore also wanted to show that it is possible to implement such a course, not with a small group of students but with a typical, large, heterogeneous class. To demonstrate that it was possible, we planned to carry out a detailed evaluation of the impact of such a course on the students and teachers involved.

In order to improve student understanding we used the following strategies:

- investigations of the relationships between mathematical objects using the DERIVE computer algebra package, to give the students opportunities to explore mathematical ideas before they were formally defined in lectures
- an emphasis on correct use of notation
- development of mathematical ideas from real world examples
- linking the graphical, numerical and symbolic representations of concepts
- using groups in the lab classes so that students could develop their ideas through discussion with their peers
- encouraging students to read the text (see [HG]) as part of their learning
- asking students to write group reports about their activities in the lab classes in which they had to explain the results of their investigations in clear English

** This project was funded by a grant from the Committee for the Advancement of University Teaching.

In accordance with these objectives, the method of assessing student learning also had to change. We used some standard examination questions but in addition we asked students questions which tested their understanding of concepts. For example, we asked them to explain why they chose certain methods and how different aspects of a problem were related.

We hoped to be able to show that students and staff would be able to adapt to this approach to learning, that the students would become convinced that their understanding of mathematics had increased and that their performance in examinations would demonstrate some improvement in learning.

Method of evaluation

The following methods were used to gather information about the project.

- Students completed surveys on their attitudes towards and beliefs about mathematics during the first week and towards the end of the project.
- Students completed a brief survey of their mathematical knowledge during the first week, two class tests and a final examination.
- Students were asked to write something about the usefulness of the labs for their learning during a lab session.
- In week 9 students respnded to a questionnaire about aspects of the course - the text, study guides, solutions to exercises, group work, lectures and tutorial assistance.
- A random sample of 20 students were interviewed individually three months after the conclusion of the project.
- All the tutors were interviewed in small groups, two months after the conclusion of the project.

Impact on students - attitudes and beliefs

A number of identical questions, about attitudes to and beliefs about mathematics, were used in the surveys at the beginning of the project and at the end. The first survey also asked about the students' experiences with mathematics at school. The statements that were repeated were:

1. I like mathematics more than other subjects.
2. In mathematics something is either right or wrong.
3. In mathematics you can be creative and discover things for yourself.
4. Real mathematics problems can be solved by common sense instead of the mathematical rules.
5. To solve mathematics problems you have to be taught the right method or you can't do anything.
6. The best way to do well in mathematics is to remember all the rules.
7. Mathematical problem solving is important for everyday life.
8. Mathematical problem solving is important in professional careers.
9. If I have the option, I will take other mathematics subjects subsequently.

Unfortunately, in both surveys, where students had to choose a response on a 4-point scale - very true, sort of true, not very true, not at all true, a majority chose the safe option, "sort of true" for almost all the questions. As a result it was very difficult to infer any changes from this information. It may also be the case that 14 weeks is too short a period to expect long-held beliefs to change significantly. The two statements that did show a change in attitude, were 5. and 6., where more students disagreed with the statements in the second survey.

Some more interesting results emerged from comparisons of the responses to the open-ended questions. In response to the question, "How do you know whether you understand something in mathematics?" there was a shift from comments about "being able to do the problems and getting the right answer" to "the ability to apply the concept to different situations" and "the ability to explain the concept to another person". The latter comments demonstrate the need for a deeper level of understanding than the former.

The open-ended questions also showed that the importance to the students of memorizing in learning mathematics declined during the semester. However when some students were interviewed about how they studied for the examination, they tended to say that they relied heavily on memorization. This suggests again that old habits are not easily dislodged.

In the second half of the semester students were surveyed about how various aspects of the course contributed to their learning. Some relevant responses were as follows.

	Percentages		
	Usually	Sometimes	Never
Do the lectures help you understand the subject?	49	46	5
Does the group work in the labs help you to learn	45	43	12
Do you find the study guides useful	62	35	3
Are the solutions to the exercises useful?	69	30	1
Is the text easy to read?	38	49	13

We also asked students if they thought they would understand the material without any lectures. Only 13% said yes, while 43% said maybe and 44% said no. Of course this was a hypothetical question because most students would have no experience in learning mathematics without a teacher.

Since our objectives place a great deal of importance on various aspects of the computer laboratories, we asked the students, during a lab class, to write something about the usefulness of the labs for their learning. The comments were classified as positive, negative and mixed in the various lab classes and in total. For the whole group (n = 225) the comments were:

Positive	Negative	Mixed	No comment
46%	12%	20%	21%

These results were not uniform throughout the groups suggesting that the approach and ability of the tutor influenced the students learning in the labs. The negative comments had three main themes:

- that the computer work was a time-consuming distraction from learning real mathematics
- that by having the computer do certain things the student wasn't learning to do them him/herself
- that it would be less confusing if the labs followed the lectures

The last comment suggests that some students were uncomfortable with the cognitive conflict which is sometimes necessary before learning can occur.

The main themes of the positive comments were:

- we have to discover rules and patterns by ourselves which helps us think about mathematical ideas
- helps us to visualize what the rules we are learning look like
- puts theory into practice, shows practical solutions
- the computer saves time on repetitive and tedious tasks

Some examples of comments which reflect our objectives are:

> "Writing reports is when you think the most and this helps you understand the work."
>
> "Labs are the most exhilarating experience I've ever had."
>
> "The lab classes give me the opportunity to construct the material that is being taught, instead of just scribbling it down off an overhead projector and never learning it again."

The students who made mixed comments showed they were carefully assessing the advantages and disadvantages of the new approach to learning.

A random sample of 20 students were interviewed during the second semester. The interviews were delayed because we wanted to find out how the students thought they were managing the next mathematics subject. However this did mean that some of their recollections about our project were becoming vague. From the interviews we found that in general the more able students were more enthusiastic about the new approach to learning than the weakest students. However by the time they were interviewed, the students knew their grades and this may have influenced their responses. The weak students would have prefered a tutor to work exercises for them. Of course this is just how they were taught at school and it did not appear to have helped them very much. We found that although many students adopted higher level learning strategies during the semester, when it came to studying for the examination, they reverted to memorization just to make sure. Most of the students at all levels claimed to have obtained some benefit from group work, even when the groups were not of uniform ability. Only two students reported difficulties with the DERIVE package, even though a number had little previous computer experience. All but one of the 10 students who were studying the next mathematics unit felt they were well prepared and were coping satisfactorily. This was an important result as most of the tutors were very sceptical about the ability of this course to prepare students for further study of mathematics.

Some mathematical outcomes

Although we had a very high response rate to the attitude surveys, in the survey of mathematical knowledge in Week 1, many students did not attempt many of the questions on topics such as the rules of logarithms or trigonometric functions. However almost all students did attempt the graphical question below.

An object is moving along a straight line. In Figure 1, the displacement s, of the object is graphed against time t.

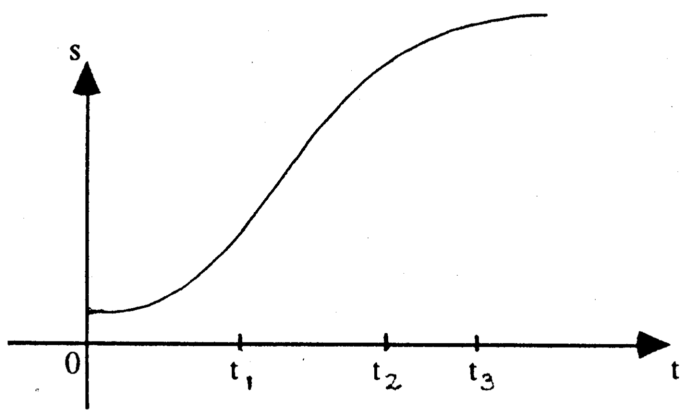


Figure 1

(a) Is the object moving faster at time t_1 or at time t_2 ? How do you know?

(b) What is the velocity of the object when $t = 0$?
How do you know?

(c) During the time interval from t_2 to t_3, is the object speeding up or slowing down? How do you know?

This question was repeated in a test towards the end of the semester with the following percentages of students giving a correct answer.

Part	Week 1 (n = 218) %	Week 11 (n = 207) %
(a)	52	86
(b)	33	66
(c)	70	90

An answer was marked correct **only** if the reason was correct, because in Week 1, many students gave the correct answer for the wrong reason. For example in part (a), they said that the object was moving faster at time t_2 because the graph was **higher** at t_2, not that it was **steeper**. The results show some improvement but part (b) is clearly a problem. The problem was that instead of basing their answer on the graph, students attempted to use a formula, even though there was no data to substitute into a formula. Unfortunately, the formula they chose was $v = s/t$, a formula which they had learned a long time ago and, which used properly, had produced correct answers. This is an example of a common problem (see [V] page 80) where students develop a concept image of some mathematical entity, in this case average velocity, and continue to use this concept image instead of modifying it to accomodate instantaneous velocity.

A different kind of formula problem occured in one of the examination questions, which we hoped the students would answer using integration. However 34% of all students correctly identified the situation and wrote down the formula $s = ut + \frac{1}{2}at^2$. Only 24% of these students were able to apply the formula correctly to the problem. The habit of responding to mathematical problems by writing down a poorly understood formula is very difficult to modify.

Another aspect of mathematics learning which we adressed in the project was the use of notation. In pencil and paper work, the teacher can frequently understand what a student is trying to communicate, even if the notation is not quite correct. However to enter mathematical statements into a computer package the student's use of notation must be perfect. There were two lab sessions on integration in which sigma notation featured prominently and although the students claimed that they understood sigma notation, they experienced great difficulty entering their data into DERIVE. In the examination students were asked to calculate

$$\sum_{i=0}^{4} f(x_i)\,\Delta x$$

from a given function and a table of its values at given points. 27% of students added either 4 or 6 terms even though in DERIVE these parameters must be specified and the terms appear on the screen before the summation occurs. A further 21% of the students were not able to correctly identify Δx with the given increments in x. This is not surprising because the examination question asked them to interpret the notation precisely instead of asking them to carry out a routine procedure. However the results indicate that the students grasp of the notation was far from complete. This raises a very important issue - if you always ask direct questions, questions which the students have practised, it is difficult to locate weaknesses in their understanding.

An even more serious problem was identified in the lab sessions on integration and elsewhere. Given a function f, a number of students were unable to identify the ordinate to the curve at b with f(b). The question must be asked - how much calculus is meaningful to students who cannot attach any meaning to the symbol f(b)? Once discovered, this problem was addressed in lectures and exercises and a very direct question about the concept was set in the examination. Even so, only 72% of students were able to correctly mark a length represented by f(b) - f(a) on a given diagram.

Examination results

The overall pass rate was 83%, which is higher than usual for this subject but only 49 of the 176 students who passed the examination elected to enrol in the next mathematics unit in the following semester. This was disappointing but the interviews suggested that a majority of students were enrolled in programs that prescribe just one mathematics unit. However the students who did continue to study mathematics showed that the project did in fact prepare them well to continue their studies. The pass rate for these 49 students, in the following unit, taught in a traditional manner, was 82% whereas over a number of recent semesters the pass rate has been 70% or less.

Impact on the School of Mathematics

Some staff of the School were opposed to a project of this nature from the outset and made no secret of their objections to its existence. Most reformed calculus projects have encountered this kind of opposition in their own universities. Many staff feel threatened when what they have been doing for many years and what they believe is the best way to do it is challenged. They hold these beliefs even though they acknowledge that conventional methods of instruction have not succeeded in developing an understanding of mathematical concepts in a majority of students. Nevertheless all the staff nominated by the Head of School to act as tutors agreed to attend a two-day training session and to do what was required of them as tutors. In addition they all worked hard and tried to help the students as much as possible although they were working in an unfamiliar environment.

The tutors' main concerns were that such a course could not adequately prepare students to study more mathematics, that an emphasis on symbol manipulation was of prime importance in learning mathematics and that computer investigations keep the students interested but waste time. None of the tutors subscribed to the constructivist view of learning on which the project was based, although this was discussed with them during the initial training course. However the tutors all reported that students attended the lab classes regularly, worked very hard on the assignments, asked a lot of questions and that most of the students appeared to enjoy the laboratory work. In fact they thought that the students liked the lab work so much that they concentrated on this and neglected paper and pencil exercises.

The tutors' more personal concerns were that marking students' written, as opposed to symbolic work, was time-consuming and that it was difficult to mark objectively. This is a particular concern for mathematicians, who normally only ask questions that involve symbolic answers, which are relatively easy to mark. In virtually all other disciplines marking essays is a fact of life. No doubt with more practice on the part of the tutors, this problem could be overcome.

The Head of School was supportive of the project but has two major concerns regarding its implementation in the future. Firstly, laboratory-based courses use more resources, both human resources and equipment than traditional mathematics courses. If all aspects of this project were implemented in all our courses we would require substantially more funding. Secondly, he is concerned about the difficulty many members of staff have adapting to a new teaching environment and in particular to making use of technology in their teaching. Both of these issues, the need for additional funding and the reluctance of staff to change are fundamental unresolved problems for projects which aim to improve teaching and learning.

References:

[C] Cornwall, M., Putting it into Practice: Promoting Independent Learning in a Traditional Institution, in D. Boud (ed), *Student Autonomy in Learning*, Kogan Page, 1981.

[H] Heid, M.K., Calculus with muMath: Implications for Curriculum Reform, The Computing Teacher, Nov. 1983.

[HG] Hughes-Hallett, D. and Gleason, A.M., *Calculus*, Wiley, 1992.

[R] Rothery, A., Re-organising introductory calculus using computer algebra, *International DERIVE Journal,* Vol. 1 No. 3, 1994.

[V] Vinner, S., The Role of definitions in the Teaching and Learning mathematics, in D. Tall (ed) *Advanced Mathematical Thinking*, Kluwer, 1991.

A MATHEMATICAL MODEL OF EVALUATION FOR LEVEL OF TEACHING IN CLASSROOM

Jin JingPing

(Jilin Teacher's College, Jilin China)

Key words: evaluation, level of teaching, matrix, target factor

Abstract In this paper, we give a mathematical model of evaluation for level of teaching in classroom according to the theory of evaluation and mathematical method.

An important task of modern education is to cultivate teachers and improve their level of teaching in classroom. So how to evaluate the level of teaching which deserves to be studied is an important problem.

In this paper, we give a mathematical model of evaluation for level of teaching in classroom according to theory of evaluation and mathematical method.

一、The basis of setting up the model

The evaluation for level of teaching in classroom is one of the categories of evaluation of education. The evaluation of teaching is the judgement of evaluation of teaching actioities by using scientific methods according to the object of teaching. So the theory of this mathematical model of evaluation for level of teaching in classroom is on the base of evaluation theory of education and the rule of teaching in classroom which is clarified by the teaching theory. Moreover it is the key of setting up the model that how to select the teaching method of evaluation rightly.

According to the evaluation theory of education and evaluation practice of teaching, how to set up a scientific evaluation target system can guarantee to realize the scientific evaluation. The sub-target of evaluation target is also called target. The target embodies a concentrated reflection of the nature and character of evaluation target. The target system refers to the target group which is formed by the discomposed and then transformed evaluation target. A target system includes not only several targets but also the portion of each targets (i. e. the portion in the whole target system). The rules in setting up a target system are:

Leading feature: The particular content of target system and the assignment to portion both are the demand for evaluation object and very expression of evaluation of leading function. When you are formulating the target system, you must pay attention to apply this rule rightly in order to develope the rule well.

Compatibility: The targets can not contradict one another in the target system. Other wise the evaluation function of target system can not be elaborated.

Integration: Evaluation targets is a system. Targets differentiate and contact one another. All the targets must stress the nature of the evaluation target. The sum of all of the targets is 1.

Objectivety: The target system should found on the fact by means of following to the objective law in the evaluation field. Then we can achieve the expected goal of evaluation.

Simplicity: Every target of the target system should be convenient and practical so as to be easy to evaluate. So we should grasp the main target factors and lost some non-important secondary factors. Don't be complicated.

According to the laws metioned before we still need to combine the features of teaching in classroom with the evaluation for level of teaching in classroom together to follow the objective laws in the field of evaluation. To analysize on the view of the theory of system, the process of teaching in classroom is a control system which deliver the information purposely and orderly. Its main characters are.

Purpose character: According to the subject teaching programm, the teaching object and demand of each chapter or class must be clear. The process of teaching in classroom is the process of realizing the teaching object.

Character of plan: According to the scheduled teaching object, the teacher must care for the learners' psychological characters when they are learning. Then the teaching pan (period) can by worked out at length.

Character of time: The teaching time of each class is fixed strictly, it can neither be extended nor curtailed. The teacher must guid the students to make use of every classroom minute efficiently according to the teaching plan.

Character of practice: Teaching in classroom is a process of practical activities. The teacher plays the role of guidance and the students are the studying body. So it is called bilateral activities between the students and the teachers.

Character of integration/union: The process of teaching in the classroom is the good combination of teaching and learning. It is also the union of mind education、knowledge teaching、skill training, abilities improving and educating object.

Following the characters and objective laws we mentioned before we give a mathematical model of education for level of teaching in classroom according to theory of evaluation and experience of evaluation activities, which include five target factors: content of teaching; construction of teaching, method of teaching; innate of teacher; emotion of teacher. The concrete demands of each target factors are explained as follows:

Content of teaching: According to the teaching programm the teacher must teach the content of teaching like these: make the teaching object clear; properly deal with the teaching book; correctly explain knowledge; stress the important part; break the difficulties; grasp the key; discover the initial law of teaching content.

Constrution of teaching: According to the scheduled teaching object and demand, the teacher must choose the appropriate class type, moreover design the proper teaching canstruction. The teaching steps must be clear and sysmatic. The content of teaching in classroom is due to the teaching time. The teaching time must be arranged reasonably.

Method of teaching: According to the object and content of teaching and studying psycological character. The teacher must adapt proper teaching method such as elicitation method to cultivate the students' interest and desire to learn. Pay attention to improve teaching method and bring forth new ideas to it.

innate of teacher: It embodies the teacher's inillectual factors in teaching in the classroom, which is reflected by oral expression, written expression, teaching organization and reaction etc. The teacher's teaching language must be accurate, concise, simple, clear and vivid; The written language must be standard and neat; The teacher must speak standard language i. e. Putongtlua. And the teacher must be capable and fexible to organize teaching in the classroom.

emotion of teacher: This is reflection of the teacher's non-intellectual factors in the teaching activities. The teacher must be in full responsibility and warm-hearted to treat the teaching in the classroom. They must be inspiring, mature, kind, dignified as a model.

The five demands can be made up or stressed properly according to different level and type of school. Fox example, in a cllege we can demand addly that the teacher must introduce the latest academic activity or achievement in scientific research in their content of teaching. In a secondary school we can give higher demand to the teacher in their standard language and handwriting which belongs to the innate of teacher.

二、The method of setting up the model

There is no accurate quantitative limits to evaluate the teacher's teaching level. So you had better adapt the synthetical judgement of fuzzy mathematics to evaluate the fuzzy conception which is the method used to set up this methmatical model. It can model the spycholoyical process of teaching evaluation somewhat distinctly and it can deal with some problem meticulously and concisly. Now it is used widely and very objectively.

三、model of evaluation

Let set U of taget factor and set V of evaluation grade:

$$U = (content, construction, method, innate, emotion)$$

(1) (2) (3) (4) (5)

$$V = (excellent, good, middle, pass, bad)$$

(1) (2) (3) (4)(5)

In set U of target factor, we decide distribution of proportion coefficient of varied factor as follows:

$$T_1 = (a_1, a_2, a_3, a_4, a_5)$$

$$(\sum_{i=1}^{5} a_i = 1, 0 \langle a_i \langle 1, i = 1,2,3,4,5)$$

First, we give evaluation of a factor, Each of the specialists, teacheres and students in classroom fill in a blank form as follows:

V U	(1)	(2)	(3)	(4)	(5)
(1)					
(2)					
(3)					
(4)					
(5)					

We operate the resule of table for the specialists, teachers and students in classroom respeclively.

Let the number of speciaists are A and that of teacheres are B and that of students are C, we replace each number of statistical result by ratios of the each number and number of evaluater A, B and C respeclively. Then we obtain three matrix of evaluation a factor R_1, R_2 and R_3:

$$R_1 = \begin{bmatrix} \frac{b_{11}}{A} & \frac{b_{12}}{A} & \frac{b_{13}}{A} & \frac{b_{14}}{A} & \frac{b_{15}}{A} \\ \frac{b_{21}}{A} & \frac{b_{22}}{A} & \frac{b_{23}}{A} & \frac{b_{24}}{A} & \frac{b_{25}}{A} \\ \frac{b_{31}}{A} & \frac{b_{32}}{A} & \frac{b_{33}}{A} & \frac{b_{34}}{A} & \frac{b_{35}}{A} \\ \frac{b_{41}}{A} & \frac{b_{42}}{A} & \frac{b_{43}}{A} & \frac{b_{44}}{A} & \frac{b_{45}}{A} \\ \frac{b_{51}}{A} & \frac{b_{52}}{A} & \frac{b_{53}}{A} & \frac{b_{54}}{A} & \frac{b_{55}}{A} \end{bmatrix}$$

$$R_2 = \begin{bmatrix} \frac{c_{11}}{B} & \frac{c_{12}}{B} & \frac{c_{13}}{B} & \frac{c_{14}}{B} & \frac{c_{15}}{B} \\ \frac{c_{21}}{B} & \frac{c_{22}}{B} & \frac{c_{23}}{B} & \frac{c_{24}}{B} & \frac{c_{25}}{B} \\ \frac{c_{31}}{B} & \frac{c_{32}}{B} & \frac{c_{33}}{B} & \frac{c_{34}}{B} & \frac{c_{35}}{B} \\ \frac{c_{41}}{B} & \frac{c_{42}}{B} & \frac{c_{43}}{B} & \frac{c_{44}}{B} & \frac{c_{45}}{B} \\ \frac{c_{51}}{B} & \frac{c_{52}}{B} & \frac{c_{53}}{B} & \frac{c_{54}}{B} & \frac{c_{55}}{B} \end{bmatrix}$$

$$R_3 = \begin{bmatrix} \frac{d_{11}}{C} & \frac{d_{12}}{C} & \frac{d_{13}}{C} & \frac{d_{14}}{C} & \frac{d_{15}}{C} \\ \frac{d_{21}}{C} & \frac{d_{22}}{C} & \frac{d_{23}}{C} & \frac{d_{24}}{C} & \frac{d_{25}}{C} \\ \frac{d_{31}}{C} & \frac{d_{32}}{C} & \frac{d_{33}}{C} & \frac{d_{34}}{C} & \frac{d_{35}}{C} \\ \frac{d_{41}}{C} & \frac{d_{42}}{C} & \frac{d_{43}}{C} & \frac{d_{44}}{C} & \frac{d_{45}}{C} \\ \frac{d_{51}}{C} & \frac{d_{52}}{C} & \frac{d_{53}}{C} & \frac{d_{54}}{C} & \frac{d_{55}}{C} \end{bmatrix}$$

In matrixes above, the b_{ij}、c_{ij}、d_{ij} $(i, j = 1, 2, 3, 4, 5)$ are the sum of each one which is filled in the form as table above, for specialists, teacheres and students respeclively.

The matrix of comprehensive evaluation W_1, W_2, W_3 are obtained by matrix mutiplition:

$$W_1 = T_1 \cdot R_1 = (r_{11}, r_{12}, r_{13}, r_{14}, r_{15})$$
$$W_2 = T_1 \cdot R_2 = (r_{21}, r_{22}, r_{23}, r_{24}, r_{25})$$
$$W_3 = T_1 \cdot R_3 = (r_{31}, r_{32}, r_{33}, r_{34}, r_{35})$$

We decide distribution of proportion coefficient of varied evaluater as specialist、teacher and student as follow:

$$T_2 = (e_1, e_2, e_3), (\sum_{i=1}^{3} e_i = 1, 0 \langle e_i \langle 1, i = 1, 2, 3)$$

Then comprehensive evaluation matrix W can be obtained:

$$W = (e_1, e_2, e_3) \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} \\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} \end{bmatrix} = (p_1, p_2, p_3, p_4, p_5)$$

Let $W' = (q_1, q_2, q_3, q_4, q_5), q_i = \frac{p_i}{\sum_{i=1}^{5} p_i}$, Then $0 \langle q_i \langle 1, \sum_{i=1}^{5} q_i = 1$.

The evaluation fraction P can be obtained by matrix mutiplition:

$$P = (q_1, q_2, q_3, q_4, q_5) \begin{bmatrix} 95 \\ 85 \\ 75 \\ 65 \\ 50 \end{bmatrix}$$

The example by method above, space does not permit inclusion of the details.

四、Some illustrations

(一)How to ascertain proportion coefficient

In the above mathematical model, we have given out two proportion coefficients: $T_1 = (a_1, a_2, a_3, a_4)$, $T_2 = (e_1, e_2, e_3)$. The determination of the concrete number in it can be fixed in the following ways:

1. Specialists' estimation: It's the way to take the average value of the proportion coefficient of target system which have already planned out, after having taken

the statistics of specialists' views. Such method is very simple and convenient, but in some means it is subjective and casual.

2. Comparing the means:

Respectively base on the target which has the smallest important degree in the targets of T_1, T_2. marked its value as 1. So the other target according to their importance can be indicated as the multiples of its. Then sum them up and get the proportion coefficients.

3. Deffier consultation:

It is also a kind of specialists' consultation. It needs doing the estimation back and forth for several times, which avoided the casualness of the specialist's estimation. This differs from the specialists' estimation.

4. order analysis:

This method was introduced into this field, firstly by Americans. Its steps are: first, compare each target which formed a set of target (eg. a_1, a_2, a_3, a_4 in T_1; e_1, e_2, e_3 in T_2) in pairs. Then, distinguish the relative degree that different order targets have influnced on evaluated targets' realization and set up judge Matrix. At last, base on the Matrix, take some appropriative count and get the proportion coefficient of targets on the same class(order).

Due to limited space, here we give no examples to the above methods.

(二)the illustration on the teacher's emotion target factor

In the avaluation model, one of target factors which evaluate the level of teaching in classroom is emotion factor. It reflects one character of the evaluation tartet system in the model. teacher's emotion factor (it is also called unintellectual factor). These are expressions of the teacher's ethusiasm, interest and determination to the teaching. This factor has an inneglected effect on the teacher's classroom teaching. And conversation, observation and investigation can be used to evaluate the factor.

(三)Some noticable problems in using the model

1. The communation of the certain quantity and quanlity determined the evaluation method.

The process of classroom teaching is a complex process of teaching practice. In using the model, give a definite class to every target of target factor collection of set of target. In fact it is a certain quanlity process of evaluation. In summing up the evaluated informations of different evaluaters, the certain quantity mathematical way is appled. The two methods are not isolated. Only connecting them both can we get a scientific result of evaluation.

2. Pay attention to the charactor of different kinds of evaluation. Make full use of the effect of evaluater.

In the process of applying the evaluation model there are three kinds of evaluation. They are students evaluation, teachers evaluation and specialists evaluation. Try to grasp the characters of the three evaluations and praise its stronge points as well as avoid its shortcomings. Make full use of the effect of every evaluaters and get the result of evaluation objective and adjusted. Give the proportion which come from the three results (e_1, e_2, e_3 in T_2) analysize according to its own situation of the school of different kinds and different class. And definified them by using the ways metioned above.

3. A computer propramm can be made by the application of the results of model statistics evaluation. Calculation can be done by computer.

References

[1] Liu bengu:《theory of educational evaluation》. Dongbei Normal University, 1988. 10.

[2] Yi Peijun, Sueng dazang, jin jingping:《theory of mathematical teaching method》, 1987. 10.

[3] jin jingping, Yan guangjie:《education mathematics》, 1991. 12.

CAI/L IN FUNDAMENTAL STATISTICS
REVISING CURRICULUM & INSTRUCTION THROUGH TECHNOLOGY

Marissa P. Justan
Ateneo de Manila University
P.O. Box 154, Manila 1099, Phil.
(Email: mpjustan@mathsci.admu.edu.ph)

Cynthia B. Omampo
Trinity College
P.O. Box 10363, Broadway Centrum, Quezon City, Phil.
(Email:tcqc@philtap.tool.nl)

ABSTRACT

It is hardly necessary to argue that statistical literacy has become indispendable in almost any field today. An introductory knowledge of statistics can prove very useful. In recognition of the need to provide the necessary statistical background for 21st century students, the primary objective of this project was to adopt the computer-aided instruction/learning (CAI/L) method to facilitate the teaching of statistics to the underachievers in the tertiary level. In the Philippines, this method is still on its infancy stage. The problem of articulating and integrating them with instruction remains, therefore, to be adequately solved.

The next issue that confronted the reseachers was exactly which statistics software should be used. It should be simple, student-oriented, and must not cost too much. For this project, a special software named STAT121 was developed.

This paper reports on the results of a collaborative work between two tertiary math educators for the purpose of improving the quality of teaching, raising statistical literacy to a higher level and promoting technological awareness.

CAI/L IN FUNDAMENTAL STATISTICS: REVISING CURRICULUM & INSTRUCTION THROUGH TECHNOLOGY

Marissa P. Justan
Ateneo de Manila University
P.O. Box 154, Manila 1099, Phil.
(Email: mpjustan@mathsci.admu.edu.ph)

Cynthia B. Omampo
Trinity College
P.O. Box 10363, Broadway Centrum, Quezon City, Phil.
(Email:tcqc@philtap.tool.nl)

INTRODUCTION

The emphasis which is now being given to computers to improve instruction undoubtedly increases over the last 5 years. Some of the conditions contributing to the local CAI/L trend are the following: (1) *Increased emphasis upon science and mathematics*. The rate of development of science and technology is accelerating so fast, emphasizing the necessity for providing more up-to-date science education for the masses as well as for training scientists and technologists who would conduct research and apply findings to industry, medicine, home, life and defense; (2) *More modern course materials*. There has been a very serious effort to improve the content of a number of the subjects in all levels of education. Efforts have been made to bring up to date the content of courses in physics, chemistry, mathematics, social sciences, as well as in other fields; (3) *Increased amount and variety of experimentation*. There has been in the past few years a greatly increased amount and variety of "action research" and controlled experimentation in education. The experiments involve every aspect of education - subject requirements, new subject content, laboratories, and other devices for adapting learning materials and activities; (4) *The trend towards technology education* [HR,p.2]. Students should be made aware of personal, moral, social, environmental and economic implications of technology.

THE PROJECT & ITS RATIONALE

Despite recommendations by the Department of Education and Culture (DECS) and Department of Science and Technology (DOST) to make use of computers in mathematics curricula at all school levels, there are very few actual examples of computer use being systematically incorporated into the core mathematics curricula because of (1) the prohibitive cost of installing and maintaining of computers in the classrooms, (2) the lack of qualified teacher, and (3) the prohibitive cost of good educational softwares. Only a few schools in the Philippines have the necessary resources mentioned and most of these schools cater to middle and upper strata of the society and they have a reputation for high-academic standards and innovative mathematics teaching.

This project was carried out in a different setting involving one of the private colleges in metropolitan school system. Its primary aim was to integrate computer use with student's learning activities in statistics. The motivation for change rose from the concerns of teachers about the poor performance of their students in statistics.

The use of computers was deemed inevitable. The driving force behind this experiment was to lessen the rigours of computations so that even the underachievers and the not-so-highly motivated students can study and appreciate statistics.

Computer use, to be effective, must be integrated with student's learning activities. The use of computers should be appropriate and should accomplish some significant event. The accomplishment of this purpose requires considerable care in organizing the materials for instructional and learning purposes.

METHODOLOGY

The "joint work" established here involved mutual planning, observation and evaluation. The teaching and mentoring role was performed by the teacher directly involved in the project. The writing of the software and the computer exercises were done by the other researcher.

Sample

The school in this project is one of the middle-class private schools in Metro Manila. Two (Fundamental) statistics classes of varied mathematical abilities were chosen for the experiment. One class was randomly picked as the experimental group (EG) while the other class served as the control group (CG). Table 1 shows the frequency distribution of students according to their mathematical abilities (verified through the school's entrance test). Very few students of the EG had computer backgrounds.

Table 1. Frequency Distribution

CLASSIFICATION	CG	EG
Below Average	13	16
Average	17	17
Above Average	10	9
TOTAL	40	42

Procedure

The project was conducted over 6 months in the following stages: (1) investigation and development (2) implementation (3) Evaluation and follow-up.

Stage 1. Investigation & Development

During this stage, all the current instructional materials were gathered. The course syllabus and the textbook was a perfect match. The next issue was finding the software that would not only match the textbook but would also incorporate some psychological principles of learning. For experimentation purposes, programs were developed to suit these needs. This collection of programs was named STAT121 after Math121, the course code of fundamental statistics. Then, the STAT121 manual was prepared to contain the outline form of the topics to be discussed, computer exercises, and the guide to STAT121 program.

STAT121 program was written in Foxplus language and compiled in CLIPPER. It runs on personal computers based on the Intel 8086 or compatible microprocessors, on an IBM personal computer or PC compatible computer running PC-DOS or MS-DOS operating system software, at least version 2.1, with at least 256 K of computer memory, and a DSDD 5.25 inch or 3.5 inch drive.

The next issue was the availability of computers. Should there be enough so that every student has his own, or how many students should share in one? Since this was a pilot study, the administration suggested that students should share computers at the moment.

Stage 2. Implementation

The teacher handled both the CG and the EG. Each groups met three hours per week over 18 weeks. The CG used calculators in class while the EG used both calculators and computers. The following is a sample session conducted by the teacher in teaching the EG the technique in performing simple analysis of variance using data in a one-way classification model.

Compare the cleaning action of three detergents on the basis of the following whiteness readings made on 15 switches of white cloth, which were first soiled with India ink and then washed in agitator-type medium with respective detergents.

Referring to the data on Table 2, test the null hypothesis that the cleansing action of the three detergents are not significantly different from each, either at 95% or 99% level of significance using simple analysis of variance.

Table 2. Whiteness Readings

Detergent	A	B	C
Sample 1	77	72	76
Sample 2	81	58	85
Sample 3	71	74	82
Sample 4	76	66	80
Sample 5	80	70	77

First, the teacher demonstrated to the EG how to compute the F value using only the calculator. This was a tedious task and consumed one session. After the calculator work, the teacher demonstrated how to use the STAT121 program. To do this, the teacher followed the guide set in the STAT121 Manual.

Type the command "STAT121" to run the program. Figure 1 shows the main menu. Choose "Simple Analysis of Variance" option to go to the ANOVA subprogram (Figure 2). If there were existing ANOVA file, then choose "Erase Data File" option; otherwise, choose the "Add/Edit" option to add or change entries (Figure 3). After all the data has been entered, choose "Display Results" option. Figure 4(a) shows the ANOVA worksheet. To see the computed value of F, hit the < enter > key (Figure 4(b)). Hit any key to go back to Figure 2.

The EG expressed positive attitudes towards the use of computers despite the lack of, or limited nature of, their personal experiences with such technology. When the teacher assigned computer exercises, the emphasis was more on interpreting and analyzing the results.

Stage 3. Evaluation & Follow-Up

Throughout the whole experiment , same tests were administered to both groups. Since the number of lessons taken by the EG was more than that taken by the CG, the teacher included only the common topics in the tests. The EG group was given extra quizzes on the other lessons not covered in their common tests.

The teacher selected the students with below average abilities from both groups and compared their post-test results. Table 3 shows the mean scores and the standard deviations. Applying t-test at 5% significance level, the computed t- value is equal to 1.7072 which is greater than the tabular value 1.697. Then, the teacher randomly selected 16 students from the EG and 16 students from the CG and compared their post-test results (see Table 4). Again, applying t-test at 5% significance level, the computed t- value is equal to -2.814 while the tabular value of t is 1.697. Thus, there is sufficient evidence that the CAI/L method is more effective to these types of students.

Table 3. Below Average Post-Test Results

Group	Sample	Mean	Std.Dev.
Experimental	16	78.3	4.15
Control	13	75.6	4.34

Table 4. Mixed Abilities Post-Test Results

Group	Sample	Mean	Std.Dev.
Experimental	16	86.25	3.90
Control	16	81.38	5.73

At the end of the course, the researchers interviewed both student groups. Most of the answers favored the CAI/L method. Based on these findings, the administration decided to require teachers to use STAT121 program for classroom teaching and bought more computers to accommodate these additional computer users.

Figure 1.
Main Menu

Program

- Collection, Organization of Data
- Frequency Table
- Measures of Central Tendency
- Measures of Variation

- Probability Distributions
- Simple Tests of Hypothesis
- Simple Analysis of Variance
- Simple Regression, Correlation
- EXIT

Figure 2.
ANOVA Option

Program

- Collection, Organization of Data
- Frequency Table
- Measures of Central
- Measures of Variatio

- Probability Distributio
- Simple Tests of Hypo
- Simple Analysis of Variance
- Simple Regression, Correlation
- EXIT

- Erase Data File
- Add/Edit Data
- Display Results
- Exit

Figure 3.
Add/Edit Data Option

Record no.	:	1
X_1	:	77.0000
X_2	:	72.0000
X_3	:	76.0000

Add Change Delete Next Past Exit

Figure 4.
Display Results Option

X_1	X_2	X_3	X_1^2	X_2^2	X_2^3
77.00	72.00	76.00	5,929.00	5,184.00	5,776.00
81.00	58.00	85.00	6,561.00	3,364.00	7,225.00
71.00	74.00	82.00	5,041.00	5,476.00	6,724.00
76.00	66.00	80.00	5,776.00	4,356.00	6,400.00
80.00	70.00	77.00	6,400.00	4,900.00	5,929.00
385.00	340.00	400.00	29,707.00	23,280.00	32,054.00

(a)

Source of variation	Sum of squares	df	mean sum of squares
between -column	390.00	2	195.00
within-column	276.00	12	23.00
TOTAL	666.00	14	
computed F value	8.478		
tabular F value	3.880 (5%)		6.930 (1%)
CONCLUSION	reject		reject

(b)

RESULTS & DISCUSSIONS

CAI/L method is more effective

This study established the effectiveness of CAI/L method in teaching statistics to students of mixed abilities. The teacher and the students took advantage of the power and flexibility of the computers. The use of computers allow them to apply statistical concepts and processes more conveniently. It also helps emphasize statistical thinking and less emphasize the arithmetic of statistics. This shift on emphasis has been the important aspect of using computers in statistics in classrooms.

Revised Syllabus

The use of computers have made considerable impact in the completion of their course syllabus. It has long been a tradition for their average statistics class to end the term without covering all the lessons, but through the facilitation of the CAI/L method, the experimental group was able to take up the whole syllabus and performed a little better.

Old MATH121 Course Outline

- Introduction
- Collection, Organization of Data
- Frequency Distributions
- Measures of Central Tendency
  - Mean, Median, Mode
  - Quantiles
- Measures of Variation
  - Range, Quartile Deviation
  - Mean Absolute Deviation
  - Standard Deviation
  - Variance
- Counting Techniques
- Permutation & Combination
- Probability Concepts
- *Probability Distribution
  - Expected Value & Variance
  - Binomial Distribution
- Simple Test of Hypothesis
  - Z-Test
- *Simple ANOVA
- *Simple Regression
- *Simple Correlation Analysis

*not covered in class

New MATH121 Course Outline

- Introduction
- Collection, Organization of Data
- Frequency Distributions
- Measures of Central Tendency
  - Mean, Median, Mode
  - Quantiles
- Measures of Variation
  - Range, Quartile Deviation
  - Mean Absolute Deviation
  - Standard Deviation
  - Variance
- Probability Distributions
  - Probability Concepts
  - Permutation & Combination
  - Expected Value & Variance of a Probability Distribution
  - Binomial Distribution
  - Poisson Distribution
  - Normal Distribution
- Simple Tests of Hypothesis
  - Z-Test
  - T-Test
  - Chi-Square Test
- Simple ANOVA
- Simple Regression Analysis
- Simple Correlation Analysis

The STAT121 Program

Thru the project, this school was able to develop its own statistics software tailor-fit to the needs and backgrounds of their students. However, the teacher does not limit his students to the use of STAT121. Students are allowed to use other available softwares, local or foreign, outside their class laboratory.

Some Feedbacks

The following sample interviews indicated student's positive attitudes towards the use of computers. They agree that it is becoming important as part of their education and gave them a perspective of what technology can accomplished for them personally.

Sophomore, B.S. Biology: Statistics is very confusing but the (STAT121) manual helps me understand... The (STAT121) program needs improvement but I guess it is a good idea having it as part of our study.

Sophomore, A.B. Mass Communication: I find Statistics very interesting and a little difficult. STAT121 program is very advantageous to students like us since it helps us to understand more clearly.

Freshman, B.S. Hotel & Restaurant Management: Statistics is hard to understand. STAT121 program is also hard to understand but it is very interesting to use. It gives us better understanding of the subject.

Junior, B.S. Computer Science: Statistics is very difficult. STAT121 program provides us the correct answers. It also shows some portions of the solutions.

CONCLUSION

The results of this project appear to support the efficacy of the CAI/L method in teaching statistics. This method was superior than the traditional method in terms of time, preparation, ease of delivery, motivation and confidence. The EG appeared to have performed better and was much more motivated because they tried and computed all the problems. This was not the case with the CG. It was clear that the EG possessed a better understanding of the concepts and techniques which gave them the confidence and the motivation to tackle more problems.

STAT121 makes no claim to being flawless or all-encopassing in its scope. For an experiment, it helps fill a need for a change in istructional strategy. However, it must also pass the tests of efficiency. Although a handful of foreign and local statistical packages are available in the market, this project provided the opportunity for developing the school's own statistics program. The teacher found it very convenient since it is a very simple, student-oriented program and it is in series with their lessons. Since this program was specially developed for this project, there was really no problem about its continuous improvement and software copyright.

Many factors could have affected the success of the CAI/L in this project but the outcomes of this study appear to lend support to Vistro-Yu's findings that computer use can be integrated in mathematics curricula and improve student's learning as well.

References

[D] Douglass, Trends & Issues in Secondary Education, Prentice-Hall of India (Private) Ltd., 1965.

[PG] Pagoso and Garcia, Fundamental Statistics, Rex Printing Co., Inc., Phil., 1985.

[HR] Harding & Rennie, Technology Education in Science & Mathematics, Key Centre for School Science & Mathematics, Curtin University of Technology, Western Australia, 1987.

[V] Vistro-Yu,Teaching Calculus with Computer applications, Ateneo De Manila University, Phil., 1994.

[OJ] Omampo and Justan, MATH121 Manual, TRI Publishing Co., Phil., 1994.

THE IMPORTANCE OF BEING ACCESSIBLE: THE GRAPHICS CALCULATOR IN MATHEMATICS EDUCATION

Barry Kissane
Murdoch University
Murdoch, Western Australia, 6150
kissane@murdoch.edu.au

The first decade of the availability of graphics calculators in secondary schools has just concluded, although evidence for this is easier to find in some countries and schools than in others, since there are gross socio-economic differences in both cases. It is now almost the end of the second decade since the invention of microcomputers and their appearance in mathematics educational settings. Most of the interest in technology for mathematics education has been concerned with microcomputers. But there has been a steady increase in interest in graphics calculators by students, teachers, curriculum developers and examination authorities, in growing recognition that *accessibility* of technology at the level of the individual student is the key factor in responding appropriately to technological change; the experience of the last decade suggests very strongly that mathematics teachers are well advised to pay more attention to graphics calculators than to microcomputers.

There are clear signs that the commercial marketplace, especially in the United States, is acutely aware of this trend. It was recently reported that current US sales of graphics calculators are around six million units per year, and rising. There are now four major corporations developing products aimed directly at the high school market, with all four producing graphics calculators of high quality and beginning to understand the educational needs of students and their teachers. To get some evidence of this interest, I scanned a recent issue (April 1995) of *The Mathematics Teacher*, the NCTM journal focussed on high school mathematics. The evidence was very strong: of almost 20 full pages devoted to paid advertising, nine featured graphics calculators, while only two featured computer products, with two more featuring both computers and graphics calculators.

The main purposes of this paper are to explain and justify this heightened level of interest in graphics calculators at the secondary school level, and to identify some of the resulting implications for mathematics education, both generally, and in the South-East Asian region.

DESCRIPTIONS AND DEFINITIONS

In some respects, graphics calculators are similar to other calculators, such as scientific calculators and four-function calculators, which have become quite familiar to mathematics teachers over the last two decades. The most notable similarities are that each is small enough to be hand-held, has an independent power source, is operated by pressing keys, has a monochromatic numerical display and is quiet in operation. Because of these surface similarities, it is not surprising then that the term 'calculator' was used by manufacturers to describe this new form of technology.

However, the differences between graphics calculators and other kinds of calculators are much greater than the similarities, and has given rise to substantial misgivings about the use of the term 'calculator' to describe such different devices. Perhaps the most obvious difference is the graphics display screen. As the name suggests, a graphics calculator screen can be used to display graphs of functions or of statistical data, both clearly of considerable value in mathematical work. This property, together with typical advertising images, has given some mathematics teachers the impression that a graph-drawing capability is the distinguishing feature of a graphics calculator.

This is not the case, however. From the perspective of mathematics education, the most important difference between a graphics calculator and its ancestors is not the display screen, but the mathematical capabilities built into the device. (As an aside, it is interesting that the developers of the first 'supercalculator', Hewlett Packard's HP-28, added the graphing capabilities as an afterthought to the rest of the mathematical software embedded in the calculator, and certainly did not think of them as the major innovation of the device.) Although there are many differences among them, most currently available models contain a significant suite of mathematical capabilities, including the standard functions found on scientific calculators, together with function graphing (on rectangular or polar coordinates, with explicitly, parametrically or recursively defined functions), manipulation of graphs, numerical equation solving and root-finding, data analysis (both numerically and graphically), matrix manipulation, operations with sequences and series, complex number arithmetic and numerical differentiation and integration. In addition, all are programmable, and have considerable memory for longer-term storage of programs, data, matrices and images. (Indeed, several calculators on today's market have more user-accessible memory than did microcomputers of the late 1970's.) Some calculators have limited symbolic manipulation capabilities as well. Modern graphics calculators have some communication capabilities, to other calculators, computers, printers or overhead projection units.

A major consequence of these kinds of capabilities is that a graphics calculator can be used to *analyse* a mathematical situation rather than merely to perform a computation, an enormous surge of mathematical power, when compared with the scientific calculator. The scientific calculator actually provided students with little more than the four-function calculator. The most obvious advances were table facilities (replacing the previous need to have trigonometric, logarithmic and exponential tables), statistical facilities (which only replaced arithmetical aspects, since actual data were not stored for analysis) and programming facilities (which in fact were rarely used in schools). One could be forgiven for thinking that the graphics calculator is, like the scientific calculator before it, merely a slight advancement in terms of mathematical power. But one would be wrong.

Alternative names for graphics calculators have been suggested. Two in everyday use are 'graphing calculator' and 'graphical calculator'. There are also compound names, such as 'programmable graphical calculator', 'graphing scientific calculator' or 'advanced graphic calculator', but it is not clear whether these are attempts to grapple with the conceptual problem of adequately defining the devices or whether they are motivated by marketing issues. More expressive suggestions have been made, including 'supercalculator' and 'compulator', each of which acknowledges a growing unease with the term, 'calculator'. It is worth noting the historical precedent for difficulties with describing devices of these kinds. There is not much 'scientific' about scientific calculators and four-function calculators almost always have more than four functions.

Whatever it is called, however, it is preferable to think of the graphics calculator as a small, portable computer with inbuilt mathematical software that costs substantially less than other kinds of computers. It is arguably the most potent influence on the secondary mathematics curriculum of today, and particularly on the curriculum of tomorrow.

GRAPHICS CALCULATORS OR COMPUTERS?

Until recently, technology in mathematics education referred mainly to computing technology, fuelled by the extraordinary rise to prominence and reduction in price of the microcomputer over the past two decades. In many respects, such 'high' technology continues to excite both mathematics educators and mathematicians alike. This is especially so in the past decade when a number of significant developments in software for mathematics and mathematics education have been developed. Some examples of such software are *Derive*, *Mathematica*, *Theorist*, *Maple*, *Logo*, *Cabri-géomètre*, *Geometer's Sketchpad*, *AutoGraph*, *MousePlotter*, *ANUGraph* and *MiniTab*. In affluent western countries, many secondary schools have acquired significant microcomputer resources, usually in the form of laboratories full of machines. A small number of elite schools have even begun to equip each of their students with portable laptop computers and to plan their curriculum accordingly. Schools with less resources have opted for a configuration of a demonstration computer in each classroom, or a mobile computer that can be shared between classrooms.

While such developments at the leading edge of high technology are exciting and challenging, the reality for many students in more affluent countries, and for almost all students in less affluent countries is that microcomputers are too expensive and hence not a significant force on mathematics education. In short, the technology has been, and continues to be, inaccessible to the great bulk of students for almost all of their time in school. Further, since agencies responsible for official curricula naturally attend to the circumstances surrounding the mass of students, rather than an elite few, there have been almost no significant effects of microcomputers on the mathematics curriculum prior to undergraduate education.

As the title of this paper suggests, the key to understanding the significance of graphics calculators is their potential for increasing the accessibility of technology to individual students. There are two aspects to this accessibility. In the first place, the purchase price of graphics calculators, while still too high for many individual students, places them within reach of many more classrooms than do microcomputers. Schools can purchase a class set of graphics calculators for around the same price as a single microcomputer sufficiently powerful to operate modern innovative software. This is certainly the case if the cost of the software is taken into account, since graphics calculators come complete with their own mathematical software, while computers demand that the software be purchased separately. Even for individual students, the cost in present terms of a graphics calculator, spread over several years of schooling has become comparable with the cost of scientific calculators late in the 1970's, especially if students do not need to purchase a scientific calculator as well. The remarkable surge in recent sales of graphics calculators in affluent western countries suggests that many schools and individuals find them affordable.

The second aspect of accessibility is a consequence of the physical size of graphics calculators. Small, light, battery-operated computers are clearly much more portable than are large, heavy, electrically-powered computers. Graphics calculators are as potentially mobile as the students for whom they were designed. They can easily be taken home and they can accompany students to an examination room or on a field trip. They can easily be moved around a school; in my institution, we have a briefcase containing a set of graphics calculators, which allows our 'computer laboratory' to be wherever we and the students are, rather than the *much* more difficult problem of transporting the students to a computer laboratory. (Bradley, Kemp & Kissane, 1994) As long ago as 1986, the University of Chicago School Mathematics Project, developing an innovative 11th grade course (Rubenstein et al., 1992) based on a premise of significant computer access, found that schools were much more likely to be able to acquire access to graphics calculators than to computers, which were often used by computing subjects, were located inconveniently in computer laboratories and required too much advance booking of rooms to be a realistic option.

Despite their relative inaccessibility and price disadvantage, it should be acknowledged that microcomputers enjoy some significant advantages over graphics calculators for secondary mathematics education. They are much more powerful, are faster and can use much more mathematically and educationally sophisticated software. They are much more versatile, in the sense that the same computer can be used for many different purposes. Computer screens are larger and have higher resolution than current graphics calculator screens, and are generally coloured while (most) graphics calculator screens are monochromatic; thus more information can be presented more effectively. Computer software is more easily upgraded and modified than is graphics calculator software. Computers rely on electricity rather than batteries, which are a nuisance to replace. (Although this is not always an advantage for computers; I heard recently of one South-East Asian country in which many schools were issued computers from a central government, although they lacked adequate electricity supplies to operate them.) It is usually easier to print from a computer than a graphics calculator. However, these many advantages of microcomputers over graphics calculators evaporate and are merely of academic interest if students do not enjoy ready access to machines.

While it would be incorrect to claim that there is not a high technology element in today's graphics calculators, and no more correct to regard them as 'low' technology, it seems more reasonable to regard them as an example of what Schumacher referred to as '... *intermediate technology* to signify that it is vastly superior to the primitive technology of bygone ages but at the same time much simpler, cheaper and freer than the super-technology of the rich.' (1974, p.128) Graphics calculators are *much* more appropriate than are microcomputers to the realities and constraints of most students in most classrooms in most countries at this moment in time, and for at least the next few years, and are thus a form of intermediate technology for school mathematics education.

METAPHORS FOR GRAPHICS CALCULATORS

The previous section indicated flaws in using the metaphor of a calculator rather than that of a computer to think about graphics calculators. In fact, it seems that people invoke a number of metaphors to help them come to terms with graphics calculators and other computers relevant to mathematics education. These are described in detail in Kissane (1995c) and include the following:

Laboratory. The calculator provides opportunities for exploration of mathematical ideas and situations, akin to the explorations characteristic of scientists in a laboratory. Both doing and learning to do mathematics have significant elements of personal exploration associated with them, and the calculator provides a powerful environment for such things to occur.

Tool. The calculator provides a tool for doing a particular mathematical task, so that learning when to use it, when not to use it and how to use it well are as important for students learning mathematics as the equivalent learning is for people learning the tools of other trades.

Teaching aid. Since computers are often regarded as teaching aids (in part because of their limited availability), it is not surprising that some people think of graphics calculators as devices to help teachers to teach more than as devices to help students to learn. The availability of overhead projection capabilities strengthens this metaphor.

Curriculum influence. The accessibility and portability of the calculator demand that serious attention be paid to whether we are teaching the right things, in the right way at the right time to the right people. It is especially significant that it is the *central* elements of the secondary curriculum that are most susceptible to the influence of the graphics calculator, notably the traditional algebra, trigonometry, calculus sequence as well as statistics and probability, as elaborated below.

Status symbol. As for other areas of technology, it is inevitable that some will focus attention on next year's models or the features that one calculator lacks in comparison with others. Marketing people are naturally sensitive to this orientation, which is certainly not restricted to graphics calculators, but is quite evident with other computers too.

Cheating device. There is a strong tradition within mathematics education communities for doing mathematics the 'right way', and hardly surprising that some people's orientation to new ways of doing things is to regard them as illicit. The continuing strong influence of formal examinations in mathematics is also a factor here.

In recent work with first year undergraduate students, (Kissane, Kemp & Bradley, 1995) found evidence of each of these underlying metaphors for thinking about graphics calculators, and also suggested that an additional metaphor of the graphics calculator as a *nuisance* seemed to characterise the reactions of some students, who regarded the intrusion of graphics calculators into a course as merely adding to the burden of what is to be learned in an already crowded curriculum.

The significance of these various metaphors is that they help us to understand why there is a range of reactions from both students and their teachers to the use of graphics calculators, along a spectrum from unbridled optimism to hostility and derision. It seems important for *each* of these metaphors to be brought to the forefront in discussions about technology in general, and graphics calculators in particular, in order to achieve a balanced appraisal of each of the prospects, pitfalls and possibilities associated with technological change.

CURRICULUM IMPLICATIONS

There are considerable implications for the curriculum of the widespread availability and use of graphics calculators, as Romberg (1992) recently observed:

> Computers and calculators have changed the world of mathematics profoundly. They have affected not only what mathematics is important but also how mathematics is done. It is now possible to execute almost all of the mathematical techniques taught from kindergarten through the first 2 years of college on hand-held calculators. This fact alone must have significant effects on the mathematics curriculum. ... the changes in mathematics brought about by computers and calculators are so profound as to require readjustment in the balance and approach to virtually every topic in school mathematics. (p.772)

There are implications for what is taught in secondary school mathematics, for how it is taught and for when it is taught. There is not space in this paper to do justice to the complete range of these implications; rather a selection is made.

The algebra curriculum has long been the central core of high school mathematics, and has concentrated upon symbolic manipulation to deal with both expressions and equations. Development of manipulative skill in algebra has been a major goal, based on the premise that no further progress in mathematics is possible without a fluent grasp of these skills. Evidence for success of this approach has been generally disappointing, with many students apparently acquiring the skills at the expense of the associated understanding of algebraic concepts and even the whole notion of generalisation. Until quite recently, there was limited emphasis on graphs and graphing in school algebra for the practical reason that it takes students so long to draw graphs that there is little time left to make use of them.

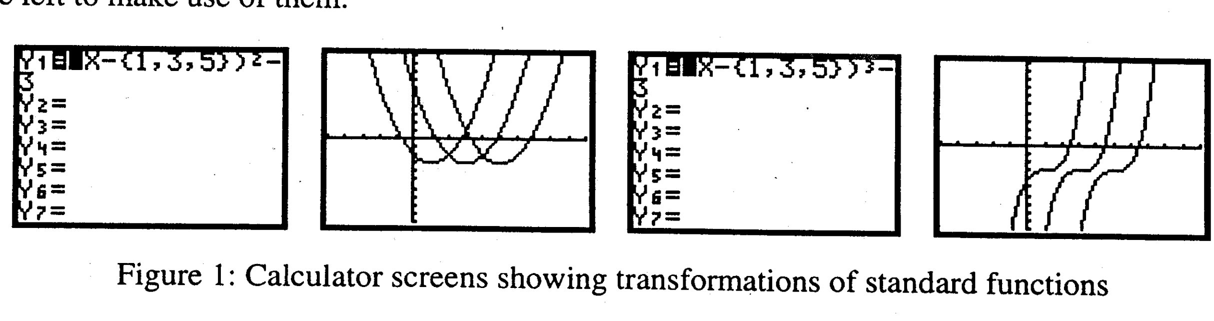


Figure 1: Calculator screens showing transformations of standard functions

The graphics calculator has allowed much more emphasis to be placed on graphs and their interpretation, both to help students understand key ideas (such as those of function, gradient and transformation) and to deal with practical algebraic problems (such as solving equations). The ease with which calculators can draw graphs means that students can concentrate on the meanings inherent in graphs instead of the mechanics of producing them. That is, the important curriculum task becomes to *make use of* a graph, rather than to *produce* a graph, which is a profound shift. Figure 1 (and other calculator screens in this paper) shows some calculator screens from a Texas Instruments TI-82™ graphics calculator. These screens illustrate how students might easily graph families of functions to help them understand horizontal transformations by studying many examples.

Modern graphics calculators provide a capability not unlike that of some innovative computer software to provide users with ready access to three different representations of functions simultaneously. These three aspects, sometimes referred to as 'the rule of three' are the symbolic, graphical and numerical respectively represented on calculators by symbols, graphs and tables of values. Figure 2 shows an example of this. There are clear advantages to understanding the nature of generalisations associated with seeing their particular manifestations in graphical or tabular (numerical) form, and also seeing the effects of changing the symbolic representation on the graphical and the numerical representations. In addition, the availability of graphics calculators suggests a better use of classroom time for exploration and conceptual development as well as applications of algebra to real situations, rather than an exclusive focus on the development of algebraic skills.

Another key idea in school algebra is that of equations. The graphics calculator has considerable impact on ways of dealing with this. Both graphical and the numerical representations of situations give rise to efficient and insightful ways of seeking solutions to equations. In addition, modern graphics calculators have an automatic solve command, so that numerical solutions to elementary equations are provided. An extended description of the range of possibilities is given in Kissane (1995b). Prior to the availability of technology such as graphics calculators, algebra curricula were mainly constrained to the solution of linear equations or equations for which factorising was appropriate. Iterative and approximate solution strategies were not technically feasible, and so were neglected. Now, there are serious doubts on whether factorisation is worth the trouble and time that it takes to teach and learn, at least if high levels of skill are expected. There has long been a perfectly good alternative to factorising quadratic expressions in order to solve quadratic equations – the quadratic formula – but this does not seem to have diminished enthusiasm for teaching young students about factorising trinomials. Only time will tell whether other sacred cows of the mathematics curriculum, such as exact values of trigonometric functions and a preoccupation with trigonometric identities will endure. However, the curriculum continues to be a zero-sum game, so that if new techniques and ideas are to be included, something must be removed to make room.

At present, symbolic manipulation capabilities of graphics calculators are rather limited, where they exist at all, but this is a temporary state. Recent developments such as the Texas Instruments TI-92™ and the Hewlett Packard HP-38G™ calculators have inbuilt symbolic manipulation capabilities and it seems reasonable to expect this trend to increase. Once again, although the capabilities are likely to be much more limited than those of fully fledged computer algebra systems, the calculator accessibility advantage is critical. Opinions will be divided for a while on the merits of allowing students access to automatic symbolic manipulation For example, Waits & Demana (1992) suggested that graphics calculators provided more support for algebraic intuition than did symbolic manipulation software at that stage while French (1993) anticipated the forthcoming debate:

> The implications for the teaching and learning of algebra are immense, because so much of the traditional development of manipulative skills must be called into question. If simplification, factorisation and equation solution are available at the press of a key, as well as the means of plotting graphs, we are again faced by questions about the algebraic understanding and skills that children need to develop and the opportunities, as well as the difficulties, created by such a powerful tool. (p. 18)

Some recent curriculum development projects concerned with algebra have proceeded on an assumption of access to graphics calculator technology, including senior school courses such as the Nuffield Advanced Mathematics Project (Nuffield Foundation, 1994) in the UK and the University of Chicago School Mathematics Project (Rubenstein et al., 1992) in the USA as well as lower secondary school courses, such as *Access to algebra* (e.g., (Lowe et al. 1994; Lowe et al., 1994) in Australia. Indeed, Heid (1995) suggests that the availability of technology to students will affect profoundly the algebra curriculum across the whole range of schooling. A clear shift in emphasis away from symbolic manipulation for its own sake and towards the use of algebra to understand and model real situations is evident in all of these curriculum development initiatives.

Although it is still too early to judge attempts to use graphics calculators in these ways, the early signs are encouraging. In a recent review of research, Dunham & Dick (1994) concluded:

> The early reports from research indicate that graphing calculators have the potential dramatically to affect teaching and learning mathematics, particularly in the fundamental areas of functions and graphs. Graphing calculators can empower students to be better problem solvers. Graphing calculators can facilitate changes in students' and teachers' classroom roles, resulting in more interactive and exploratory learning environments. (p. 444)

It is still rather difficult to get good research evidence on the desirability of various kinds of curriculum change in algebra, however, since the contexts of the research are still often fairly artificial. At present, it is still necessary to make inferences about the impact of graphics calculators on the learning of students in classrooms organised to take advantage of the technology, over substantial periods of time and following curricula that have been designed to incorporate it appropriately, since direct evidence is not yet available.

As for algebra, there are considerable implications for elementary calculus of the availability of graphics calculators. Like algebra, calculus in high school and the early undergraduate years has long been learned (although not necessarily taught) as a collection of symbolic manipulative skills, with too little time left to focus on the important concepts involved. This has not gone unnoticed, of course. As one senior Australian mathematician recently put it, 'Students come into my university with well-developed skills at integration, but with almost no idea of what integration is *for*.' Access to graphics calculators has the potential to concentrate student attention on the big ideas of the calculus, and attend to the development of appropriate manipulative skills later.

An example of this is the notion of a derivative. Standard calculus courses deal with the slope of a tangent to a curve as the limiting case of secants to the curve through a particular point. Derivations from first principles are generally unconvincing to students, who focus instead on their symbolic consequences, consisting of various rules for finding derivative functions. With personal access to graphics calculators, it is possible for students to explore the gradient of a curve *directly* by successively zooming in on a curve. Functions that have derivatives are 'locally straight' and so it is technically unnecessary to deal with the tangent at all. Modern graphics calculators allow students to trace a curve, numerically evaluating the derivative at any point, leading naturally to the idea of a derivative *function*, which can be automatically graphed, as shown in Figure 2.

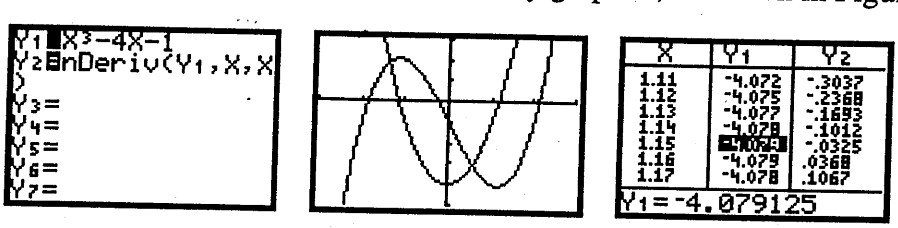


Figure 2: Three representations of a function and its derivative

In the past, many students seemed to regard derivatives as *expressions* rather than as functions; the conceptual value of graphing a function and its derivative simultaneously has only recently been appreciated. In Figure 2, the turning points of function Y_1 are just the points where its derivative crosses the *x*-axis. A tabular representation of the same information is also shown, giving numerical values for the function and its derivative at points near the relative minimum point. Students can toggle between these three representations to help them make connections between them, and can readily see the effects of changing the original function on the graphs or table. With such representations at their fingertips, students have much richer access to the important ideas of the calculus than are afforded by our traditional preoccupation with symbolic manipulation.

A major endpoint of studying the derivative in elementary calculus is to find relative extrema of functions. Modern graphics calculator allow these to be found numerically and directly, long before calculus is studied. For example, the topic of optimisation appears naturally in Lowe et al. (1994, p.103), where it is treated numerically, graphically and informally, typically a full year before it would appear in a calculus course. In addition, most graphics calculators have various commands for finding extreme values numerically, such as the two illustrated in Figure 3.

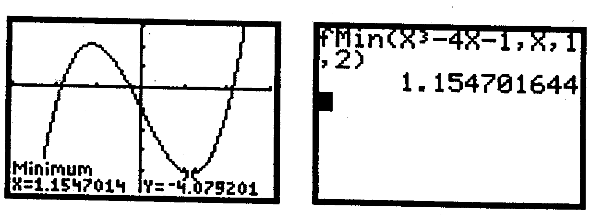


Figure 3: Automatic determination of a relative minimum

Numerical procedures do not provide exact answers, of course, which explains why the two results shown in Figure 3 are accurate to only five places of decimals. However, such a level of accuracy would meet any genuinely practical purpose plausibe to students. The numerical procedures also provide students with a good intuitive grounding and some motivation for later study of exact procedures through a symbolic calculus

There are a number of other ways in which the calculus curriculum might be affected (generally for the better) by student access to graphics calculators. Some details are given in Kissane (1995a). The calculus reform movement in the USA has been substantially influenced by the development of graphics calculator technology. One of the people involved in this movement (Kennedy, 1994) summed up the sentiments of many fellow reformers recently:

> We have been teaching the math literate of tomorrow with the problems of yesterday, while explaining to them that they will *need this mathematics in the future*. My friends, this is the stuff of which the Emperor's New Clothes are made! While we have been spinning golden oldies on the phonograph, perhaps more accurately the victrola, the world or Mathematical Reality has gone CD." (p.607)

Finally, a third example of the implications of graphics calculators for school mathematics concerns data analysis. A scientific calculator allows students to derive some numerical statistics (such as means, standard deviations and regression coefficients) from a set of data. In contrast, and like other computer packages, a graphics calculator actually *stores* the data so that transcription and entering errors can be detected, data can be sorted, alternative analyses can be performed, graphs can be produced, relationships examined using scatterplots, outliers removed and the data generally *analysed*, rather than merely summarised. Hackett & Kissane (1993) explore the implications of these sorts of capabilities in more detail. Figure 4 shows some examples of the data analytic capabilities of the TI-82™, showing some bivariate data (with one set sorted), a histogram and statistics for the first variable, and a pair of box plots allowing the two distributions to be compared.

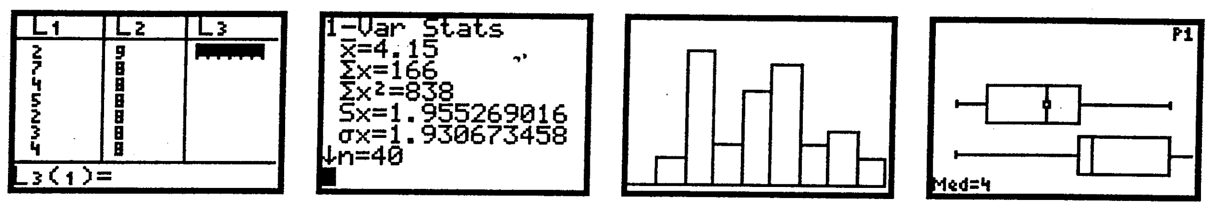


Figure 4: Some data representations from a graphics calculator

These four screens illustrate the general point that a modern graphics calculator allows students to engage in the same kinds of exploratory data analysis as they would do on a microcomputer; once again, the relative accessibility of the graphics calculator gives it a significant advantage.

This section of the paper is already long, yet contains only a few examples of possible influences of graphics calculators on the mathematics curriculum . It is important to bear in mind that this development does not impact only on aspects of the curriculum of marginal and passing interest; many aspects of the traditional *core* of the secondary curriculum can be taught and learned differently, and may be subject to a new kind of curriculum scrutiny in the light of this technology, as the earlier quote from Romberg suggests. This kind of scrutiny has not been undertaken with microcomputers in mind, since there have rarely been enough physical facilities available for the results to justify the considerable effort needed to obtain them. But that has now changed. If it becomes possible that most, or even very many, students can have personal access to graphics calculators, then a serious examination of the curriculum implications will be worth undertaking.

ASSESSMENT ISSUES

Successful incorporation of technology into mathematics curricula is only possible if careful thought is given to issues of assessment. The critical idea is that there should be a coherence between the conditions in which students normally learn and do mathematics and those in which their achievements are formally assessed. In framing their US mathematics curriculum standards, the National Council of Teachers of Mathematics (1989) was unambiguous in its recommendation regarding technology, suggesting that all high school students should have access to a scientific calculator with graphing capabilities at all times. Without a high level of integrity of this kind, there is little prospect of the technology being regarded as important by teachers or students alike. Indeed, the sorry history of the microcomputer in mathematics education is readily explained with reference to this precise point, since it has been too hard to provide students with enough computer access to realistically assess their mathematical achievements. The portability of graphics calculators resolves much of this problem, but at the same time exposes other problems to be addressed.

Assessment can take many forms in mathematics, ranging on a continuum from casual informal observation of students in their classrooms, through submitted work such as assignments and projects to more formal means of assessment such as classroom tests and timed external examinations. There would seem to be few problems associated with integrating graphics calculators into less formal assessment processes; the major issues are related to the more formal settings of tests and examinations. The three most pressing issues concern the styles of questions appropriate to graphics calculator use, the possibility of some students having an advantage over other students because of the use of different models of graphics calculator with different features and the emerging problems associated with symbolic manipulation capabilities of calculators. These issues are much less problematic at a local level (say, an individual classroom or school) than they are at a regional (state) or national level. Of course, they are not problems at all for settings in which mathematics examinations are not relied upon for assessment, but there are still very few of these.

Experience related to the first issue is accumulating, since UK Examination Boards have permitted the use of graphics calculators in A-level examinations for some years now, and there have been isolated experiences of other kinds reported as well (E.g., Kissane, Bradley & Kemp, 1994). Significantly, the US College Board has allowed graphics calculator use for the Advanced Placement Calculus AB and BC examinations since May 1995. The new arrangements include some questions (both multiple-choice and free response) for which calculators are *required* and others for which students are not permitted to use a graphics calculator. In addition, rather than demanding that students clear their calculator memories before the examination begins (a practice apparently used in the UK examinations, although it is hard to imagine how it can be successfully invigilated), students are encouraged to use the programming features of calculators to ensure that their calculator has a minimum set of capabilities. The College Board itself provides suitable programs for students to enter into their calculators for this purpose, and suggests that these be well understood before the examinations begin. In this way, the examiners can be reassured that each student has available to them a facility for graphing, numerical equation solution, numerical integration and numerical differentiation. Although not a complete solution, this seems to be a good way of reducing some of the perceived inequities among students.

As far as question styles are concerned, there is a considerable difference between *requiring* and *allowing* a graphics calculator to be used. In the former case, questions are asked which cannot be reasonably be answered by students without using a graphics calculator, and it is expected that part of the assessment task is to decide when calculator use is a good idea and when it isn't. When calculators are merely allowed on examinations, there is a tendency to try to ask calculator-neutral questions. Bradley (1995) has noted that one common way of doing this is to replace some numbers with algebraic symbols, which unfortunately may have the effect of making questions harder than intended. In addition, setting examination questions to avoid any advantage being conferred on graphics calculator use is not consistent with sound integration of technology into the curriculum referred to above, and may ultimately be counter-productive. Presumably, the reason for 'allowing' rather than 'requiring' graphics calculator use is that some students do not have graphics calculators yet, or have not yet had sufficient time to become fluent with them. If this is a temporary problem, its solution will be delayed by avoiding the issue of what is important when *everyone* has a graphics calculator available to them. One of the few advantages of timed external examinations is the prospect that they might be used as a fairly powerful form of encouragement to schools to change in some ways. However, the use of calculator neutral papers would seem to undermine this potential. Once again, solving one educational problem seems to create others.

Care is needed to deal with what students actually write down, since graphics calculator use can often permit students to provide a numerical answer only; if an explanation is to be given, it seems important that students be explicitly advised about this. Similarly, students need to be informed when exact answers are expected and when a numerical approximation is adequate, and when they should make (and defend) a choice between these. These and other issues of these kinds are also referred to by Kemp & Kissane (1995) and Kissane et al. (1994). The little experience so far formally gathered and published on the use of graphics calculators in examinations suggests that students do not generally make best use of the potential of the devices, according to Bradley (1995)

The issues associated with symbolic manipulation are more difficult to resolve, as noted by Bradley (1995), but are already in need of urgent attention with the latest batch of new models of graphics calculators including significant symbolic capabilities. The 'solutions' of prohibiting the calculators for examination use or of disabling the symbolic capabilities may provide a temporary respite from the problem, but certainly are not sufficient for the longer term.

A particular problem in many places is the speed of change; technological change tends to be frighteningly fast while educational change tends to be extraordinarily slow. Thus, in most Australian states, schools are given at least two years advance notice of significant changes to examinations; in the UK, examination papers tend to be set a full two years in advance of their administration. For many elements of society, even two years notice of impending change is inadequate. School systems, publishing companies and many teachers, students and their parents have great difficulty coming to terms with significant change. On the other hand, graphics calculators are prone to change enormously over the course of two years, especially now that there are four significant international corporations competing for the same market.

Predicting the future is hazardous at the best of times; predicting the future developments in graphics calculators seems an especially error-prone activity, and not for the faint-hearted. It is clear that the development of symbolic algebraic capabilities on inexpensive graphics calculators raises difficult problems for assessment. The same is likely to be true for the development of dynamic geometry systems, spreadsheet capabilities and improved screen resolution, as well as other enhancements of graphics calculators. The next few years are likely to be difficult ones for people responsible for mathematics examinations, trying to chart a course between changing too slowly and changing too rapidly, when it is not really clear which is the greater evil.

CONDITIONS FOR IMPLEMENTATION

There are remarkably few voices in the professional literature opposed to the integration of technology into mathematics education, and even today's suite of available graphics calculators have been well enough designed to be enormously useful to students learning mathematics and teachers teaching them. The main barriers to the more widespread educational exploitation of this technology are human and financial rather than technical. To describe the conditions for successful implementation in detail would require a companion paper to this one, but this section will summarise the main requirements evident at the moment.

The most critical factor is likely to be the financial one. Graphics calculators remain relatively expensive (compared with paper and pencil, chalk, textbooks and other bare necessities of mathematics education) even though they are relatively inexpensive compared with microcomputers and their software. Although there is now a relatively inexpensive graphics calculator available from one manufacturer (the TI-80™ from Texas Instruments), most of the development seems to be focussed on producing calculators with improved features. As for microcomputers, prices have stabilised and the products are improving. For many schools and their students, the stabilised price is too high, however, especially in developing countries. At the very least, strategies of directing educational resources towards graphics calculators instead of computers would seem to be wise, but a companion strategy of producing inexpensive calculators with basic features may also be a good idea. There is scope here for productive partnerships between education and industry.

The provision of graphics calculators hardware is a necessary but not a sufficient condition for effective change, however. The frequently forgotten element in curriculum development, the classroom teacher, is crucial. She will need help, support and encouragement to use graphics calculators well. Professional development support in the form of courses, time to attend them, and associated materials are all needed. (E.g., Andrews & Kissane, 1994) Teachers need to have a personal graphics calculator and some time to become familiar with it before they will be confident to use one in their classrooms; there seems to be no effective substitute for this personal experience. Curriculum development to take better account of the potential of graphics calculators is also needed, some examples of which are hinted at earlier in this paper. It is unwise to leave all use of the graphics calculator until the final year or two of secondary schooling. Suitable student text materials, advice for teaching in an environment that includes ready access to graphics calculators, and an assessment environment that accommodates graphics calculators well are also crucial.

This is a formidable list, which helps to explain why educational change is so hard to bring about. But there is scope to learn from each other about many aspects of curriculum development, and a redirection of some energies and resources away from microcomputers towards the more appropriate technology of graphics calculators may help to bring about change. We are well reminded, in trying to bring about change of a recent suggestion by Leinwand (1994):

> It is unreasonable to ask a professional to change much more than 10 percent a year, but it is unprofessional to change by much less than 10 percent a year. (p. 393)

CONCLUSION

Graphics calculators are small portable computers with built-in mathematical software. It is restricting and unwise to think of them as calculators. Their potential for school mathematics education is not restricted to their graphic capabilities. While they make good mathematical tools for students, there are other appropriate metaphors for them. Accessibility of technology is critical, but means more than accessibility to the physical device. Educational environments designed and nurtured to take good educational advantage of the potential of graphics calculators in curriculum, teaching and assessment are needed. The issue of accessibility to technology suggests a need for more thoughtful allocation of resources for school mathematics. Graphics calculators represent a more appropriate technology for mathematics education than do microcomputers. While this seems to be the case generally, it seems especially so for less affluent nations, including developing nations in the South-East Asian region. Finally, to gain maximum educational benefit from this form of intermediate technology, careful attention to the needs of classroom teachers of mathematics is required.

REFERENCES

Andrews, T., & Kissane, B. (Eds.). (1994). *Graphics calculators in the classroom.* Adelaide: Australian Association of Mathematics Teachers.

Bradley, J. (1995). *Technology and assessment – An international experience.* Paper presented at the 5th Annual Australian Bridging Mathematics Network Conference, Batchelor, NT.

Bradley, J., Kemp, M., & Kissane, B. (1994). Graphics calculators: A (brief) case of technology. *Australian Senior Mathematics Journal. 8*(2), 23-30.

Dunham, P. H., & Dick, T. P. (1994). Research on graphing calculators. *Mathematics Teacher, 87*(6), 440-5.

French, D. (1993). Automated algebra: What is there left to learn? *Micromath, 9*(3), 18-19.

Hackett, P., & Kissane, B. (1993). The graphics calculator: A hand-held statistics package. In T. Herrington (Ed.), *New horizons, new challenges,* (pp. 131-143). Perth: Australian Association of Mathematics Teachers.

Heid, M. K. (1995). *Algebra in a technological world.* Reston, VA: National Council of Teachers of Mathematics.

Kemp, M., & Kissane, B. (1995). Graphics calculators and assessment. In A. Richards (Ed.), *FLAIR: Forging links and integrating resources,* . Darwin: Australian Association of Mathematics Teachers.

Kennedy, D. (1994). A tale of two CD's. *The American Mathematical Monthly, 101*(7), 603-608.

Kissane, B. (1995a). Calculus and the graphics calculator. *Australian Senior Mathematics Journal, 9*(2).

Kissane, B. (1995b). How to solve an equation. *Australian Mathematics Teacher, 51*(3), 38-42.

Kissane, B. (1995). Technology in secondary school mathematics—the graphics calculator as personal mathematical assistant. In Hunting, R., Fitzsimmons, G., Clarkson, P. & Bishop, A. (Eds.) *Proceedings of the International Commission on Mathematics Instruction Conference on Regional Collaboration.* (pp 383-392) Melbourne: Monash University.

Kissane, B., Bradley, J., & Kemp, M. (1994). Graphics calculators, equity and assessment. *Australian Senior Mathematics Journal, 8*(2), 31-43.

Kissane, B., Kemp, M., & Bradley, J. (1995). Student reactions to the use of graphics calculators. In B. Atweh & S. Flavel (Eds.) *Proceedings of 18th Annual Conference of the Mathematics Education Research Group of Australasia* (pp 366-372) Darwin, NT: MERGA.

Leinwand, S. (1994). Four teacher-friendly postulates for thriving in a sea of change. *Mathematics Teacher, 87*(6).

Lowe, I., Willis, S., Grace, N., & Kissane, B. (1994). *Access to algebra Book 3.* Melbourne: Curriculum Corporation.

Lowe, I., Willis, S., Kissane, B., & Grace, N. (1994). *Access to algebra Book 4.* Melbourne: Curriculum Corporation.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: NCTM.

Romberg, T. A. (1992). Problematic features of the school mathematics curriculum. In P. W. Jackson (Ed.), *Handbook of research on curriculum,* (pp. 749-788). New York: Macmillan.

Rubenstein, R., Schultz, J. E., Senk, S. L., Hackworth, M., McConnell, J. W., Viktora, S., Aksoy, D., Flanders, J., Kissane, B., & Usiskin, Z. (1992). *Functions, statistics and trigonometry.* Glenview IL: Scott, Foresman.

Schumacher, E. F. (1974) *Small is beautiful: A study of economics as if people mattered.* London, Sphere Books.

Waits, B. K., & Demana, F. (1992). A case against computer symbolic manipulation in school mathematics today. *Mathematics Teacher, 85*(3), 180-183.

Some Calculator Activities to Develop Children's Number Sense

by

Dr. Ramakrishnan Menon, Division of Mathematics,
School of Science, National Institute of Education,
Nanyang Technological University, 469 Bukit Timah Road,
Singapore 1025. Fax: 4698952
e-mail: MENONR@NIEVAX.NIE.AC.SG

Abstract

Six calculator activities designed to develop children's number sense through a search for patterns and mental arithmetic skills are described.

Most people equate technology in mathematics with the computer, little realising the potential of the humble, inexpensive four-function calculator to help primary school children become more competent in mathematics. In this paper, I will describe some calculator activities that can develop children's number sense. But before that, let me outline a number of advantages in using the calculator, as well as clarify what I mean by "number sense."

Advantages of using a calculator

First of all, it is much more interesting than laboriously writing numbers and computing answers using finger-counting. Secondly, it gives children a sense of power, of initiating, controlling and accomplishing something, as they are the ones who press the keys on the calculator to get the fascinating, almost instantaneous, resulting display. Thirdly, pupils can generate a series of numbers in a short time and be guided to discover patterns. Fourthly, problems involving computations with real-life numbers can be attended to, without having to sacrifice realism for easy-to-compute numbers. Finally, pupils will tend to ask questions and explore topics such as decimal fractions and negative numbers, even before being taught those topics formally.

Number sense

What is number sense? The Standards (NCTM 1989, p. 38) characterise it as:

1. understanding the meaning of numbers
2. having an awareness of multiple relationships among numbers
3. recognizing the relative magnitude of numbers
4. knowing the relative effect of operating on numbers, and
5. possessing referents for measures of common objects and situations in the environment

Greenes, Schulman and Spungin (1993) state that number sense is a "sound understanding of relationships among numbers, their uses, and problem context"

(p. 279). For the purposes of this paper, I am taking number sense to mean an intuitive "feel" for numbers, about the appropriateness of numerical answers, as well as flexible, meaningful and efficient methods of computing.

Number sense and mental arithmetic

Because many confuse mental arithmetic with arithmetic drill, I will differentiate the two, before discussing number sense and mental arithmetic. Mental arithmetic is the solving of arithmetic problems/sums mentally, without recourse to paper and pencil or other manipulatives such as fingers or the calculator, while arithmetic drill is the repetitive practice of number bonds and the like under strictly controlled, short periods of time. While the purpose of arithmetic drill is to get fast, accurate answers from memory, the purpose of mental arithmetic is to get answers, preferably speedily, but more importantly, to get answers that consciously use mathematical links. While drill emphasizes the product (answer), mental arithmetic emphasizes both the process and the product, as not only is the answer given but how the answer (product) was arrived at should be explained as well. According to Sowder and Kelin (1993), mental arithmetic skills enhance number sense. This is so because of the various links that have to be made in order to mentally arrive at a sensible answer. For example, to find 50% of $40, the student who relates 50% to 1/2 and then computes 1/2 of $40 mentally shows more number sense than one who computes it by writing 50% of $40 as 50/100 x $40, and proceeding algorithmically to "cancel" etc. Although Sowder and Kelin (1993) list three types of mental arithmetic activities, in this paper I focus on only the skill of mental computation. Specifically, I will describe six calculator activities that can develop such mental computation skills and hence enhance number sense. The activities are:

1. adding/subtracting tens.
2. adding/subtracting nines.
3. building the five and ten times tables.
4. multiplying and dividing by tens.
5. subtracting by keeping the difference constant.
6. the "look before you leap" activity

Mental computation

By mental computation, I mean computing, mentally but accurately, answers to numerical problems. That is, there are three criteria for a mental computation activity, namely mental (as opposed to written), accuracy (as opposed to estimation) and numerical problems (as opposed to word problems). Some examples are, to mentally compute 42 + 9, 42 -11, 93 - 37, 6000 - 3459, and 432 x 10.

What are the advantages of mental computation over written computation? Researchers (e.g. Plunkett, 1979; and Sowder & Kelin, 1993) state that while written algorithms are efficient, they seldom correspond to the way people think, and are inappropriate for mental computation. As well, mental algorithms are specific to the problem at hand, and require a good understanding of numbers and relations between numbers (i.e. number sense).

Researchers (e.g. Hope & Sherrill, 1987; and Markowits & Sowder, 1988) also state that skilled mental computers use a variety of strategies quite different from the written algorithms learned. For example, they generally work from left to right and use number properties and relations which indicate an understanding of the structure of the number system, thereby evidencing number sense.

Calculator activities

Adding/subtracting tens

Adding tens is a calculator activity well-suited to Primary 1 children. The teacher asks the students to input a two-digit number, say 23. Then the students are to input +10, and state the display on pressing the = sign. (It would be good to have enough calculators, at least one calculator per pair of students). Then the students are asked to repeat the process a number of times. Each time the students read out the number displayed on the calculator, the teacher writes the number on the board, and soon the following series of numbers would have been written on the board by the teacher: 23, 33, 43, 53, 63, 73, 83, 93. The teacher should use a different colour to highlight the tens digits and guide the students to see an emerging pattern of increasing numbers in the tens place, while the ones digits remain unchanged.

The teacher could then lead a discussion as to why such a pattern occurs and guide the students to generalise (by generating many different series of two-digit numbers resulting from adding ten successively) that each addition of a ten increases the tens digit by a ten, leaving the ones digit constant. In reality, even when generating one series of two-digit numbers by adding tens, the students will start shouting out the answers before inputting numbers in the calculator, as they would have an intuitive feel for the number pattern generated. This will be especially true if the teacher has kept track of the numbers generated by writing them on the board. In other words, students can *mentally add* ten to a two-digit number because of the pattern of increasing digits in the tens place. The discussion will further reinforce the idea of adding tens, by focussing on the "why" of it, thereby enhancing number sense and understanding.

This is a far cry from writing, say, 23 + 10 in the usual vertical form and saying that 3 + 0 = 1, 2 + 1 = 3 etc., where the 23 and 10 can be perceived as isolated pairs of one digit number to be added. That is, understanding place value is not a prerequisite in the column addition of such two digit numbers, as the addition could be successfully carried out by thinking of 23 as 2 and 3, the 10 as 1 and 0, and so 2 + 1 is 3 (and not necessarily 3 tens). This adding tens activity could then be extended to three and four digit numbers, such as 523 + 10 and 3526 + 10, even without a formal or extended exposure to three and four digit numbers (for example, the teacher could read and at the same time, write, "five hundred twenty three add ten," and ask the students to input the numbers etc.) A further extension would be to *subtract ten*, using the calculator, until students can mentally compute sums such as 92 - 10.

Adding/subtracting nines

Adding nines and *adding elevens* are activities related to adding tens, but since they are similar, I will describe only the former. The activity can be carried out as follows: the teacher writes on the board, say, 23 + 10 (in horizontal form), and writes down the answer (33) given by the students. Then students compute 23 + 9, using the calculator. The answer (32) is written next to the answer (33) to 23 + 10. Repeating this process, the following will be generated:

$$23 + 10 = 33, 23 + 9 = 32; 33 + 10 = 43, 33 + 9 = 42;$$

A well-guided discussion based on the pattern of answers should then lead to the generalisation that to add nine, one has to add ten and then "go back" (subtract) one. For example, 33 + 9 would be computed mentally, beginning by subvocalising "43, 42," and finally verbalising "42" as the answer. The students would not have to rely on the calculator to get an answer, once the pattern is seen and understood. Indeed, after some time, the students would not use the calculator to compute such addition, as they would realise that the answer could be obtained much faster using mental computation and number sense than by spending time "pushing buttons" on the calculator.

A similar activity could be used for subtracting nines, but it must be cautioned that this is conceptually harder, as the result is one *more* than subtracting ten (e.g. 33 - 9 is "23, 24," that is, 24), and students might find it difficult, initially, to reconcile the *adding* of one to the operation of *subtraction*, as they expect a smaller number. In the case of adding nine, students seem to find it easier to understand that as nine is less than ten, the answer is one less than when ten is added.

Multiplying

Activities similar to the ones described could be done for multiplication as well. I will describe two activities, one for the "five times" table, and another for mentally computing the product of a number and ten.

Five/ten times tables

Because the ten times table is easier and can be similarly explored, I describe only the building up of the five times table. Students can begin by entering 5, followed by repeating +5, and calling out the number displayed, to be recorded on the board by the teacher. A look at the pattern of numbers generated (the alternate fives and zeroes in the ones place), together with a discussion (led by questions such as "how many fives have you added?") showing the relationship between the repeated addition of five and the five times table can lead students to easily recall the five times table. Moreover, students can also be guided to see that, for example, six fives gives the same result as five sixes, and that too should help students compute mentally the product of, say, five nines, by relating it to nine fives as one five more than eight fives. (An additional advantage is that the five times table can then be related to the counting of five-minute intervals on the clock). It should be noted that the patterns in the 5 and 10 times table make them much easier to learn, and there is no need to always start with the two times table as is usually done. Additionally, once multiplication is seen as repeated addition with the 5 and 10 times table, students can build up their own time table for the other numbers, using the calculator and recording the results (especially interesting are the patterns for the 9 and 11 times tables), and be able to recall any specific multiplication fact by using mental computation that links multiplication, addition and commutativity, thereby making use of, and developing number sense.

Product/Quotient of a number and ten

All too often, students are just told that "when you multiply by 10, just add 0." Strictly speaking, adding zero should leave the number unchanged (e.g. 26 added to 0 gives 26), but even if the rule were given with the word "annex" substituting "add," it still remains a (teacher-given) rule. But when students use the calculator to explore the results of multiplying 10 by whole numbers like 7, 17, 24, 327, 300, etc., and the products are displayed on the board by the teacher, they very soon begin to discover the pattern for themselves, and will assume ownership of the rule. Students should also be encouraged to use the calculator to show that ten seventeens and seventeen tens are equivalent, and explain why this is so.

For students at a higher Primary level, this activity could be extended to multiplying by nines, by computing, say, 9 x 19 and comparing it with (10 x 19) - (1 x 19) and discussing why the results are equivalent. Such discussions will enable students to mentally compute sums such as 9 x 19, by going through, mentally, steps such as the following: 9 x 19, 10 x 19, 190--minus 20, 170, plus 1, equals 171. Another extension would be to explore the products of 10 and decimal fraction numbers, thereby leading students to have better number sense of place value. Division by 10 can be explored similarly.

Subtracting by using constant difference

Subtracting by using the idea of *constant difference* is also amenable to a calculator activity. For example, 94 - 36 can be computed by computing 98 - 40, as the relative differences are the same. This can be achieved by computing 94 - 36 and 98 - 40--which is (94 + 4) - (36 + 4)--on the calculator, comparing the answers, generating a similar series and discussing why the answers happen to be the same. Subtraction with renaming across zeroes, a notoriously difficult operation, can then be shown as an extension of this idea of constant difference. For example, 6000 - 3459 could first be computed by directly inputting the numbers in the calculator. Then 5999 - 3458--which is (6000 - 1) - (3459-1)--could be computed on the calculator and the answers compared. A series of such examples, recorded on the board and discussed should lead to students being able to mentally compute the answers to such numerical problems, by realising that as long as the differences do not vary, the answer can be obtained by suitably modifying the original numbers--a good example of number sense in action.

The "look before you leap" activity

Many people use calculators for the most trivial of computations, such as 24 x 10, so the activity outlined below will hopefully sensitize students to the need to decide whether some computations may be more efficiently done mentally than by mindlessly inputting numbers in the calculator. The activity, modified from the one of that same title (p. 41, William, 1992), consists of a series of sums--some calculator-appropriate, others more appropriate for mental arithmetic--which are flashed to the students. Students have to get the answer as quickly as they can. In so doing they will realise that some can be computed faster mentally than by using the calculator.

For example, a set of flash cards with the following sums 13 x 14, 14 x 13, 13 x 1, 13 x 17, and 13 x 0 can be flashed, one at a time, and students asked to call out the answers as quickly as they can, using the calculator if they want to. Initially, students might use the calculator for every sum, but when some students, say, call out the answer to 13 x 14 (and for 13 x 1 and 13 x 0) even before the others have entered the appropriate numbers and operations into their calculator, students begin to appreciate that sometimes it is better to use mental computation and number sense than to rely on the calculator.

Just as for any other mental computation activity to be successful, for this activity, too, the teacher has to set aside time for discussing how the sums were computed. In the above example, the answer to 14 x 13 is obtained mentally because commutativity can be applied to the 13 x 14 which was computed using the calculator. Similarly, an awareness of the multiplicative property of 1 (and 0) would render the use of the calculator for 13 x1 (and 13 x 0), unnecessary. This activity can be suitably modified to meet the needs of different levels (e.g. Primary 2 or Primary 5) by focussing and revising specific properties of numbers and operations.

Conclusion

In this paper, I have shown six calculator activities designed to help develop mental computation skills and hence number sense. I have emphasized that students should be made aware that using the calculator does not always give answers faster or more easily than by using mental computation and number sense. Finally, I would like to say that I cannot overemphasize that discussion should be part and parcel of such activities, if these activities are to be optimally useful to help develop students' number sense, which is critical to learning mathematics meaningfully.

References

Hope, J. A., & Sherrill, J. M. (1987). Characteristics of unskilled and skilled mental calculators. *Journal for Research in Mathematics Education*, **18**(2), 98-111.

Markowits, Z., & Sowder, J. (1988). Mental computation and number sense. In M. J. Behr, C. B. Lacampagne, & M. M. Wheeler (Eds.), *Proceedings of the Tenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 58-64), DeKalb, IL: Northern Illinois University.

Plunkett, S. (1979). Decomposition and all that rot. *Mathematics in Schools*, **8**(3), 2-5.

Williams, D. E. (1992). *Creative mathematics teaching with calculators*. Mountain View, CA: Stokes Publishing Co.

A SURVEY INTO THE USE OF MICROCOMPUTERS IN THE TEACHING OF MATHEMATICS IN PENANG SECONDARY SCHOOLS, MALAYSIA.

Munirah Ghazali
School of Educational Studies
Universiti Sains Malaysia
11800, Penang Malaysia.

Introduction

Much has been written about the roles of computers in the mathematics classroom (NCTM, 1980; Cockcroft, 1982; Tall 1986; Ediger, 1989; Ganguli, 1990; Kalman, 1994) as well as other subjects. Malaysia too is heading towards implementing the use of computer in the curriculum. It was recently reported that the Ministry of Education is devising a book-free school system (The Star, 1995) in which computer disks will replace textbooks.

Cockcroft (1982) lay out two ways in which the impact of the computers in the mathematics classroom will be felt. The first concern the ways in which microcomputers can be used to assist and improve the teaching of mathematics in the classroom. The second concerns the extent to which the availability of microcomputers in the mathematics classroom should change the content of what is taught or the relative stress which is placed on different topics within the mathematics syllabus. Professional bodies of mathematics educators too recognize the fact that computers will have a tremendous impact on the mathematics curriculum, both in content and in pedagogy, and therefore steps should be taken to prepare for this onslaught.

Purpose

The main purpose of the survey is to find about the use of microcomputers in the mathematics classrooms in Penang secondary schools. The objectives of the survey are to determine:

1. The availability of microcomputers in schools

2. The availability of microcomputers for teaching and learning of mathematics.

3. The types of software available for the purpose of using microcomputers in the mathematics classrooms.

4. The type of support teachers received for using microcomputers in the mathematics classroom.

5. The major problems and issues faced by teachers regarding the use of microcomputers in the teaching of mathematics.

Method

A survey was conducted to obtain data from 20 secondary schools in Penang, Malaysia. Mathematics teachers were requested to complete a questionnaire containing items about the availability of microcomputers in the schools, information about computer usage, the types of software available and the form of support or training the teachers received in the use of microcomputers in the mathematics classroom.

Subject

A total of 20 schools participated in the survey. However out of the 20, only 9 schools reported that they have computers in the school. 11 schools reported that they do not have computers in their school and therefore could not respond to the survey.

Results

The results were obtained from the 9 schools that responded. The responses are tabulated as follows:

i. Information about computers available in the schools surveyed.
ii. Information about computer usage for selected Penang secondary schools
iii. The types of software available for selected Penang secondary schools
iv. Government agencies that offer support to teachers in using computers in the mathematics classroom

Table 1. Information about computers available in selected Penang secondary schools.

School	Number of computers available	Types of computer	Speed	year acquired
A	5	IBM compatible	286	-
B	11	IBM compatible	386&486	1993
C	14	IBM compatible	386	94/95
D	8	IBM compatible	286&386	1986
E	15	IBM compatible	XT, 286&386&486	-
F	6	IBM compatible	XT, 386&486	1989-1994
G	15	IBM compatible	286&386	1992
H	4	IBM compatible	286,386&486	1995
I	4	IBM compatible	286,386&486	93,94&95

Table 1 shows information about computers available in selected Penang secondary schools. Out of the 20 secondary schools surveyed, 9 have computers in their school. The least number of computers available was 4 and some schools have up to 15. All are of IBM compatible types either of XT, 286, 386 as well as 486 speed. Almost all of the computers were acquired quite recently i.e. in the 1990's. Most of the computers were obtained through the efforts of the Parents Teachers Association (PTA) and through the schools own effort. None of the schools surveyed received the computers from the Ministry of Education. This is probably due to the fact that the Ministry of Education's priorities are for schools in the rural areas.

Table 2. Information about computer usage for selected Penang secondary schools

School	computer club activities	computer literacy class(as a co-curricular activities)	computer literacy as a school subject	computers in the mathematics classroom
A	yes	yes	no	no
B	yes	yes	yes	no
C	yes	yes	no	no
D	yes	yes	no	no
E	yes	yes	no	no
F	yes	no	no	no
G	yes	yes	yes	no
H	no	no	no	no(use for school administration)
I	yes	yes	no	no

Table 2 gives the information about computer usage in schools. All but one schools reported that they use the computers for their computer club activities and also for computer literacy classes. Both are co-curricular activities. Only two schools have computer literacy as a school subject while the rest of the schools surveyed do not offer such facility. Meanwhile, it is rather disappointing to report that none of the schools surveyed actually use computers in their mathematics classroom. Perhaps one of the reason could be found in table 3 that is none of the schools surveyed have software or CAI packages that deal specifically for the teaching of mathematics.

Table 3. The types of software available for selected Penang secondary schools

School	Word Processing	Spreadsheet	Database	Graphic package	Programmin g language	CAI for mathematics
A	yes	yes	yes	yes	no	no
B	yes	yes	yes	yes	no	no
C	yes	yes	yes	yes	BASIC	no
D	yes	yes	yes	yes	no	no
E	yes	yes	yes	yes	BASIC	no
F	yes	yes	yes	yes	no	no
G	yes	yes	yes	yes	no	no
H	yes	yes	no	no	no	no
I	yes	yes	yes	yes	no	no

Table 3 shows the types of software available for selected Penang secondary schools. It is interesting to note that all the schools reported that they have either word processing or spreadsheet packages. Meanwhile, all except one schools reported that they have graphic packages. There are two schools that also have BASIC as a programming language installed in their computers. As stated above, none of the schools have CAI packages for mathematics. Consequently the teachers thought that there is no place for microcomputers in their mathematics lesson.

Table 4. Agencies that has been involved in training to teachers to use computers.

School	your school	local education authority	State education department	Ministry of education	others
A	no	no	no	no	no
B	yes	yes	no	no	no
C	no	no	no	no	no
D	no	no	no	no	no
E	no	no	no	no	no
F	no	no	no	no	Local teachers' resource center
G	no	yes	yes	no	no
H	no	no	no	no	no
I	no	no	no	no	no

Table 4 shows agencies that has been involved in training to teachers to use computers. Out of the 9 schools , 6 reported that they have no training in the use of computers from any of the agencies stated in the table above. It is assumed the rest of the teachers get the skills from their own personal efforts. Two schools reported that they have help from either their own school, local education authority or the State education department. Meanwhile, one school reported that they were given help by the local teachers' resource center.

Discussion

It would be inappropriate to discuss the results the survey above without first considering some information about the Computers in Education Unit. This unit was set up by the Ministry of Education to carry out three major activities:

i. Pilot study in the teaching and learning of computer literacy.

This project began in 1992 and involved 60 secondary schools from the rural area. At the moment three rural schools from the State of Penang has been chosen. All three are located on the mainland. Each school were supplied with 21 PC's that are served on LAN. The software supplied were similar to those that were available to the Penang secondary schools. The software packages available are DRDOS, PowerBASIC, Wordperfect,Lotus 123, dBase IV and Drawperfect.

ii. Pilot study in the teaching and learning of CAI - in all school subjects with priority given to the use of CAI in Mathematics and English.

This project began in 1993 and was aimed at primary schools in rural areas in the state of Selangor only as a starting point. The schools were similarly provided with computers set up as described in activity (i) above.

iii. Plan and carry out pilot project in educational networking (Jaringan Pendidikan).

At the moment 4 schools in the Federal Territory area has been earmarked for this project. Three schools will be chosen from the State of Penang to participate in this project.

It is also the aim of the Ministry of Education that by the year 2000, 60% of both primary and secondary schools in the country will be equipped with at least one computer laboratory with 21 computers that will fulfill the three activities outlined above.

Perhaps the small sample size of schools taking part in this survey might not represent a true picture of the use of microcomputers in the classrooms for the secondary schools in Penang, Malaysia. However, from the survey and also the discussion above, it can be gathered that there are serious efforts both from the schools and the Ministry of Education towards using microcomputers in the classrooms, mostly towards computer literacy rather than for the use of computers in any particular subject area such as mathematics.

From table 2, it could be concluded that at the moment, none of the schools surveyed is using microcomputers in their mathematics classroom. This is rather unfortunate considering changes that are taking place in the mathematics classrooms in other parts of the world. One reason might be lack of facilities such as computers etc. However, it could be seen from table 1, even schools with 15 computers may not be using them in their mathematics classroom. As discussed before, perhaps one of the factors may be lack of suitable mathematical software as could be seen in table 3 where none of the schools surveyed reported having CAI packages for mathematics.

From table 3 and from activity (i) of the Computers in Education Unit under The Ministry of Education, it could be concluded that almost all the schools that have computers (whether funded by the schools or the Ministry) have spreadsheets packages.

Conclusion

Current literature on the use of microcomputers in the mathematics classroom is littered with various ways of using spreadsheets in the mathematics classroom (Arganbright, 1984; Catterall and Lewis, 1985; Corkill and Robinson, 1988; McDonald, 1988; Dye, 1994; Ainley and Pratt, 1994). These literature has suggested various ways in which spreadsheets packages could be used in the mathematics classroom as well as suggesting that the use of a spreadsheets may even help alleviate some of the misconceptions associated with some mathematical concepts (Catterall and Lewis, 1985; Poletti, 1988). It is about time that the mathematics teachers in Malaysia harness the spreadsheets packages that is available rather than wait for some CAI packages.

As an example of one way of using spreadsheet in the mathematics classroom is an investigation of the maxbox problem by using a spreadsheet package carried out by the writer with a group of 16 year old students. The maxbox problem states that:

An open box is made from a 25 cm by 25 cm square of card by cutting a small square from each corner and then folding up the edges. What is the greatest volume that the box can hold? If the box is to have a volume of $847cm^3$, how big should the small squares be ?

It is not my intention to elaborate the spreadsheet's procedure in this paper. However, I would like to quote some of the students comments after using the spreadsheet's procedure to solve the above problem.

" *Initially I differentiated the formula we derived for the volume, but on using the spreadsheets it became simpler, if less accurate. I was a bit afraid to use the spreadsheets as it seemed overcomplicated and lengthy, but when I started the problem resolved very quickly. The worst thing was having to stop as I was grasping the problem.*"

(Student D)

" *In the differentiation method, it was quicker and more accurate to use standard methods then using spreadsheets, although it wasn't as abstract a problem when the spreadsheets was used.*"

(Student E)

" *The method is a very interesting/fascinating way of finding the answer. I found it more exciting than straight forward differentiation. This was probably because the situation was more practically sought out using the spreadsheets.*"

(Student C)

References

Ainley, J. and Pratt, D (1994) Unpacking Maxbox. Micromath Vol 10 no. 3 pp14-19.

Arganbright, D.E. Mathematical Applications of an Electronic Spreadsheet. In Computers in Mathematics Education, 1984 Yearbook of the National Council of Teachers of Mathematics, pp 184-93, Reston, Va.

Catterall, P. & Lewis, R. (1985) Problem Solving wiht spreadsheets, Journal of Computer Assisted Learning, Vol.1 No.3 pp. 167-169.

Cockcroft, W.H.(1982) Mathematics Counts, Report of the Committee of Inquiry into the Teaching of Mathematics in Schools. HMSO, London.

Corkill, D. and Robinson,N. (1988), Using Spreadsheets to solve a Mathematical Problem, Micromath, Vol.14 No.1 pp.13

Dye, B. (1994). 1,2,3,4: four digits that dwarfed the universe. Micromath Vol 10 no. 3 pp12-13.

Ediger, M. (1989). Microcomputer in the mathematics curriculum. Journal of Computers in Mathematics and Science Teaching, 9(1), pp45-50

Ganguli, A.B. (1990). Microcomputer as a Demonstration Tool for Instructional in Mathematics. Journal for Research in Mathematics Education. 21(2), pp 154-159.

Kalman,R. (1994) Future Classrooms: A personal Vision. The Mathematics Teacher, Vol87, No.7 pp. 486-487.

McDonald, J.L (1988), Integrating Spreadsheets into the mathematics classroom. The Mathematics Teacher, Vol.81 No. 8 pp. 615-622.

Munirah, G.(1991), An investigation into the introduction of integration as the limit sum of areas under a curve by using a spreadsheet (Multiplan). Unpublished M.Ed dissertation, University of Birmingham.

National Council of Teachers of Mathematics (1980). An Agenda for Action-Recommendations for Schools Mathematics of the 1980's. Reston, Virginia: NCTM.

Poletti, C.D, (1988), An investigation of the potential of the electronic spreadsheet as a medium for introducing early algebraic concepts. Unpublished M.Ed dissertation, University of Birmingham.

Tall, D.O. (1986) A graphical approach to integration. Mathematics Teaching, No. 114 pp. 58-51

ALGEBRAIC MANIPULATION LANGUAGES IN PROBLEM SOLVING

A Case study with Maple-V

P.M.Nobar,
Mathematics and Science Centre,
Ngee Ann Polytechnic,
Republic of Singapore

J.J.Williams
Civil Engineering Department,
Queen Mary and Westfield College,
University of London,
United Kingdom

ABSTRACT

The problem of designing an appropriate mathematics curruculum and methods of effective implementation of its syllabi have always been a matter of open discussion in all educational institutions. The main reason is that the problem is complicated and has a large numbers of variables. Moreover, some of these variables are continuously changing as society changes with time. Above all, the problem involves learning media and the personality of educators. Although a large number of the variables in the design process are the same, so far no acceptable solution has been produced in any country. There are still sensitive variables which make the implementation of a successful system in one country impossible in another country. In this paper the authors initially review the most important factors in designing a mathematics curruculum and proceed to set out the most common ways of teaching problem solving in mathematics. The review makes the contribution of using algebraic processing languages such as Maple-V in teaching mathematics more appropriate. The paper also highlights the complementary aspect of algebraic processing languages with numerical analysis software. Finally, the potential of using algebraic processing languages in Singapore polytechnics has been investigated.

Key words:
Mathematics curricula, Psychological demands, behaviouristic theories, Holistic theories, Sociological demands, Structural demands, Instructional methods, Expository approach, Inquiry approach, Pure learner oriented approach, Evolution of mathematics curricula, Meaning versus skill, Problem solving, What is a problem, The process of problem solving, Intuition, Polya's model, Lester's model, Factors in problem solving, Pattern matching, Adjunct Questions, Advanced organizers, Learning hierarchies, Teaching tool skills, Teaching heuristic strategies, Algebraic processing languages, Maple-V

1-Introduction

Mathematics occupies a well established position in the education system of any country. While there is wide acceptance of the importance of mathematics as a topic in an education system, there is also a lack of consensus regarding the content of the curriculum in different countries and the overall objectives for its study. In general:

- The public expresses concern about the mathematical competence of learners, which causes educators to examine the achievement of learners more closely.
- Mathematics educators are concerned about the limited treatment of content that they view as important for learners.
- The curriculum developers are concerned about the fit between knowledge of the subject matter and their view of the overall objectives of education.

2-Mathematics curricula

The selection of the mathematical content in any education establishment is based on three different criteria: psychological, sociological and structural. Ideally these categories play an equal and complementary role in the design of curricula. However, from time to time, one of the three demands seems to influence curriculum developers in mathematics to a greater extent than the others. For example, during times of economic pressure, the educators lean toward sociological demands as a means of justifying the curriculum.

Psychological demands

The selection of mathematical content falls under the influence of two broad psychological theories:

a-behavouristic theories

These theories view the learner as mastering pieces of mathematical ideas that collectively produce the whole learning. In other words the goal of drill instruction is to subdivide the mathematical content into segments of knowledge that must be mastered in a sequential order for the learner to succeed.

b-Holistic or field theories

These theories view the learner as comprehending the entirety of learning usually through insight or occasionally through sudden inspiration. Following this theory, the educator has to provide the fabric that spreads learning and provides a cognitive structure.

Most educators subscribe to an eclectic point of view, believing that certain kinds of mathematical problems such as solving quadratic equations should be dealt with using a behaviouristic theory; whereas other mathematical problems such as understanding the geometrical proofs should be guided by a holistic theory. Namely, in teaching an algorithm, it is important to identify the sequential steps of the process and to have learners master these steps. While in teaching geometrical proofs, it is important to provide a selection of viewpoints for examining the problem with the objective that one of these viewpoints will lead to a solution.

Research studies conducted during the 1930's and 1940's attempted to test the effectiveness of these two theories. In the beginning it seemed that drill (practice) and meaningful instruction were opposite to each other, but later it became apparent that the two procedures are in fact complementary and that, the drill should be preceded by meaningful instruction

Sociological demands

Sociological demands in developing a curriculum stress the individual's needs for a certain mathematical content. According to this, a learner must learn a kind of mathematics which will satisfy certain vocational demands such as those required by mathematicians, physicists, engineers and technicians in different fields. In general, psychological theories involve academic questions that educators and curriculum developers can address in the laboratory, while sociological demands are more diverse to identify and to provide for in the curriculum.

Structural demands

The structure of mathematical science is often the main guide for developing a curriculum. During the modern mathematics movement in most of the European countries and the United States, the structure of mathematics was the main criterion in the development of curricula in mathematics. Most modern mathematics courses were concerned with structural questions such as What is a number?, What are the axioms of rational numbers? What is a variable? What is a function? What is a group? All these questions were aimed at understanding the structure of the discipline.

3-Instructional methods

The term instructional strategies is occasionally used in place of the more traditional term, instructional methods. A strategy is the educator's approach to using information, selecting resources and defining the role of the learners. It includes specific practices used to accomplish a teaching objective. An instructional method is defined as a systematic plan for presenting information. Instructional methods can be divided into three categories:

Expository approach

In this category educators have full control of the classrooms in terms of the sequencing of events. The curriculum designed for this type of educators, provides an outline of the topics, but leaves the manner of presentation and class activity to the discretion of the educators. Moreover, problem sets and exercises are also provided but the selection of them are left to the educators. This type of instructional procedure is most common at educational institutions in European countries and the United States. It is the most efficient and practical form of instructional procedure. The source of information most frequently used are the textbook and other reference materials and audiovisual materials.

Inquiry approach

In the inquiry or discovery mode, the educator assumes the role of facilitator of learning experiences and arranges conditions in such a manner that learners raise questions about a topic or event. In other words, the educator attempts to engage learners in the learning activities. Unlike the previous method, the sequence of events and the manner of presentation are also included in the curriculum. Examples of resources are textbooks, documents, statistical data, films or slides. and Computer Based Learning (CBL) programs. In this method, learners raise questions about the content of the materials and attempt to organize this information. The learners are active participants as they develop hypotheses which are later tested by use of additional data. This method is more effective than the previous one, but requires more human and financial resources.

Pure learner oriented approach

In this method the learner develops a curriculum according to her/his own needs. The educator's role is one of counselor and provider of initial directions. The method has been used in only few colleges, primarily at the graduate level. The curriculum is highly flexible in a sense that the individual freedom and creativity are emphasized. This method requires extremely competent and mature educators.

4-Problem solving

Anyone who has studied mathematics at any level knows that learning to solve problems is the principal reason for studying mathematics. However, because of the complex nature of problem solving its role in educational institutions is less clear. The importance of problem solving in mathematics can be demonstrated by a large number of text books, formal reports and journal articles. In addition, the problem solving has been the focus of several conference proceedings. Henderson[1] provides a set of criteria for a *problem* concerning an individual:

•The individual has a clearly defined set of objectives which s(he) is consciously aware and whose attainment s(he) desires.

•Blocking of the path toward the objectives occurs and the individual's fixed pattern of behaviour or habitual responses are not sufficient for removing the block.

•Deliberation takes place. The individual becomes aware of the problem, defines it, identifies various solutions and test them for feasibility.

The above criteria implies that, the individual must be aware of the *existence of a situation* that needs a solution, s(he) must have an *interest* in finding a solution, a procedure for obtaining the solution must not be immediately available and the individual must consider the problem in order to develop a clear understanding of the problem and how to proceed to find a solution to the problem. The summary of these criteria is shown in Fig-1. By the above definition, a mathematics problem is simply a problem for which the solution involves the use of mathematical concepts and skills. A common question is how can a situation be a problem for one person and not for another? To answer this consider the following example.

If $\sin A=1/2$, $\sin B=1/3$ and both the angles are acute, determine $\sin(A+B)$.

This problem poses different levels of difficulties depending upon the problem solver's mathematical abilities:

•For an individual who is not familiar with the trigonometric functions, this is, a

Aware of the situation
Interested in resolving situation
be unable to proceed directly
Make a deliberate attempt to find a solution

Fig-1-Criteria for having a problem

formidable problem.

•For a person who knows about the trigonometric functions but is not familiar with the trigonometric identities

$$\cos\theta=(1-\sin^2\theta)^{1/2},\quad \theta\leq\pi/2^c$$

$$\sin(\alpha+\beta)=\sin\alpha\cos\beta+\cos\alpha\sin\beta$$

the problem is complicated.

•For a person who is familiar with the above identities the solution reduces to a simple substitution. That is, this poses no problem to that person.

•For an individual who has no interest in this area, this also poses no problem. A common attitude toward a problem like this one is indifference. A reaction like "Who cares what is the value of sin(A+B)?

5-The process of problem solving

In general the process of problem solving involves: coordination of knowledge, previous experience, intuition, various analytical abilities and various visual abilities. In some contexts, intuition is referred to as a source of true knowledge. It is generally in this sense that the term intuition is used in the works of Descarts[2]. For him, in a world of misleading appearances and futile interpretations, intuition remains the ultimate reliable source of certain truths. Bergson[3] advocates that intuition as a method is a sort of mental strategy which is able to reach the essence of phenomena. Bergson[3] argues that intelligence addresses itself to the world of objects, solids, static realities. Namely, the uninterrupted flow of real phenomena is cut into a sequence of static representations mainly expressed in concepts. However, the essence of motion or life which involves duration cannot be reached this way. According to Bergson[3] it is through intuition that one is able to grasp the essence of living and changing phenomena. The term intuition is used by Beth[4] for indicating a certain category of cognitions, i.e., cognitions which are directly grasped without any need for explicit justification or interpretation. Very often intuition means an elementary common sense, i.e. a popular and a primitive form of knowledge as opposed to scientific conceptions and interpretations. In contrast, according to some philosophers like Spinoza[5], intuition is the highest form of knowledge through which the essence of things may be revealed. According to Poincare[6], no genuine creative activity is possible in science and in mathematics without intuition. On the other hand, Hahn[7] argues that intuition is mainly a source of misconception and should be eliminated from a serious scientific study. In the pedagogical literature, intuition is often related to sensorial knowledge as the first necessary basis for a further intellectual education. Sometimes people use the term insight for indicating a sudden, global rearrangement of data in the cognitive field which would allow a new solution in the given conditions. The terms revelation and inspiration are also used as synonymous with intuition. Also, very often, common sense, naive reasoning, empirical interpretation are used in reference to intuitive knowledge.

The most perplexing aspect of the problem solving is that two individuals can obtain the same solution using different but correct methods. This characteristic of problem solving makes it difficult to decide on the best procedures useful for its teaching. A popular model of problem solving has been proposed by Polya[8,9]. This model has four distinct phases through which a successful problem solver goes: understanding the problem, devising a plan, carrying out the plan and looking back. Although the model is not much help in specifying the mental processes involved in successful problem solving, it can be valuable in problem solving methodologies. Other models that attempt to explain problem solving behaviour in terms of cognitive process have also been developed by information processing researchers such as Newell[10]. However, the effort of most of the researchers have focused on problems involving puzzles. These types of problems do not represent the majority of problems that confront mathematics learners. An alternative model for problem solving was developed by Lester[11].

1-Problem awareness

A situation has been posed for the learner.Before the situation is regarded as a problem:

(i)-The learner must realize that a difficulty exists in a sense that the situation cannot be readily resolved. This recognition often follows from initial failure to attain a goal.

(ii)-The learner must show willingness to try solving the problem.

If the learner neither recognizes a difficulty nor shows willingness to try solving the problem, there no point to proceed any further.

2-Problem comprehension

Once the first criteria has been satisfied, the problem solver begins the task of making sense out of the problem. This step involves:

(i)-Translation or interpretation of the provided information into terms that have a meaning for the learner.

(ii)-Internalization of the provided information, that is, the problem solver has to sort out the relevant information and determine how the information is interrelated.

At the end of this step the problem solver has an internal representation of the problem. Although the representation may not be an accurate one, but it does furnish a means of establishing objectives and priorities for working on the problem. The accuracy of the problem solver's internal representation may increase as progress is made toward a solution. Thus the degree of problem comprehension will be a factor in the remaining stages of the solution process.

3-Goal analysis

Goal analysis concerns the formulation of the problem so that familiar strategies and techniques can be applied. It can also involve identification of the different components of the problem. Namely, it is a process that moves from the goal itself backward in order to separate the different components of the problem.

4-Plan development

During this stage, the problem solver gives conscious attention to devising a plan to attack the problem. It is perhaps this stage which causes more difficulty than any other stage. It is common to hear mathematics learners proclaiming after watching their educator solve a problem "How did you think of that?". The main sources of difficulties for learners at this stage are:

a-They are prone to give up if a task cannot be done easily.

b-They are unable to devise appropriate plans because they have few plans at their disposal.

c-They are unable to work out what to do first and organize their ideas

5-Plan implementation

At this stage the problem solver tries out a plan that has been devised. The main pitfalls of this stage are that the problem solver:

a-who correctly decides to make a table and search for a pattern may fail to see the pattern due to computational errors.

b-may forget the plan or become confused as the plan is carried out.

c-may not be able to fit the various parts of the plan together.

6-Solution evaluation

Successful problem solving relies upon:

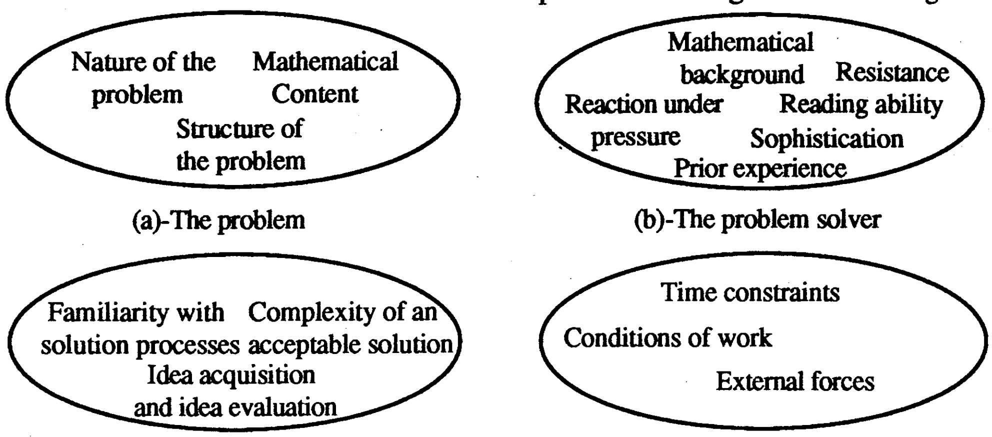
a-a systematic evaluation of the appropriateness of the decisions made during the problem solving process

b-a thoughtful examination of the results obtained

6-Factors in problem solving

Problem solving models such as Lester's model do not describe the factors that make a problem difficult or influence success in problem solving. These factors fall into four categories.

•The problem
A summary of the points which affects the problem are shown in Fig-2a.
•The problem solver
A summary of the characteristics of an individual which play an important role in problem solving are shown in Fig-2b.
•The problem solving process
A collection of the behaviuor of an individual during problem solving is shown in Fig-2c.
•The problem solving environment
A collection of features that are external to the problem solving is shown in Fig-2d.



(a)-The problem (b)-The problem solver

(c)-The problem solving process (d)-The problem solving environment

Fig-2-Factors influencing success in problem solving

7-Instruction in problem solving

A number of differing view points regarding instruction in problem solving have been proposed. The most common procedures are as follows:

1-Pattern matching

In pattern matching, the educator presents a model problem to the class and leaves the model on display as a pattern for working other problems. It is expected that, by doing a sufficient number of problems of this kind, learners memorize the pattern or the algorithm for solving a particular type of problems. However, it has two major disadvantages:

(i)-The learners are not really thinking through the problem but rather are engaged in a matching process.

(ii)-The pattern or algorithm is usually forgotten after a short while and some learners are not able to apply the matching procedure for slightly modified problems.

A more effective variant is to guide the learners by key questions designed to highlight or emphasize the essential elements of a problem and encourage the learners to consider these elements. For further details see Hatfield[12].

2-Adjunct questions

In this methodology a set of questions would follow a written paragraph. These questions are selected to focus the attention of the learners on the important features of the paragraph. So far, mathematics educators have not given sufficient attention to the possible uses of adjunct questions. Rothkopf's[13] research on 159 college learners suggests that if adjunct questions are given after reading a passage, studying these questions can have visible facultative effects on post-reading performance.

3-Prior organizers

Experienced educators usually provide an introduction before starting a new topic. The introduction consists of two phases:

1-A review which helps learners to recall the relevant principles, concepts and facts.

2-An overview which helps learners appreciate the outcome of learning the new material. Ausubel's[14] research shows that the prior organizers facilitates meaningful learning in three ways:

(i)-They mobilize relevant anchoring concepts already established in the learner's cognitive structure.

(ii)-They provide optional anchorage which promotes initial learning and resistance to later loss.
(iii)-They reduce rote memorization which learners often resort to because they lack sufficient numbers of key-anchoring ideas.
Bloom[15] and his team have tried to design an organizer to complement the cognitive structure of a learner. Bloom[15] identifies cognitive entry characteristics as specific knowledge, abilities or skills that are prerequisites for a particular learning task.

4-Learning hierarchies

The learning hierarchy is a means of organizing learning tasks to achieve a set of final objectives. The process requires describing the mathematical objective and asking what the learners should be able to do in order to achieve that objective. Answering such a question raises the necessity of defining sub-tasks and sub-tasks of these sub-tasks. By following this procedure a network of tasks will be generated. The achievement of the final task dependents upon achieving all the sub-tasks. For further details see Pillips[16].

5-Teaching tool skills

Tool-skills are those that assist the problem solver in using a strategy. In general tool-skills approach attempts to teach a few carefully chosen facilitating techniques. Subsequently, learners are asked to solve several problems for which solutions are made easier by using these facilitating techniques.

6-Teaching heuristic strategies

This method of teaching is proposed by Polya[17]. It is the most popular method which has been studied in recent years by researchers in mathematics education. This popularity stems from a general acceptance of Polya's[17] method of teaching problem solving. This method involves the presentation of the content of mathematics in the context of problems to be solved. For further details see Polya[17] and Rubinstein[18].

8-Valuable results

From several years of research on methods of instruction in problem solving some valuable results have been gathered:
a-There is a psychological organization of the objects of the curriculum, that is, an organization that arises from the analysis and study of the mathematical understanding of the learners. This organization may be quite different from its axiomatic, logical and mathematical organization. This implies that, the educators must also be aware of the psychological organization of the curriculum as well as its mathematical organization. Psychological organization has only been given serious consideration in 20th century.
b-It is impossible to fit all mathematical learning into a single model. Namely, the most sensible approach for instructing learners is some combination of different models. Certainly, the learners attitude towards problem solving is the most important factor which cannot be totally altered by using different instruction methodologies.
c-It seems that the problem solving performance will be increased if learners see competent problem solving behaviour exhibited by their educators.

9-Classification of mathematical problems

In terms of solvability, mathematical problems can be divided into two categories:
1-Problems that can be solved in a reasonable amount of time using pencil and paper.
2-Problems which require computers.
Numerical methods of analysis have been available in digital computers for the last four decades. These programs have been used extensively in solving large and more realistic problems in different fields of science. Numerical methods have two major disadvantages:
•Errors due to the limited number of digits in the storage and the execution of numerical values.
•Inability to manipulate algebraic expressions.
In the last two decades a set of algebraic processing packages have been developed which are capable of manipulating symbols as well as numbers using exact arithmetic. In addition, some of these packages also have the numerical implementation of a large number of numerical algorithms. The use of these packages have increased drastically

in the last decade in research and teaching mathematics in different science departments at universities or polytechnics. This is partially due to increase in the power of microcomputers.

10-Use of algebraic processing software

Algebraic processing languages can be used in mathematics courses in different ways:
•By designating the task of performing the tedious algebraic operations to the software, one can ask learners to explore their ideas, i.e. to concentrate on "how to do the things" without becoming too much embroiled in methods of doing.
•The results of theories can be examined in details by applying them to general problems. This can assist learners to understand the details of the theories and their implications.
•The learners can examine the assumptions of the theories and visualize their effects if they are discarded.
•It creates a practical dimension to the theoretical work, in a sense that learners can test their theoretical theorems and formulae.
•In the applied mathematical sciences, the symbolic solution of a problem can assist learners in analyzing the effects of individual properties in the final results. This type of analysis is not feasible when numerical methods are used.
•The structure of the commands can help learners in understanding the different components of the problem at hand.
•More realistic problems can be solved.
•The software can be used for mathematical modelling.
•It can enhance the computer programming capabilities of the learners.
The algebraic processing languages can be used to increase the success rate of problem solving at the plan development and plan implementation levels of Lester's[10] model. This is due to the fact, a solution can be produced with no errors. Moreover, these languages can be used to solve different models quickly, so that the problem solver can check his(her) formulation at any stage without any difficulties.

11-Use of Maple-V

Maple is a computer environment for performing mathematical operations. It is interactive and easy to use, moreover it offers a variety of support in different areas of mathematics:

•Algebraic/symbolic manipulations

This is the most important component of Maple-V, it enables users to perform extremely tedious and repetitive operations. By allowing the constant values to remain in their exact form, Maple-V provides exact answers with more accuracy than numerical approximation methods. If a floating point result is needed, it can be calculated at the end of the computation. Topics covered: *Calculus*, *Linesar algebra*, *Differential equations*, *Geometry* and *Logic*.

•Numerical computations

Numerical algorithms provide alternative methods when a symbolic manipulating algorithm either does not exist or is too slow. Symbolic constants and rational numbers can be evaluated to an accuracy of any number of digits.

•Graphical representations

univariable functions and double variable functions can be represented graphically. There are many options for customizing the way plots are displayed.
Internally, Maple consists of three components: *Kernel*, *Library* and *Interface*.
•The kernel is the mathematical engine which performs the majority of the basic computations carried out by the system. It consists of a set of highly optimized routines written in programming language C and then complied.
•Most of the Maple-V's commands reside in the Maple library and are written in its own programming language. A Program written in Maple is interpreted as it is entered. This allows the user to communicate with the program interactively within a session. A user can enhance the existing standard library by adding his/her own procedures to the system.

•The interface defines to a large extent, how a user interacts with commands and procedures. Depending upon the available hardware, Maple-V can offer a rather simple or sophisticated interface.
Numbers, constants, strings and names are the simplest objects in Maple-V. Maple has a wealth of built-in commands stored in the Maple library. Each command is called as:

commandname(par1,par2,par3,.....,parn)

Command names have been chosen to best represent the functionality of a command and at the same time require the least amount of typing. The majority of Maple-V commands are written entirely with lowercase letters. Each Maple-V command takes a parameter sequence as input. This sequence may contain several numbers, expressions, sets, lists or no parameters at all. Some examples of the commands are:

```
isprime(111);
```
false
```
diff(3*x+x^2-1,x);
```
$3+2x$
```
int(x^3*y^2,x);
```
$x^4y^2/4$

When Maple starts up, it does not have any commands entirely loaded into memory. However, many standard commands that have pointers to their location is loaded, so that when they are invoked for the first time, Maple knows where to go to load them. There are other functions that reside in the library and must be explicitly loaded with the command such as

```
readlib(factors);
```

Once a command has been loaded into memory, it does not need to be reloaded during the current Maple session.

12-Mathematical content and application of Maple-V

The main components of Mpale-V have been shown in Fig-3.

Calculus Linear Algebra	Solving Equations	Programming and system commands
Polynomials and Common transformations	Geometry Combinatorics	Expression Manipulation
Number Theory	Standard functions and constants	Plotting

Fig-3-contents of the Maple-V package

For a complete set of instructions see Redfern[19]. Large and more diverse classes, reduction in staff/learner ratios, different learner work habits and expectations from higher educational institutions makes traditional teaching more problematic and difficult. About 20 of the Australian and New Zealand universities are in a Macintosh Consortium. Maple for the Mac has been available since 1987; overall about a quarter of the Australian universities use Maple-V in the first and second years for teaching mathematics with many other users in research and advanced teaching. Maple has also been integrated into the mathematics course of some universities in United Kingdom. The authors believe that algebraic processing languages can also be integrated into the mathematics curricula of polytechnics in Singapore. For example, Maple-V can be used to assist teaching topics such as: Introduction to numbers, Sets, Functions, especially polynomial functions, Quadratic equations, Major operations on polynomials, Matrices, Trigonometry, Complex numbers, Solution of linear and non-linear equations, Differentiation, Integration, Optimisation, Solution of differential equations, Fourier series and Fourier transforms, Laplace transforms
Maple can support learners in two ways:
1-The software can be used to explore ideas. For example, learners can try to see the effect of different transformations in integrating a function.
2-The effect of variation of one or two parameters can be easily investigated without tedious algebra hampering the direct testing. For example: in quadratic equations,

the learners can investigate the effect of reducing the magnitude of the coefficient a over the roots of the equation.
3-The use of the software encourages learners to discover partial answers for questions which require investigations.

13-Conclusion

Because of the complex nature of the problem solving process, no single method can be treated as the best one for teaching all types of learners. Algebraic processing languages can help learners in problem solving and can be easily integrated into the existing curricula at polytechnics in Singapore.

14-References

1-Henderson,K.B.,Pingry,R.E.,(1953),"Problem solving in mathematics" In the Learning of Maths: Its theory and practice. Twenty-first Yearbook of the National Council of Teachers of Mathematics, Washington,D.C.

2-Descarts,R.(1967), "The philosophical Works", Vol-1, Translated by E.S. Haldane and G.R.T. Ross, The university Press, Cambridge.

3-Bergson,H.(1954), "Creative evolution", Translated by A. Mitchell, Macmillan &Co, Ltd., London.

4-Beth,E.W.,Piaget,J.,(1966), "Mathematical Epistemology and Psychology", Reidel, Dordrecht.

5-Spinoza,B.(1967), "Ethics and Treatise on the Correction of the Understanding (Translated by A.Boyle). Everyman's library Dent: London.

6-Poincare,H.,(1913), The foundation of science", (trans. Halsted) Science Press, New York.

7-Hahn, H.,(1956)," The crisis in intuition" in J.R. Newman(Ed), The world of mathematics, Simon and Schuster, New York.

8-Polya,G.,(1962), "Mathematical Discovery: On understanding learning and teaching problem solving ", Vol(1), New York: Wiley.

9-Polya,G.,(1965), "Mathematical Discovery: On understanding learning and teaching problem solving ", Vol(2), New York: Wiley.

10-Newell,A.,Simon, H.,(1972)," Human problem solving", Englewood Cliffs, N.J.:, Prentice Hall.

11-Lester,F.K.,(1978), 'Mathematical problem solving in the Elementary school: Some educational and psychological considerations", In mathematical problem solving: Papers from a research workshop, edited by L.L.Hatfield and D.A. Bradbard. Columbus, Ohio: ERIC/SMEAC.

12-Hitfield,L.L.(1978), "Heuristical Emphases in the Instruction or Mathematical Problem Solving: Rationales and Research." In mathematical problem solving: Papers for a research workshop, edited by L.L.Hatfield and D.A. Bradbard. Columbus, Ohio: ERIC/SMEAC.

13-Rothkoft,E,,Z.,(1966),"Learning from written instructional materials: An exploration of the control of Inspection Behaviour by Test-Like Events." American Educational Research Journal 3.

14-Ausubel,D.P.,(1968),"Educational psychology: A cognitive view", New York: Holt, Rinehart and Winston.

15-Bloom,B.S., (1980),"The new direction in Educational Research", Alterable Variables."Phi Delta Kappan 61, pp(382-385).

16-Phillips,E.,Ray, Kane,R.B.,(1973),"Validating learning hierarchies for sequencing mathematical tasks in elementary school mathematics", Journal for research in Mathematics Education 4.

17-Polya,G., (1957),"How to solve it", 3rd ed. Garden city, N.J.: Doubleday.

18-Rubinstein,M.F.,(1975),"Pattern of problem solving", Englewood Cliffs, N.J.: Prentice-Hall.

19-Redfern, D., (1993),"The Maple, Handbook", Springer-Verlag, New York.

INTERACTIVE MULTIMEDIA IN TEACHING MATHEMATICS

P.M.Nobar
Mathematics and Science Centre,
Ngee Ann Polytechnic,
Singapore.

A.J.Crilly
Education Development Centre
Ngee Ann Polytechnic
Singapore.

K.Iynkaran
Mechanical Engineering Department
Ngee Ann Polytechnic,
Singapore.

ABSTRACT

The role of mathematics and its level of mastery have always been important issues in designing mathematics curricula. So far no unique set of answers have been found for any of these questions. The common way to know the world is to experience it directly. However, as the world becomes more complex, it becomes increasingly difficult to provide learning with direct experience. Therefore, some compromise must be found to teach learners to understand their environment. In this paper the authors review one of the commonly accepted criteria for the role of mathematics in engineering science and extend these criteria to cover the field of numerical computation. The process of learning and teaching in education can be viewed as a communication model. The engineering and psychological models of communication have been discussed. It has been shown that, there are a number of problems in the communication process which hinder learning. One of the proposals for eliminating these problems has been to make the lessons more interesting by employing a number of suitable media. This paper shows that the interactive multimedia programs can be a suitable media for teaching and learning. In pursuit of this claim, the authors have developed an interactive multimedia program which can be used in teaching numerical analysis. However, the cost of the development and quality of software are two factors which have slowed down their acceptance in education at large. This paper discusses the formation of a group consisting of academic staff and students which can help to develop high quality software with minimum cost.

Key words

Engineering mathematics, what mathematics an engineer should know, communication, engineering model, psychological model, barriers to successful communication, excessive verbalization, confusion over referents, field of experience, day dreaming, perceptual difficulties, physical discomfort, deficiency in language, learning process, behavioural objectives, Skinner's model, Bruner's model, CBL, Interactive multimedia, Basic CBL hardware, Modern components, Numerical analysis, Algorithms, enactive, iconic, symbolic, autonomy, variety.

1-Introduction

The content and methods of teaching mathematics to undergraduate and diploma students at universities and polytechnics have been debated for several years and a number of attempts have been made to rationalize and define the precise needs of engineers. It is now widely accepted that an engineer cannot be an expert in mathematics and the content of engineering mathematics must reflect the present and future demands. It is thought that there is no need for mathematical rigour but special attention must be given to the following points:

1-An engineer must be able solve basic engineering related mathematical models which requires straightforward application of mathematics and statistics.

2-An engineer must be able to communicate both in writing and orally the results of analytical and statistical information as well as being able to read and understand literature that contains application of basic mathematics and statistics.

3-An engineer must be able to understand the mathematical modelling procedures and be able to use computer packages and interpret the results sensibly.

4-An engineer must be able to continue professional development in future by means of access to relevant technical literature.

The development of powerful, inexpensive desktop computers presents both an opportunity and a challenge to engineering educators. Computers can now be brought into classrooms to expand the educators' abilities to illustrate concepts and present more realistic exercises. The challenge for academic staff is to exploit computers to reduce the burden of teaching and so concentrate more on research and development, without reducing the quality of teaching and increasing the cost of delivery. The main objectives in incorporating computers in numerical analysis or in computational mathematics courses have been:

- Some of the algorithms have a relatively large number of steps. Using conventional media to solve realistic problems requires a large amount of time and occasionally the amount numerical calculations prohibits such practice in the classrooms.
- For some algorithms when written explanations are enriched with verbal comments, their understanding become much easier.
- Tedious numerical calculation can be delegated to computers, so that students can fully concentrate on the steps and structure of the algorithm.
- The programs can be accessed during tutorial hours by all students, making them more active and in addition making supervision more uniform and productive.
- Students can access the programs outside formal tutorial hours in order to do their tutorial work.

The achievement of the above objectives have always been problematic and no practical solutions have been discovered so far to deal with them. In this paper the basic problems with teaching and learning will be highlighted and possible use of interactive multimedia programs will be investigated.

2-Communication and learning

The common way to know the world is to experience it directly. However, as the world becomes more complex, it becomes increasingly difficult to provide learning with direct experience. Therefore, some compromise must be found to teach learners to understand their environment. The transfer of information from educators to learners is referred to as communication flow. Obviously effective education cannot take place unless communication takes place. There are two types of models for communication:

1-The Engineering model

The first engineering model for communication was developed by Shannon&Weaver[1]. The model is useful in a sense that it allows the identification and analysis of the critical stages of educational communication. In the model which is shown in Fig-1. Shannon introduces the information as binary bits and derives formulae for determining the number of bits in a message and lays out rules for handling the bits of information under varying conditions of transmutation. The messages travel from the sender toward the receiver. The route along which they travel may be thought of as communication

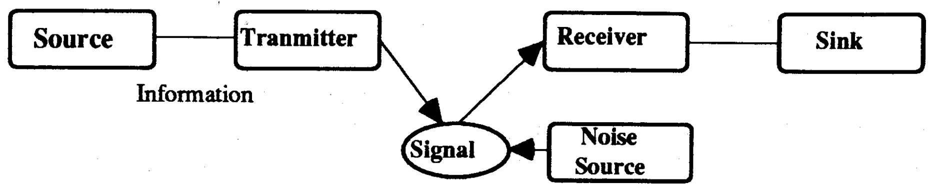


Fig-1-Shannon&Weaver model

channel. A message may consist of a statement made by a learner or an educator, a drawing on a white board or printed material. The receiver of this message may react in some way by listening, looking etc. Ideally a communication channel carries both messages and counter-messages. Any other irrelevant or distracting messages being received are referred to as noise in the communication system.The objective of effective communication is to maintain the maximum signal to noise ratio in the system.
The development of educational communication theories has led to the design and production of machines which learn or teach. The science of cybernetics is mainly concerned with the application of communication theory to explain and simulate human thought. So far, complex devices have been built which adapt the course of training on the basis of an individual learner's error pattern.

2-The Psychological model

The psychological model of communication is also concerned with the effect of a message travelling away from its source, but the main interest is what happens to the recipient ? In an ideal learning environment, messages travel along a channel clear of any interference. However, in the real world, other media also influence the messages. In addition psychological barriers to communication are created between educators and learners by conditions inside and outside classrooms as shown in Fig-2. The barriers can be divided into the following groups:

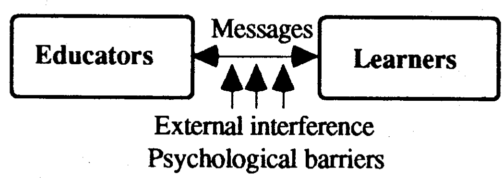


Fig-2-A simple model of communication with noise effects

- *Excessive verbalization*
  Learners are capable of switching off when their educator is uninteresting. Experienced educators report that the continued use of any one kind of teaching method results in decreasing interest. Namely, the effectiveness of words declines as learners sit through lecture after lecture. Use of interactive multimedia programs can reduce drastically the adverse influence of this phenomenon.
- *Confusion over referents*
  It is important to note that in communication one cannot transmit meaning. What is actually transmitted are symbols of meaning, such as words or pictures. In receiving messages it is natural to turn automatically to related experiences in order to understand something new. When reference to previously learned material can help one to understand the new material, a positive transfer of insight is said to have taken place. Often application of previous experience results in referent confusion. That is, the experience a learner draws on to digest the new material is not relevant. Schramm[2] adapts Shannon&Weaver's model to explain this concept more clearly. As shown in Fig-3, the material presented should be sufficiently within the learner's field of experience so that the learner can grasp what needs to be learned. However the material must be enough outside the field of experience to challenge and extend that field. The range of proximal development before confusion sets in depends upon many factors among which one is the ability of the learner. Able learners

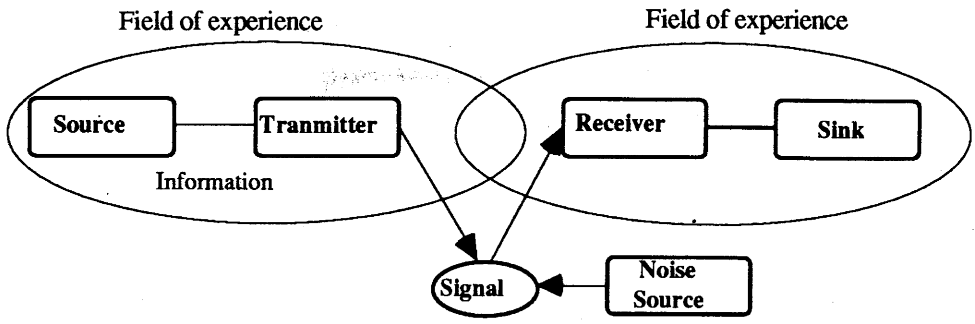


Fig-3-Schramm model

usually accept larger gaps between their field of experience and their educators. This implies that every learner is unique and as a result the mass education systems, i.e., lectures are bound to be ineffective in higher levels of learning. For further details see Bloom[3]. The interactive multimedia programs can improve the situation by providing more background material enabling every learner to take part in learning process in his/her own pace.

- *Daydreaming*
  Many learners are adept in daydreaming. This kind of distraction can be eliminated by making the lectures more interesting and by seeking the student's involvement. One again interactive multimedia programs can be more effective in keeping learners' attention longer by providing animation, graphic and audio.
- *Perceptual difficulties*
  Science has acquired a great deal of information about the physical aspects of perception. However the psychological processes by which the brain receives and deals with messages through eyes and ears are less clear. Some learners have poor hearing and poor sight. Some are verbal rather than visual learners. These differences can affect the learners' reactions. Experience shows that, interactive multimedia programs can reduce the effect of this phenomena.
- *Physical discomfort*
  The fact that comfortable physical surroundings are desirable in education and industry has been generally accepted. In recent years classrooms and factories have been remodelled for comfort. For further information see Brum[4].
- *Deficiency in language*
  If the receiver of the message is not sufficiently proficient in a particular language, the communication is bound to breaks down. Whereas this is relatively easy to diagnose in the case of formal spoken languages, it is not always easy to identify when symbolic systems such as mathematics are used. Undoubtedly, the interactive multimedia programs can reduce the effect of language deficiencies by offering more help through text, audio or Video.

3-The learning process

It has already been shown that, the education is a two-way communication process in which:

- the educator transmits a variety of messages;
- the learner by performing certain tasks, communicates back to the educator that learning is taking place.
- The received information is interpreted by the educator who decides whether any corrective message or any other action should be taken. These actions are then translated into feedback information to the learner. Fig-4 shows how the educational communication process would cycle one or more times until learning objectives are achieved. Learning is one of the most important areas in the psychology of education. It is a concept which is extremely difficult to define. Science needs an observable or measurable subject, within the science of psychology, the subject is behaviour, psychology has tended to become a behavioural science. Hence,

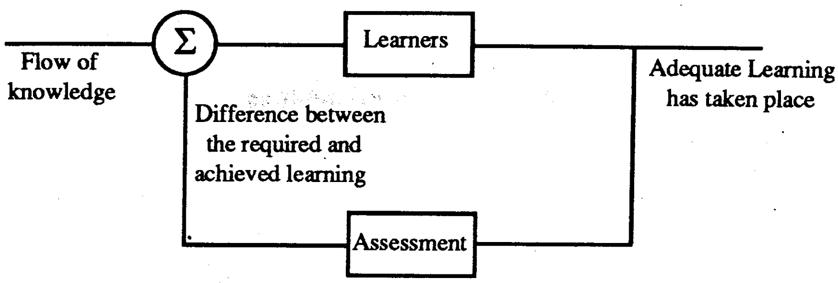


Fig-4-A feedback control model for teaching and learning

every study in psychology must be expressed through behaviour. However, this does not mean that behaviour is learning. Psychologists study behaviour in order to make inferences concerning the process believed to be causing behavioural changes. This process is referred to as learning. Most learning theories agree that the learning process cannot be studied directly; instead its nature is to be inferred from changes in behaviour. Skinner(1904-1990)[5,6] and his followers are a group of psychologists who believe that behavioural changes are learning and that no further process exists. Following Skinner's theory of learning Mager[7] concluded that the objectives of a course can then be defined in terms of the behaviour patterns. Namely, learners should be able to perform certain tasks. Hence it is possible to design an appropriate test to evaluate the method of teaching. According to Mager[7] the essential ingredients of a behavioural objective are:

- A statement of what a learner should be able to do at the end of a learning session.
- The conditions under which the learner should be able to exhibit the terminal behaviour.
- The standard to which a learner should be able to perform.

In terms of teaching numerical analysis algorithms, for example, a learner should be able: To solve a system of linear equations using the Gaussian Elimination algorithm; to use the steps of the algorithm in addition to computers and finally to determine the solutions correct to **n** decimal places.

Mager[7] popularized the precise statement of objectives. Later his approach became more widely applied designing courses. Many psychologists feel that Skinner's model of human learning is incorrect or at least incomplete. In particular, it is difficult or impossible to explain how all learning takes place in terms of a chain of stimulus and response bonds. An alternative theory of learning was proposed by Bruner[8] in 1966. This is a sequence based on Bruner's interpretation of Piaget's[9,10] developmental theory. Bruner[8] is a proponent of the discovery approach in mathematical education. The discovery approach is characterized by three stages:

- enactive (manipulation of material directly);
- iconic (manipulation of the mental images of the objects);
- symbolic (strict manipulation of the symbols).

For further details about theories of learning see Hergenhahn[11]. An application of this theory in the development of an interactive multimedia program in teaching vibration to mechanical engineering student is given by Nobar[12].

4-The role of media in education

In education, media refers to items such as equipment, photographs and charts which can effectively be used for teaching learners. In communication jargon they are the carriers of messages from source to destination. These carriers of information interact with the learners through their senses. Computer based learning programs like other media used in teaching and learning have various unique qualities and are one group of the tools that can be used to widen the channel of communication. Typically a CBL package will support other printed materials,lectures, seminars or tutorials. In education there are a number of reasons for using different media in delivering a certain course:

- to choose the best media for the course;

- to create a variety to limit the effects of distractive phenomena;
- to meet the financial constraints.

Each medium, whether traditional, such as white board or more recent, such as video or CBL has its strengths and weaknesses. The task of the instructional designer is to use the different media and methods so that the strengths are complemented and the weaknesses are canceled out.

5-Interactive multimedia

The role of computer based learning software may be considered by analyzing its components. Any activity in developing and using CBL software is based upon three factors:

1-Hardware

Historically, the communication between the computer and the learner has been through a keyboard and a visual display unit. The limitation of this type of communication are:

- the presentation is restricted to the resolution of the screen;
- the presentation is strictly two dimensional;
- the animation is inferior to broadcast television;
- the input is restricted to standard set of symbols;
- the interaction is slow as most of the users are not typists.

In the last decade, considerable effort has been made to develop other complementary components to increase the bandwidth of the communication between learners and machines. These components include:

- audio facilities;
- voice recognition;
- speech synthesis;
- random access audio disks, tapes or compact disks;
- random access slide projectors;
- random access videodisks, tapes or compact disks.

A computer system equipped with all or some of the components is referred to as an interactive multimedia system. Paradoxically, although many alternatives have been tried, most CBL packages are still confined to the traditional interface. The reasons are:

- additional cost in hardware;
- more time is required to develop CBL software which uses the modern components;
- limited ability of software developers to use the new components;
- some developers who have grown up using keyboards find it the most natural way of communicating with a computer.

2-Software

The role of the CBL software is to mediate in the flow of information. This involves some or all of the following tasks:

- help to use the hardware effectively;
- present information in text form;
- control other presentations of information such as audio, video and slides;
- recognize the learner's response to questions;
- keep records of learner's progress;
- select the appropriate response to meet learner's demand;
- perform the required calculations;
- retrieve information requested by learners;
- report the learners performance.

A study of some of the released CBL packages from the commercial sectors and academic sources show that a number of CBL software: do not make use of computer graphics effectively; ignore the use of text completely; seem to use facilities just because they are available; a number of CBL software seem to have too many colours on the screen; a number of CBL software have the control keys all over the screen making it difficult to control; a number of CBL software have a large amount of repetitive material; a number of CBL software have inconsistent control buttons.

3-Management

While educators and CBL software developers are mainly concerned with the problems

of teaching and training, departmental heads and managers are conscious of the problems of resources and costs. It is well known that both design costs and quality of software are critically dependent on the availability of suitable tools for courseware development. It is possible to code courseware in one of the general purpose programming languages such as ForTran, Pascal or C. A large number of CBL programs have been coded using these languages. For a list of engineering CBL programs see Nobar[13,14,15,16,17]. However these languages were not designed to satisfy the needs of CBL authors. Some of the limitations of high level languages have been removed in modern authoring languages such as Authorware Professional, SchoolBox etc. However, these authoring development tools are large in size and less flexible. If academic staff with the help of students are involved in the development of CBL software, then use of authoring language will be restricted to the modern object oriented software products as they do not require trained computer programmers.
Creating computer based learning software is a necessarily long process and it is impossible to compare its development time with the time required for developing classroom work. When comparing costs, the effectiveness of each medium should also be considered, that is:

$$\frac{\text{effectiveness of lecture}}{\text{cost of lecture}}$$

versus

$$\frac{\text{effectiveness of CBL}}{\text{cost of CBL}}$$

Superficially, this looks an easy task, but a closer investigation shows that only the cost of CBL development can be estimated accurately and to some extend the cost of lectures. The other variables require a large number of data points which are difficult to obtain, because educational research is a difficult and controversial subject and very little conclusive research has so far been carried out.

6-Local development group

In order to minimize the cost of development, one option is to form an interest group within the institution and produce high quality software which is designed by the academic staff and implemented by a number of project students. This is a policy which has been adopted by a group of academic staff at Ngee Ann polytechnic. The group adhere to the following principles:

•*Project autonomy*

If computer assisted learning is to be adopted campus wide then it must be adapted to the personal preferences, needs and circumstances into which it is introduced. The group has accepted this need for project autonomy. This is a state of affairs which suits universities as well, given their concern for academic freedom. Project autonomy and personal involvement are likely to enhance the prospect for institutionalization.

•*Pursuit of variety*

Different departments are involved and a variety of uses of computers are being tried. The virtue of pursuing variety is: it increases the coverage of areas of potential relevance and opportunity. Projects become designed products which meet the unique requirements of the curriculum. The work of the group differs from a research programme in the sense that a research programme works by controlling variation; however, the group explores variety, but supports and exploits the natural proliferation of variants rather than create variation for its own sake.

•*Eclecticism*

The pragmatism which is apparent in the previous principles implies an eclectic approach to educational values. The question of what educational theory is best suited to CBL has been left open. Against the background of this eclecticism, one issue stands out, that is, the debate between computer assisted education CAI and CAL. By definition

CAI appears to emphasize the role of the computer as a tool for the educator; on the other hand, CAL emphasizes it as a tool for the learner. The group prefers CAL over CAI for immediate use of the product. The group members have surveyed the field of possible uses for the computer in education and chosen to implement CAL in the first instance. That is, the computer is viewed as another media in the educational context which cannot provide education by itself.

7-Numerical analysis algorithms

Numerical analysis is referred to as the "theory of constructive methods" in mathematical analysis. A constructive method refers to a procedure that permits one to obtain the solution of a mathematical problem with an arbitrary precision in a finite number of steps that can be performed rationally. A constructive method usually consists of a set of directions for the performance of certain arithmetical or logical operations in predetermined order. The set of directions to perform mathematical operations designed to lead to a solution of a given problem is known as an algorithm. The word algorithm was originally used to denote procedures that terminate after a finite number of steps. At the present time, numerical analysis or numerical computations are assuming a role of increasing importance in applied science including engineering science. It is the actual solution rather than the theoretical production of a solution is of paramount interest. More and more engineering curricula are including courses in numerical analysis. The increasing availability of computers has caused a basic change in the presentation of courses in numerical analysis. The programming of numerical algorithms can be immediately emphasized as part of such courses with three fold advantages:

- The learners know that they have mastered a given algorithm when they can program it successfully.
- The learning of programming techniques is better motivated when the learners can apply them immediately to the coding of a useful numerical algorithm .
- The learners can solve more realistic problems by employing numerical methods in analyzing the mathematical models.

8-Motivation for developing GAUELM package

In almost in every scientific discipline there are problems that give rise to a system of linear algebraic equations. For example, an electrical network with passive and active elements can be represented as a system of linear equations in order to determine the current at each branch or voltages at each node. In polynomial curve fitting, the coefficients of the least square polynomial are determined by solving a system of linear equations. In structural or machine part designs the displacements are usually determined from the solution of a system of linear equations.

There are two different methods of solving these equations: direct and indirect methods. One of the main algorithms which is included in the curriculum of mathematics for polytechnics in Singapore and other countries is the Gaussian Elimination algorithm. After teaching the algorithm for many years the authors have noticed difficulties in communicating the details of the algorithm to students. This is due to the fact that to learn the algorithm one must solve a reasonably large size problem, that is, a system of equations with at least 4 unknowns to show all the details excluding special cases. Due to the size of numerical calculation, it is practically impossible to show all the details using conventional media such as overhead projector transparency or white board. In addition the methods of written solution, on sheets of paper or on transparency sheets have been proven ineffective. The interactive multimedia program GAUELM has been developed to overcome these problems. In this program using audio, computer graphics and animation all the details of the algorithm have been explained. The program has the following sections:

- objectives;
- introduction;
- classification of the algorithms for solving a general system of linear equations;
- description of the basic algorithm on a system of equations with four unknowns;

•solution of preset examples for 2,3,4,5 and 6 unknowns;
•solution of user defined exercises 2,3,4,5 and 6 unknowns.
The program has been designed so that the other direct algorithms can be easily added to the existing program without any alterations. Some typical frames of the package are shown in Appendix-I

9-Hardware and software

The Education Development Centre at Ngee Ann polytechnic is equipped with basic PC interactive multimedia systems. A typical system consists of:

Hardware

- 80486 Intel processor (33MHz);
- CD-ROM derive and controller;
- An audio board with two speakers;
- 0.5GBytes of hard disk;
- SVGA display monitor.

Software

- Microsoft Windows graphical environment;
- Multimedia extension.

The program has been developed using Authorware Professional[18] software, this choice being based on the campus wide availability of the package at the Ngee Ann Polytechnic.

10-Statistical results

At present the package has been examined by a small pilot group of students. Therefore, no statistical results can be supplied on the performance of the software. The software is available on diskettes which can be obtained from EDC. The package will be transferred to the general library for use on the campus network and the group intends to publish the results in one of the forthcoming conferences in Singapore.

11-Conclusion

The general communication model has been reviewed and the extension of Schramm has been illustrated. The zone of proximal development has been shown and it has been pointed out that in a large class every learner will have a different zone. The elements of distraction in a classroom have been discussed and it has been shown that by employing media such as interactive multimedia programs, it is possible to unify the zones of proximal development of different students. The interactive virtues of GAUELM have been discussed as well as its advantages over conventional teaching methods. This interactive multimedia approach represents an innovation to the teaching of mathematics and numerical analysis and the results will be available and publicized shortly.

Acknowledgment

The authors would like to thank two of their students in the department of Mechanical Engineering, Yeo Heng Chai and Tan Cheng Guan, and Tommy Ong, student from the Computer Science Department, who helped us extensively with GAUELM program. Their efforts made the development of this program possible.

Appendix-I

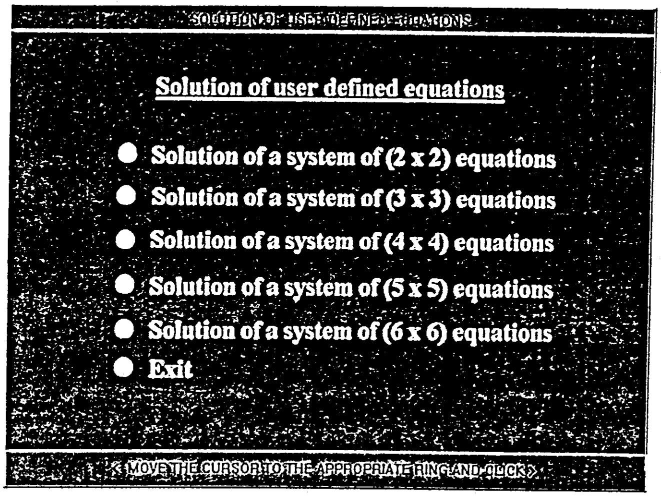


Fig-1-Selection of the order of equations

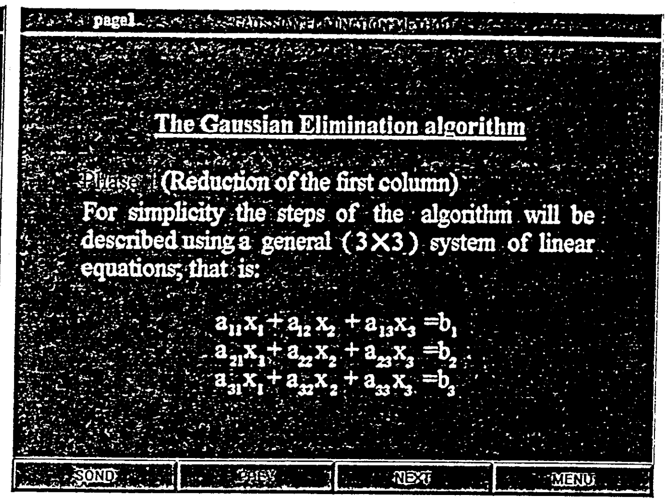


Fig-2-A typical frame from the theory

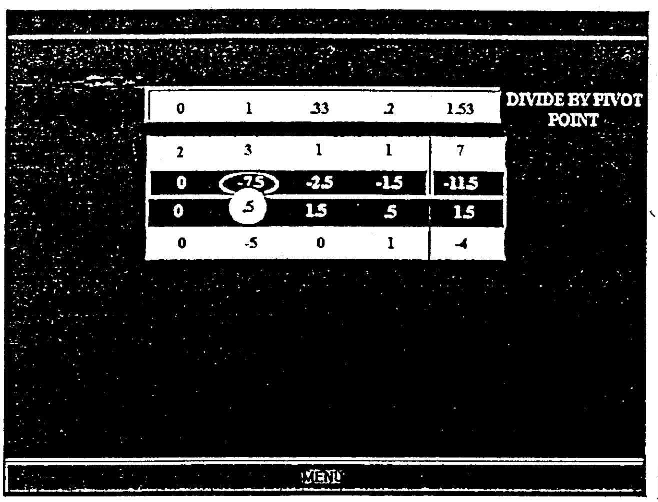


Fig-3-A frame from the iteration process

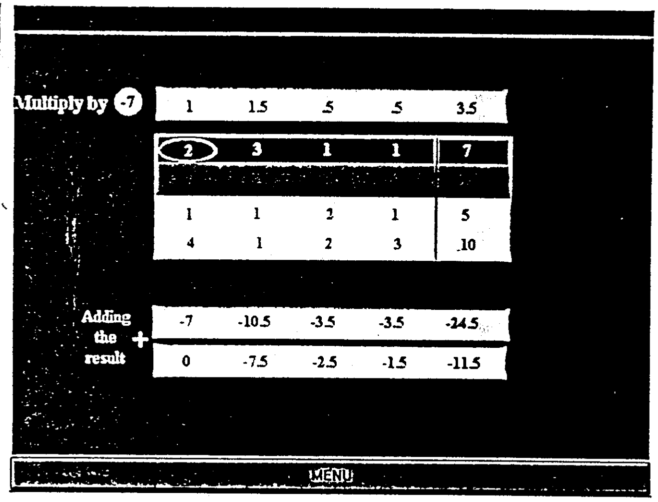


Fig-4-A frame from the iteration process

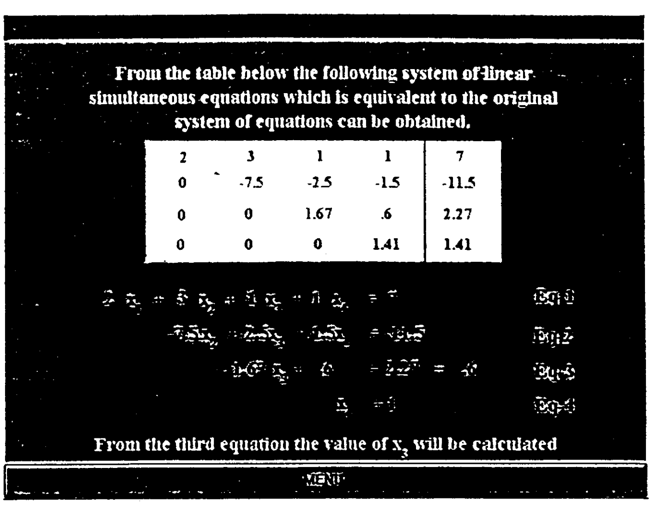


Fig-5-A frame from the back sustitution process

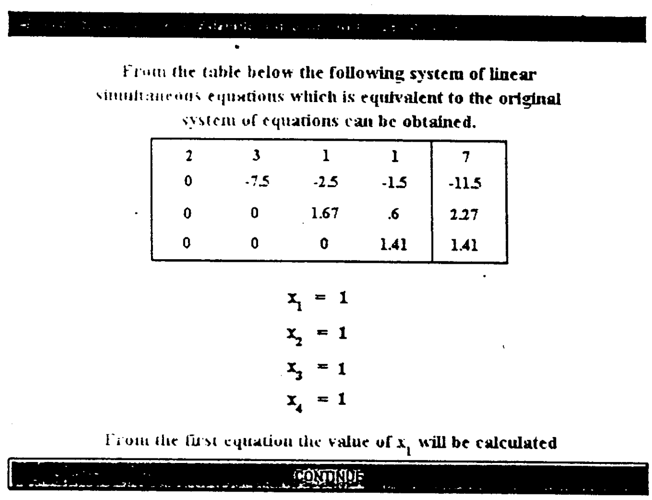


Fig-6-A frame with the final answers

References

[1]-**Shannon,C.E.,Weaver,W.,(1949),**"The mathematical theory of communication", Champaign, IL,University of Illinois Press.

[2]-**Schramm,W.,(1954),**"Procedures and Effects of Mass communication", in Mass Media and Education, ed., Nelson B. Henry(Chicago; University of Chicago Press).

[3]-**Bloom, B.S., Engelhart, M.D., Hill, W.H., Furst,E.J.,Krathwohl,D.R.,(1956),** Taxonomy of educational Objectives, Handbook I:

[4]-**Brum,F.,(1956),**"Atlantic City Educations Seminar", Carnegie Corporation, reported in Honolulu Advertiser.

[5]-**Skinner,B.F.,(1953),**"Science and human behaviour", New York Macmillan.

[6]-**Skinner,B.F.,(1961),**"Cumulative record", Methuen, New York.

[7]-**Mager,R.F.,(1962),**"Preparing Instructional objectives", Fearon Belmont,CA.

[8]-**Bruner,J.S.,(1966),**"Towards a Theory of Instruction", Norton New York.

[9]-**Piaget,J,(1957),**"Logic and Psychology", Basic Books, New York.

[10]-**Piaget,J,(1965),**"The child's conception of numbers",W>W.Norton and Company Inc.,New York.

[11]-**Hergenhahn,B.R.,(1988),**"An Introduction to Theories of Learning", Prentice Hall, International edition.

[12]-**Nobar,P.M.,Crilly,A.,J.,Iynkaran,K.(1995),**"The increasing influence of computers in engineering education" (A case study for teaching via computer based technologies)", a paper to be published in the international journal of Engineering Education.

[13]-**Nobar,P.,Williams,J.J.(1992),**"Promoting CBL software and role of tutorial programs in teaching fluid mechanics", Proceeding of ASME European Joint Conference on Engineering Systems design and analysis, New York, ASME.

[14]-**Nobar,P.,(1990),**"Computer Based Learning in Teaching Vibration", Workshop on Computer aided Open Access Education, Napier Polytechnic of Edinburgh (Edinburgh, Napier Polytechnic Press..

[15]-**Nobar,P.,Song,B.,(1990),**"Simulating the behaviour of three dimensional reticular structures with both curved and straight elements", Proceeding of the ICCAL90 conference, Hagen, Germany (Berlin, Springer-Verlag)

[16]-**Nobar,P.,Andrews,K.,(1992),** "Computer Based Learning Software For teaching structural design", Proceeding of the conference in computer training in property and construction, Salford, UK.

[17]-**Nobar,P., Williams,J.J. (1995),** "The educational analysis of computer simulation in teaching and research",(a Case study using SIMULINK). Proceeding of CBLIS95 conference, University of Opava, Czech Republic.

[18]-**Nobar,P., Pavel,F, (1995),**" The Role of Simulation in Engineering Practice", (A case study using CODAS-II), Proceeding of CBLIS95 conference, University of Opava, Czech Republic.

[19]-Authorware Professional software manual

Analyzing what had been learnt
Technology keeps an eye on students' response.

Noor Isham Sanif
Princess Elizabeth Primary School

Abstract

Today, computer can be used to facilitate our analysis of students' "cognitive process" in their attempt to solve problem sums. It can track students' solution paths in problem solving. This study explores the possibility of using the computer programme, *Macro*, to play back 41 students' answer scripts. It aims at analysing each of the students' response and causes of error in dealing with non-routine mathematical problems.

Descriptors : Macro playback, solution path, cognitive process

Introduction

Educationist and mathematics teachers have been using results obtained from product-oriented assessments, like multiple choice test, to evaluate students' understanding of mathematical concepts. In such means of assessment, these evaluators get very little information on a student's solution path and causes of errors in problem solving.

For example, in open ended and problem sum questions, evaluators tend to treat all wrong answers equally. They fail to note that although all wrong answers look alike, each student's understanding and approach to the question differs. A student, for example, may get his answer wrong due to one of the following reasons:

(a) insufficient time spent in reading the question;
(b) misconception of the question;
(c) does not check and review his answers or
(d) does not understand the question at all.

Monitoring the cause of a student's error in solving problem sums would require a teacher to sit beside him as he does his worksheets. This is however not possible in a class of say, 40 students, as it would require the deployment of 40 teachers.

Objective of study.

This study aims at using the computer programme, *Macro,* to reflect actual students' behaviors in problem solving. It reveals students' solution paths which can be used to facilitate our analysis of students' "cognitive process" in exploring mathematical questions. Students' responses and causes of errors can, hence, be interpreted.

What is Macro ?

Macro, a program within Microsoft, is capable of recording a series of actions which can be "playback" in sequence by a computer user. The series of actions recorded may range from the drag of a mouse, pressing of letters on the keyboard or performing complicated calculation. The program is very useful as lesser time is needed by a user to do a repetitive task. For example, once the Macro records the sequence of a calculation, it can execute the sequence repeatedly for any question. The user does not need to physically type in the whole sequence again. No matter how laborious a calculation may be, all the user has to do is to press only **one** button on the keyboard.

In this study, *Macro* will be used to playback 41 students' solution paths in solving 5 problem sums.

Design

Profile of Sample

This study involved 41 Primary 5 students of the Bukit Batok Mendaki Tuition Class. These students are of mixed ability. Each student was assigned with a serial number.

The study was conducted in October 94, immediately before the school holidays. At that time, the students had sat for their final year examinations and, therefore, had sufficient knowledge of the whole year Mathematics syllabus.

The students were first given a month training of how to use the Microsoft Word. They were not only taught how to type and edit their answers using the copy-paste mode of the software, but also how to draw diagrams and models using the drawing tools available. This encouraged them to solve the problem sums through visualization. The students were, however, not informed that all their responses would be automatically recorded.

Procedures for Data Collection

Using the 386 IBM compatible computers, each student was asked to solve 5 problem sums (see Appendix A) within 45 minutes. Since no rough paper was provided, students had to type out all their workings on the test screen. They were also given the permission to amend their answers as and when they like.

The test screen looks like Figure 1 on the next page. Before answering any question, the students would have to key in their serial numbers and the question number which they wished to attempt. They were then told to show whether they were ready to solve the selected sum by clicking on the 'Yes' or `No' button on the screen. Once the `Yes' button was clicked on, the computer would show the selected sum on the screen. It would also simultaneously record all the students' response. The macro would then automatically be saved under the filename D1Q101.Mac where D1 represented diagnostic test1, Q1 represented question 1 and the last two digits 01 represented the student's serial number. Each student, however, only needed to type in their serial number once, i.e. at the beginning of the test. When he was ready to answer the next question, he needed only to click on the `Next Question' button located on the Go To line. The students could also review his previous answers by clicking on the `Previous Question' button.

Figure 1

	Diagnostic Test 1	
Serial Numbers:	**01**	
Question No :	**1**	
Are you ready ?	**Yes**	**No**
Go to :	**Previous Question**	**Next Question**

1. A sum of $100 was put into three envelopes A, B and C. There was $40 in A. There was $5 more in A than in B. How much more money was in A than in C?

To see the students' answers, a teacher just had to open the files of each student. The teacher could scan through each student's answers just as if he is marking the traditional worksheets. After viewing a student's solution path, the teacher would have a better idea of his problems and causes of errors. The teacher may even show to each student his cause of errors if he wanted to. Alternatively, the teacher may only concentrate on all the wrong answers to provide diagnostic help tailored to the individual student's needs

Before discussing the test results, it is very important for me at this juncture, to let you know some predominant behavior (see Appendix C) which I observed on playing back the Macro. As each question was flashed on the test screen, most student would take at least 30 seconds to read and scan the questions. Some student, however, took more than 1 minute just to read a question before scanning for the important information in the question. **Scanning** is basically the process whereby a child underlines the key words of the question or rephrases the question to his understanding.

It was also observed that some students did not answer a question immediately. The playback showed that such students examined the question from different angles by constructing representation using dots, diagrams and algebra. This shows that the student was engaged in constructive **analysis**. If he was unable to answer a problem sum, he would try to find the answer through trial and error, or simply by making a guess. By doing so, the student is indirectly involved in **exploration**. When the students made their final deduction or statement, this means that they were already in the stage of **implementation.** At this juncture, the students would use mathematical procedure to derive their final answers. I noticed that very often, implementation may not be the last stage. This we would see, when students **reviewed** their answers. They would scan through their answers, check their workings and concepts, identify errors made and amend their answers accordingly. Sometimes, however, a review may take place before implementation.

Results and Discussion

Without using the Macro 'playback'

Question No	Completely Wrong response	Correct response	Partially correct response	Total
1	6	29	6	41
2	41	0	0	41
3	11	15	15	41
4	15	6	20	41
5	13	8	20	41

The above tabulated result was obtained without using the Macro "playback". It can be seen that all the students found question 2 the most difficult to do. None could not get the correct answer. Question 1 was found to be the easiest: 29 students got the correct answer, 6 students got the answer partially correct response.

Using the macro 'playback'

Using the Macro to playback all the students' response to question 1, it was discovered that (see Appendix A) all the students attempted the question after their first reading. 4 students went straight into implementation, of which 2 did not analyze the question further after their first reading. 4 did not attempt to explore the answer while 6 others faced difficulties during implementation. It was also found that the 6 students who got the wrong answers were those who did not review their answers.

For questions 4 and 5, it was discovered that 2 and 1 students respectively did not attempt the question at all after reading it. It was found that many students got their answers wrong due to the lack of time span in scanning the question for more information. As for question 2, 15 students did not analyze the question question 2 after their first reading. 30 students attempted the question without exploring further for the correct answer.

Out of the 20 students who responded partially correct in question 4, 15 did not attempt to analyze the question at all, 14 failed to review the answer, 5 of them spent less than 30 seconds reading and understanding the question, and 12 did not attempt to analyse the question at all. Out of those getting correct answers for question 3 all except 1 student analyze the question. It is clear from the above result that more incorrect responses were committed due to lack of good habit practised by some students when solving mathematics questions. *Macro* make it possible for teachers to identify these students and remediate accordingly.

From the playback it was also discovered that different students spent different portion of time on different part of the same question. 75% of the students spent more time at the implementation stage than any other stages. Those who spent more time on exploring and analyzing the questions than implementation of answers had been found to get the answers correct. This requires the teacher to spend more time showing the students ways to explore and analyze questions using manipulative ways so that they can easily visualize problem sum posed to them. It was also shown that not many pupils review their work after they had completed them. Reviewing will make a different to students' answers. This therefore requires teachers to train their students to review their answers more frequently.

Evaluation of the *Macro* Method

The teacher could first identify all the weak students whom he thinks need particular attention. He could then concentrate on these few students, analyse the area of weakness and apply remediation. If necessary he could review these answers with the presence of the student concerned. As the macro is being played back the teacher could find out from the student why he prefers to work out the question in that particular and advice on better means of solving the problem. Teacher could also play back the macro during lesson time to show to the class the common errors committed by students.

The task of carefully assessing students' answers is time consuming on the parts of the evaluators. Each students was given 45 minutes to answer their questions. If the evaluator were to playback each students' *Macro*, he would have to spend at least 1845 minutes (30.75 hours) on evaluating 41 answer scripts. The macro tool allows the user to playback each answers in one tenth of the actual time which means that the whole process may take at least 4 hours. It is therefore left to the initiative of the teacher .to select the students whose macro needed to be analyzed.

Evaluation requires great concentration. A good level of mathematical background and teaching experience may be essential. It was found, however, that not many primary school teachers are computer literate. Until all mathematics teachers are trained in computer application, can macro be used for this purpose.

The result has only shown the actual solution path taking place during problem solving. It did not show students' actual thinking and therefore is not an instrument to measure students' cognitive process. This method of analysis however is not without a room for improvement. Software engineers and computer programmers can develop the programme not only to record students' solution paths but also students voice. If students were given the chance to explain their solution evaluator not only able to see how they work out their answers in step-by-step mode, but also understand why students pick such answer paths. This is not impossible because more personal computers are equipped with sound cards, multimedia tools and microphones.

Conclusion

Understanding students' problem solving skill cannot be done by sheer observation of students' answer from a piece of paper answer. There is a need for a teacher to stay close to each student to observe them. This study has shown that *macro* could be used to analyse students' solution path in the area of problem solving. *Macro* can be used to analyse students answers days after they had provided the answers. If the software can be equipped with sound input facilities, students can explain as they execute their problem solving.

APPENDIX A

Analysis of wrong answers

Number of students facing the problem / Question No	Spent time reading the questions but give up	Problem with scanning of information go straight into implemetation	Did not analyse questions on the first attempt	Did not attempt to explore the answer at all	Give up after first exploration	Problem at implementation	Fail to review (even once) the answer provided	Mistakes due to careless calculation
1	0	4	2	4	0	6	6	0
2	0	22	15	30	2	41	16	0
3	0	5	2	5	2	11	9	0
4	2	12	10	2	1	15	7	0
5	1	9	12	5	1	13	11	1

Analysis of partially correct and correct response

Number of students facing the problem / Question No	Did not attempt to analyse the question at all	Failed to review the answer	Spent less than 30 seconds reading and understanding the question	Did not attempt to explore	Number of correct response	Number of partially correct response
1	3	10	0	2	29	6
2	0	0	0	0	0	0
3	14	1	1	1	15	15
4	12	14	5	3	6	20
5	10	13	10	2	8	20

APPENDIX B
THE PROBLEM SOLVING TASKS

1. A sum of $100 was put into three envelopes A, B and C. There was $40 in A. There was $5 more in A than in B. How much more money was in A than in C?

2. How can you arrange 20 counters in 4 straight line so that there are 6 counters in each line ?

3. Mrs Chen bought an equal number of green apples and red apples. She gave away
8 red apples to her friends. Then she had 3 times as many green apples as red apples left. How many apples did she buy in all ?

4. A, B and C shared $120. A received twice as much money as B. B received 3
times as much money as C. How much money did A received ?

5. Mingfa and his sister have $130 altogether. If their mother gives Mingfa another $10, Mingfa will have 3 times as much money as his sister. Can you find how much money Mingfa actually has ?

APPENDIX C

Behavior	Code
• Rereading of problem • Underlining key words • Rephrasing the question	⇒ Scanning
• Examining the problem from different angle • Construct representation • Draw diagram	⇒ Analysis
• Search for patterns • Trial and error	⇒ Exploration
• Use mathematical procedure to derive final answer • Make final deduction or statement	⇒ Implementation
• Going back to the previous answer • Recognize error • Amending the error • Return back to the previously done answer & rereading the response.	⇒ Review

REFERENCES

Lucas, J.F., Branca, N., Goldberg,D.,Kantowski, M.G.,Kelog, H.& Smith, J.P. (1980). A Process-sequence coding system for behavioural analysis of mathematical problem solving. In G.A. Golding and E. McClintock (Eds), Task Variables in Mathematical Problem Solving (pg 350-380). Columbus, Ohio: ERIC/SMEAC.

Putt, I.J. & Pounyney, G. (July,1989). Mental activity during mathematical problem solving. Mitchell College of Avdvanced Education, Bathurst, Australia.

Schoenfeld, A.H. (1983). Episodes and execuive decisions in mathematics problem solving. In R.Lesh & M. Landau (Eds), Aquisition of mathematical concepts and process(pg340-400). New York: Academic Press.

TEACHING MATHEMATICS AND MATHEMATICS EDUCATION AT A DISTANCE VIA TECHNOLOGY

Dr Ian Putt and Ms Gloria Stillman
School of Education
James Cook University of North Queensland
Townsville. Australia 4811
Ian.Putt@pop.jcu.edu.au
Gloria.Stillman@pop.jcu.edu.au

Abstract

Since 1990, the Remote Area Teacher Education Program (RATEP) has delivered university level primary teacher education via technology to Aboriginal and Torres Strait Islander students in their local communities in Queensland. The main platform for delivery has been interactive multimedia [IMM] computer courseware together with print materials. Communication technologies such as audioconferencing and facsimile have played an increasing role in facilitating lecturer-tutor, lecturer-student, and student-student interactions in both mathematics content and mathematics education subjects in the RATEP course. Features of IMM in these subjects are discussed along with the roles played by these two communication technologies. The possible future roles of other communication technologies such as electronic mail, audiographic conferencing, videoconferencing, and the World Wide Web are also examined.

Introduction

Since 1990, at James Cook University one course of teacher education has been delivered at a distance through the Remote Area Teacher Education Program (RATEP) to Aboriginal and Torres Strait Islander para professionals in communities throughout Queensland, Australia. RATEP is a social justice and equity initiative aimed at upgrading the qualifications of these para professionals to those of qualified teachers. Through a range of technologies which include interactive multimedia (IMM), audioconferencing, facsimile, video tapes and print materials this teacher education course is being delivered to students in their remote communities instead of requiring the students to attend on campus for their education.

RATEP formed part of the larger Queensland Open Learning Project (QOLP) which has established approximately 40 Open Learning Centres' (OLCs) spread throughout the state. In some locations these OLC's function as RATEP Centres while, in other locations, RATEP Centres are situated in a local school or community hall. Each RATEP Centre houses the hardware needed to run the IMM courseware, software, communication equipment for sending and receiving facsimile and teleconferences, video recorders and monitors, and facilities for students to work individually and in small groups. This fosters a close physical, social and academic environment which conforms to Aboriginal and Torres Strait Islander learning styles which generally favour group settings (Henderson, 1993).

RATEP stake holders are: James Cook University; Cairns College of Technical and Further Education (TAFE); the Queensland Department of Education; a number of Aboriginal and Torres Strait Islander committees; Community Councils at each of the RATEP sites; the Queensland Open Learning Centre Network; the Queensland Office of Higher Education; the Commonwealth Department of Employment, Education and Training; and the students.

RATEP has a number of unique characteristics. Firstly, it delivers teacher education at Certificate, Associate Diploma, Diploma and Degree levels to Aboriginal and Torres Strait

Islander students on-site in 12 locations in North Queensland and at Cherbourg in southern Queensland. Secondly, its main delivery mode is interactive multimedia (IMM) computer courseware which is characterised by the highly effective fusion of sound, animation, text, still and moving pictures, colour, interactivity, and feedback in ways that promote learning through multi sensory pathways of knowing (Putt & Henderson, 1993). Thirdly, it is the only accredited program in Australia in which IMM has a central role in all of its standard university subjects (and the TAFE equivalent). Fourthly, in Queensland, it was the first instance of a TAFE Associate Diploma being given accreditation towards a university award. Students with the two-year Associate Diploma of Education obtain the equivalent of one full year's credit towards the three-year Diploma of Teaching. The articulation of the various aspects of RATEP is shown in Table 1. Fifthly, it is a valued model of inter-institutional and community collaboration (Logan and Sachs, 1991; Willet, 1991).

Table 1
Articulation of the RATEP Program

Entry Status	Teacher Aides, Community Members	Community Teachers B	Community Teachers A		Teachers with a Diploma of Teaching
Year	1	2	3	4	5
Course	Certificate of Community Teaching	Associate Diploma of Education	[a]Diploma of Teaching (2nd Year)	Diploma of Teaching (3rd and Final Year)	BEd.(4th and Final Year)
Awarding Institution	Cairns College of TAFE	Cairns College of TAFE	JCUNQ	JCUNQ	JCUNQ
Graduation Status	Community Teacher B	Community Teacher A		Qualified Teacher (3 year trained)	Qualified Teacher (4 year trained).

[a]Community Teachers A who have the Associate Diploma of Education enter into the Diploma of Teaching at second year level.

The Students and Their Socio-Cultural Environment

Initially, the James Cook University RATEP program was aimed at upgrading the qualifications of Community Teachers in the Torres Strait, but now the program is offered to both Torres Strait Islander and Aboriginal students. Since 1991 the Associate Diploma of Education from the Cairns College of TAFE has been offered in an IMM mode to Aborigines at a number of communities in Queensland as well as Torres Strait Islander students at Bamaga.

The first cohort of eight Torres Strait Islander students graduated with a Diploma of Teaching in 1992. Since 1992, a second cohort of Torres Strait Islander students and a cohort of Aboriginal and Torres Strait Islander students have completed the course. In 1995, enrolment in the Diploma of Teaching has reached between 40 and 50. English is generally either the second or third language for all these students. In some instances traditional behaviour patterns such as the relationships between young and old as regards teaching and learning, an individual's right not to listen, group pressure not to excel beyond the group norm, and gender differences in such things as eye contact, disagreeing with elders and males in public, and working with relatives have been retained. These language differences and cultural traditions had to be acknowledged in the planning and development of the RATEP courseware materials as well as in the conduct of audio conferences and the setting of group projects in the mathematics education subjects.

Cultural Contextuality

There is a tension between Aboriginal and Torres Strait Islander desires for western education to prevent continued disenfranchisement in a modern technological society and their resistance towards such an education because it jeopardises their cultural knowledge and methodologies of teaching and learning. However, they acknowledge that cultural appropriateness for empowerment and ownership needs to include both western and indigenous knowledge and styles and conventions of learning.

The original RATEP authors attempted, in their curriculum design, to incorporate Torres Strait Islander and Aboriginal cultures, the specific requirements of an academic culture, and the cultural context of the computer. They strove for a coherent interplay between various implicit and explicit cultural logics (see Figure 1). In particular they sought to include Torres Strait Islander and Aboriginal knowledge and preferred ways of learning and doing in ways that had cognitive and affective validity. Secondly, the design sought to include the various genres and the culturally specific ways of promoting cognitive development within the lecturers' respective disciplines. Thirdly, features of the computer technology and the apparently culturally neutral software program, *Authorware Professional*, were exploited by the instructional designers of RATEP subjects to mediate, shape and facilitate what takes place in the human-computer learning paradigm. They were able to develop courseware that incorporated the first two cultural contexts to produce inter- and intra-psychological development.

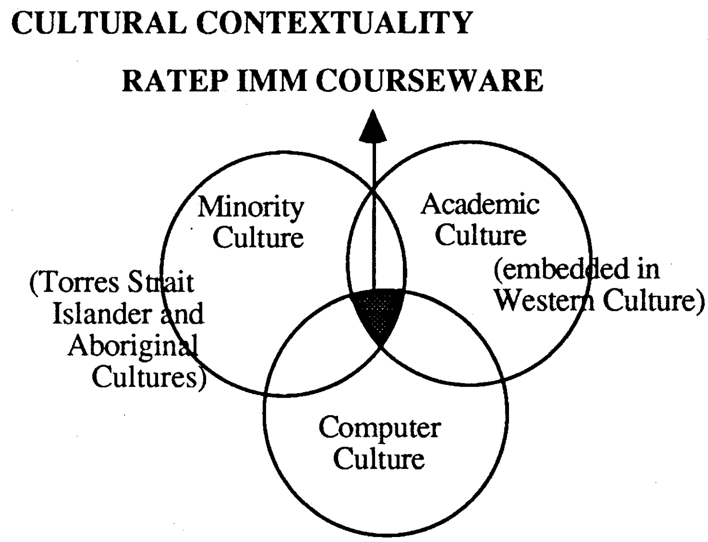
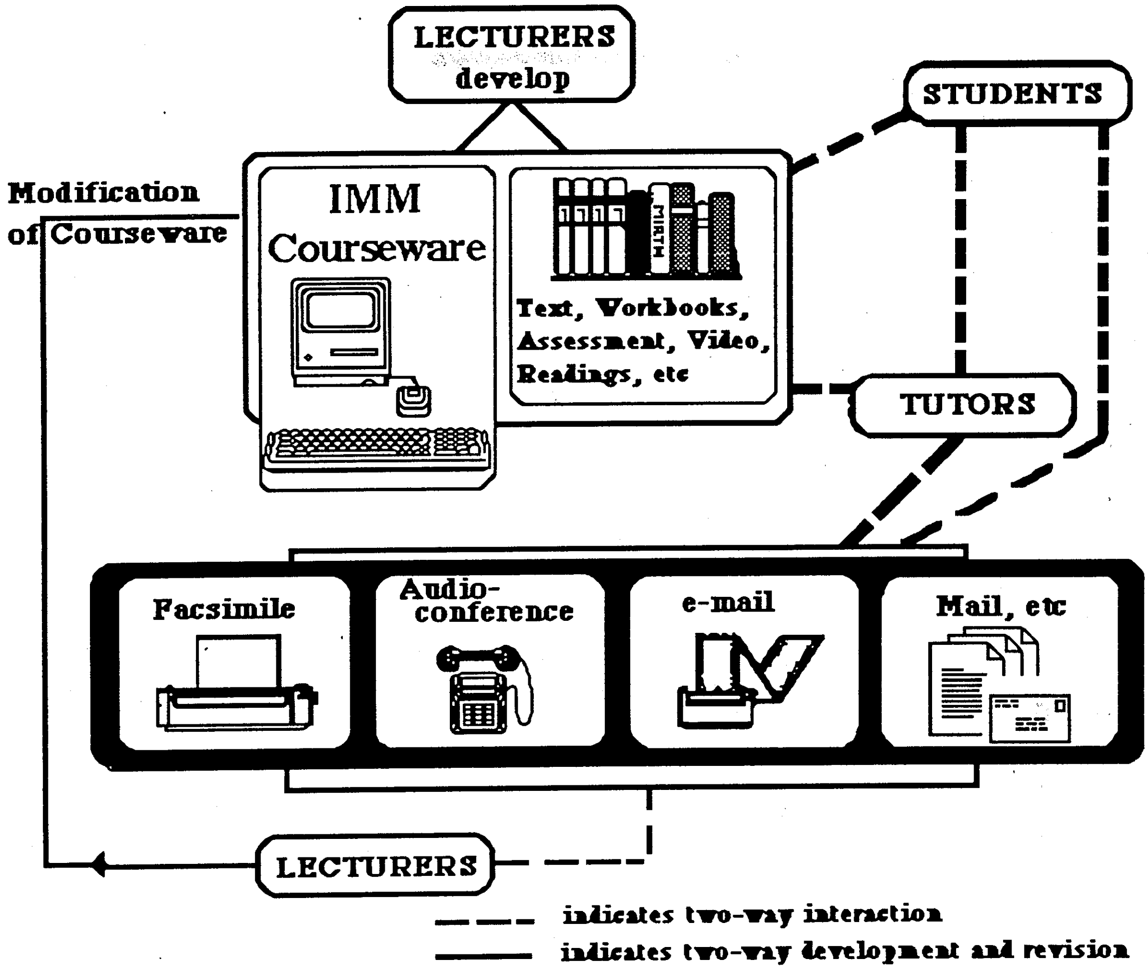


Figure 1. Cultural contextuality and RATEP IMM courseware

Course Production and Delivery

Initially, each RATEP subject was designed, produced and delivered according to the model illustrated in Figure 2 which is the curriculum paradigm developed by RATEP's first two lecturers (Macindoe & Henderson, 1991) and subsequently adopted as the basic model for subject development. Lecturers design and prepare written subject materials in collaboration with an instructional designer. These are given to the Centre for Interactive Multimedia to produce the IMM component and to prepare student workbooks and other resource materials including video if required. The completed IMM courseware is packaged on Compact Disk [CD] and delivered to the RATEP centres along with student materials and other resources. The Macintosh computers in the centres all have CD drives. Where subjects include a textbook, students obtain these from the University bookshop by mail order.



(Adapted from RATEP Feedback Network Model, Macindoe and Henderson, 1991, p.12)

Figure 2. RATEP Distance Education Curriculum Paradigm

Students work through the computer segments, workbook tasks, and assessment tasks either individually, with a peer, or in a tutorial group with the tutor. The evidence to date from students undertaking the mathematics education subjects is that the majority prefer to work alone on the computer for the initial attempt at the tasks because they want privacy if mistakes are made, they want to control the pace of their learning, and they want to test their own thinking (Putt, Henderson, & Stillman, 1995).

Communication with the lecturer and indeed with other students and/or tutors at other sites regarding any aspect of a subject can be done by facsimile, e-mail (though this has proved somewhat problematic), mail or telephone (Figure 2). Assignments are regularly submitted by facsimile. Students often rely on the facsimile machine to obtain academic support and quick feedback on their written work from their lecturers.

Lecturers conduct audio conferences either weekly or fortnightly. Their purposes include: clarification of points of difficulty, ascertaining levels of understanding, promoting discursive discussion, encouraging usage of the language of the particular discipline, discussing assessment criteria, and giving feedback on written work.

Technology in RATEP mathematics education

In the remainder of this paper we will discuss our usage of and experiences with the various technologies in our teaching of both mathematics and mathematics education subjects in RATEP. We will conclude by flagging areas of future development we see as essential to improve the delivery of these subjects and the learning of our students at a distance.

Interactive Multimedia [IMM]

As mentioned earlier, Interactive Multimedia (IMM) is defined as the effective fusion of sound, animation, text, still and moving pictures, colour, interactivity, and feedback in ways that promote learning through multi sensory pathways of knowing (Putt & Henderson, 1993). The *Authorware Professional,* software package exhibits all of these features. Following are the ways that these features were incorporated into the mathematics and mathematics education subjects in the RATEP course:

Sound

Sound effects take a number of different forms which have been classified by Bradey and Henderson (1995) as either verbal or non-verbal (see Figure 3). Verbal sounds are used for different purposes such as reading text as it appears on the screen (voice over with text); adding extra information to alert the learner to important points on the screen (voice over with extra information); commenting on graphics (voice over with graphics); and giving feedback on students' responses. Non-verbal sounds include background music; various sounds accompanying the press of a procedure button; various sounds which are heard when students type in a written response; various sounds which draw attention to text as it is revealed on the screen in point form or with animations; and sounds which may accompany feedback for a correct response. Both types of sound effects are included in the IMM courseware for the mathematics and mathematics education subjects but to date replication of the Bradey and Henderson (1995) research with these subjects has not been undertaken.

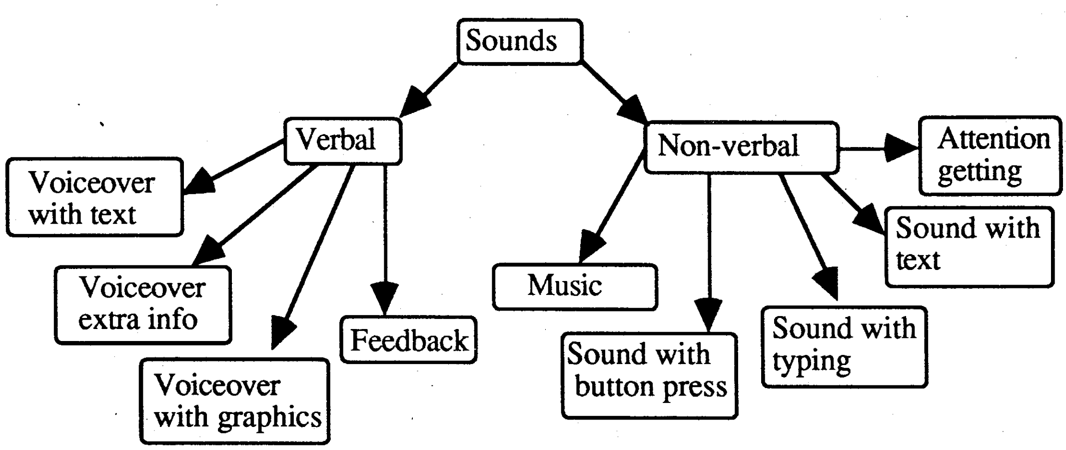


Figure 3. Sound Effects

Animation

Animation (or the ability to move text, pictures, diagrams) is used effectively in the mathematics subject for manipulation of pictures of concrete materials such as Multi base Arithmetic Blocks [MAB] when modelling regrouping for addition and decomposition for subtraction operations. When developing the concept of decimals it is used to demonstrate the breaking down of a unit into ten equal pieces to show tenths and vice versa to show trading and exchanging. In problem solving situations animation allows for the gradual build up of a diagram which may assist the student gain a clearer understanding of a problem.

Text, still and moving pictures, and colour

Text, still and moving pictures, and colour have been used separately or in combination with each other as well as with sound and animation. In some instances, photographs of the students doing mathematical activities in their own communities have been scanned into the computer courseware. Although there is scope for incorporating video segments into the courseware to provide students with vicarious experiences of mathematical situations not normally part of their lives, to date this has not been done largely because of costs.

Interactivity

Interactivity is defined by Henderson and Patching (1995) as any two-way communication or dialogue between the learner and computer. They argue that interactions can be either mathemagenic (give rise to learning) or non-mathemagenic (do not give rise to learning). Further they define mathemagenic interactions as those "computer-mediated activities that give rise to thinking and learning and which either allow or require from the user some overt response or action to which the computer can subsequently react" (Henderson & Patching, 1995, p. 1).

In the mathematics and mathematics education subjects the mathemagenic interactions consist of (i) text responses (words and/or numbers) entered into the computer by the student in answer to a query or a problem posed on the computer; (ii) click-drag activities generally involving sorting or classifying or concept attainment tasks, matching a sequence of steps in a lesson plan or ordering a sequence of steps in a problem-solving procedure; and (iii) multiple response items such as true/false, yes/no, or multiple choice. One of the distinctive features of the software is that feedback can be programmed into all of these interactions. Students receive constructive comments on both correct and incorrect text responses usually with the opportunity to 'try again' for an incorrect response. With click-drag interactions the IMM package either accepts the positioning of an object if it is correctly placed or the object is returned to its starting position for the student to have another attempt. When this happens students are challenged to examine their thinking before making another attempt.

Audioconferencing

Right from the establishment of RATEP audioconferencing was meant to be a part of the mode of delivery. However, because of poor telecommunication links with many of the islands in the Torres Strait some lecturers became disillusioned with this form of technology. Effective use of this form of communication in the mathematics and mathematics education subjects has grown as the two authors of this paper have experimented with different ways of using the audio conference time and as both students and lecturers have become more comfortable with the medium.

In the mathematics subject in first semester 1995 there were approximately 40 students from 9 different RATEP centres scattered across central and northern Queensland. The lecturer ran a weekly audio conference with four groups of 10 students plus their tutors. As the lecturer followed a problem-based approach to teaching mathematics, the audio conferences became lively discussions of students' solutions to problems which had been submitted by facsimile to her and which she then faxed to each of the audio conference groups as discussion starters for future conferences. As well as solving and discussing problems posed by the lecturer, all groups developed problem tasks themselves and shared these with other groups via the audio conferences and the facsimile machine. Despite the fact that the main means of communication is verbal and English is the second or third language of these students, they entered into mathematical discussions almost as effectively as in face-to-face situations. They also became more confident in attempting complexly worded problem tasks as they were encouraged in audio conferences to develop effective strategies to reduce the complexity of these tasks such as drawing a diagram to represent a scenario or brainstorming ideas as a group before starting an individual solution.

The lecturer in the mathematics education subjects has generally conducted audio conferences on a two-weekly basis with the main focus being discussion of any difficulties with the materials on the computer or in the students' workbooks which accompany the subjects. Requirements for student projects have always been negotiated via audio conference and facsimile and students have shared their mathematics teaching experiences while undertaking their teaching practice rounds via audio conference. A number of audio conferences in 1995 which were devoted to discussion of questions relating to required readings in the students' workbooks proved most effective in getting the students to critically examine materials which are used by teachers in schools.

Facsimile

Because some of the students live in remote areas which may receive only one airmail service per week, they see facsimile as the best way of receiving “instantaneous” support and feedback on their work. As indicated above, the facsimile facility has been used extensively in conjunction with audioconferencing to facilitate two-way communication between students and lecturers in both the mathematics and mathematics education subjects. Students also use facsimile to send most written assignments which allows them maximum time to complete the work and still meet the deadlines.

Video

Each RATEP centre has the necessary equipment to replay video tapes and most centres have access to a video camera. In one of the mathematics education subjects, the lecturer has included a video tape to demonstrate a particular type of lesson planning based on the "Excellence in Teaching" model. Students were also introduced to a model for planning and teaching problem solving lessons and have the opportunity to view the lecturer demonstrating this model with a primary class in a school near the university campus.

In 1994 the lecturer in the mathematics subject experimented with the use of audioconferencing and video taping for assessment of problem-solving tasks. These tasks and the original assessment ideas were based on the work of Lester, Kroll, Masingila, Mau, Moellwald, and Santos (1990). Students were given a group problem-solving task which yielded one solution for the group and a group mark for that solution. On completion of this task, each student was required individually to answer questions about the group solution which resulted in an individual mark. Each student therefore received two marks - one for the group solution and one for an individual solution. The next task involved students brainstorming a different problem without recording anything. The idea of the brainstorming was to enable students to get a good understanding of what the problem was about and which particular strategies may be useful in solving it. It also allowed input from the on-site tutors on any problems to do with language or culture associated with the problem context. At the completion of the brainstorming session the students were asked to solve the problem individually.

Because the lecturer found difficulty in identifying the individuals contributing to the discussion as she listened in on a conference telephone and because she was unable to view how students manipulated concrete materials or their body language, she encouraged the students to videotape their problem-solving session and to submit the tape for use in the assessment of the group solution.

Feedback from the tutor in the group that chose to participate in this video-taped session was quite positive. The video format was considered to be worthwhile by the lecturer who was able to observe students' problem-solving strategies and their overt thinking as they grappled with the problem. The fact that only one group chose to participate may be indicative of the fact that Torres Strait Islander and Aboriginal students were reluctant to be videotaped especially in a cross-cultural context where their behaviour may be misconstrued by the observer (Henderson, 1993a, 1993b, 1994). This fact exemplifies the importance of acknowledging the cultural context in which the learning takes place.

In 1995 the lecturer in the mathematics education subject required this same group of students to be videotaped teaching a lesson to a class or a small group of children to assist him in assessing their lesson planning and presentation skills. While the students were somewhat reluctant to do this, they realised that when they went into urban schools for a period of school experience later in the year they would be observed and assessed by classroom teachers who may be less sensitive to cultural contexts than the lecturer. The videotaping of these lessons was seen as a way of easing them into this experience.

Future Developments

Since the inception of RATEP, each subsequent cohort of students has become more familiar with the range of technologies used to deliver the subjects. As well as becoming computer literate with IMM which is the basic platform for the delivery of subjects, students, tutors and lecturers have developed confidence in their ability to interact using communication technologies such as the telephone and facsimile machine. As indicated in this paper, we believe that the teaching and assessing of the mathematics and mathematics education subjects have been enhanced by the use of audioconferencing and facsimile. However, we are aware also of some of the limitations of these technologies when trying to be truly interactive with the distance students.

Table 2 is an adaptation of one from Rhen and Towers (1994) which compared audioconferencing with videoconferencing and audiographics. In the left hand column are featured the types of communication and the X denotes whether we believe we have achieved this in the mathematics and mathematics education subjects using the two current media shown. A ? indicates we are unsure and a blank space means we do not believe we have done this or in some cases it is impossible. The three columns headed 'future' refer to technologies which we see as having real potential for improving the delivery of these subjects and others at a distance.

Initial attempts with e-mail via the Keylink facility of the Queensland Department of Education were unsuccessful. However, in 1995 all RATEP sites are being linked via modem to the James Cook University central computing facility which means that students, tutors and lecturers will be able to communicate with each other via e-mail. This opens up the real possibility of the authors interacting with their students using an electronic bulletin board. Furthermore, RATEP sites could be connected to the World Wide Web [WWW] which is the showpiece of the information superhighway, a global network of computers that communicate mostly via telephone lines. To date we have not explored the potential for using WWW in the delivery of mathematics and mathematics education subjects at a distance.

Table 2
<u>Usage in mathematics and mathematics education subjects</u>

		Current usage		<-------------------Future usage		----------------------->
	Communication Type	**Fax**	**Phone**	**Audiographics**	**e-mail**	**Video conference**
1.	Present information using graphic materials	X		X	X	X
2.	Negotiate	X	X	X	X	X
3.	View 3-D material			?		?
4.	Provide instruction -physical or graphic demonstration			X		X
5.	Persuade	X	X	X	X	X
6.	Affect someone	?	?	X	?	X
7.	Brainstorm	X	X	X	X	X
8.	Communicate cross-culturally	X	X	X	?	X
9.	Communicate with a group	X	X	X	X	X
10.	Review	X	X	X	X	X
11.	Develop a plan, process, project schedule model etc.	X	?	X	X	X
12.	Decide	X	X	X	X	X
13.	"Touch Base"	X	X	X	X	X
14.	Confirm	X	X	X	X	X
15.	Schedule	X	X	X	X	X
16.	Chat	X	X	X	X	X
17.	Comment	X	X	X	X	X
18.	Remind	X	X	X	X	X
19.	View computer information	X		X	?	X
20.	Interpret body language					X

Compressed video conferencing has been trialed in 1995 by one of the authors in the mathematics subject in the RATEP course. It has obvious advantages for teaching mathematics and mathematics education in that there are two way graphics as well as two way audio allowing both the students and lecturer to display graphics and text as well as to see and speak to each other. However, the high up-front costs and the limited availability of suitable specialised data lines to remote areas such as those serviced by RATEP have restricted its application. Until there is a reduction in such costs we do not see videoconferencing playing a major role in the future of RATEP.

On the other hand, audiographic conferencing appears to be a cost effective way of presenting information using graphic materials. Audiographics uses two standard telephone lines - one for audio and one for graphics - and normal telecommunications charges apply (Gooley & Towers, 1994). Users from multiple sites can be connected in a similar fashion to an audio conference. Audiographics involves the transmission of images and text between computers and is used in conjunction with an audio conference. It is thus a two way video, two way audio system by which lecturers can send textual information, diagrams or pictures to a computer so that other participants in the conference can view and edit these images. As can be seen from Table 2, audiographic conferencing has the potential to enhance existing communication as well as provide additional forms of communication but without the high up-front costs and the need for specialised data lines. The authors see this means of communication as having great potential for increasing the effectiveness of their teaching of mathematics and mathematics education.

Conclusion

As tertiary educators we are continually seeking to improve the teaching of mathematics and mathematics education particularly for students in remote areas. As we have outlined above the rapid advances in communication technology have great potential in assisting us reaching this goal.

References

Bradey, S. & Henderson, L. (1995). Voice-overs and auditory cues: Their impact and role in learning through IMM. In Proceedings of ED Media Conference 95 (in press) 17-21 June 1995, Graz, Austria.

Gooley, A., & Towers, S. (1994). Creating training opportunities for business and industry through audiographic conferencing. In Proceedings of Asia Pacific Information Technology in Training and Education Conference and Exhibition - Vol. 2, pp.523-529, 28 June - 2 July 1994, Brisbane, Australia.

Henderson, L. (1993a). Interactive multimedia and culturally appropriate ways of learning. In C. Latchem, J. Williamson, & L. Henderson-Lancett (Eds.), Interactive multimedia (pp. 165-182). London: Kogan Page.

Henderson, L. (1993b). Computers: A tool of empowerment in Aboriginal and Torres Strait Islander education.. In N. Loos, & T. Osanai (Eds.), Indigenous minorities and education: Australian and Japanese perspectives of their indigenous peoples, the Ainu, Aborigines and Torres Strait Islanders. (pp. 320-333). Tokyo: Sanyusha Publishing.

Henderson, L. (1994). Reeves' pedagogic model of interactive learning systems and cultural contextuality. In C. McBeath and, R. Atkinson (Eds.) Proceedings of the 2nd International Interactive Multimedia Symposium.(pp. 189-198). Perth, Australia: Promaco Conventions.

Henderson, M., & Patching, W. (1995). Multimedia interactivity: Relating cognitive processing to click-drag activities. In Proceedings of ED Media Conference 95 (in press) 17-21 June 1995, Graz, Austria.

Lester, F. K., Kroll, D. L., Masingila, J. O., Mau, S. T., Moellwald, F. E., & Santos, V. (1990). Mathematics for elementary teachers via problem solving. Bloomington, IN.: Mathematics Education Development Centre, Indiana University.

Logan, L., & Sachs, J. (1991). It opened my head: An evaluation of the first phase of the RATEP. Brisbane: University of Queensland.

Macindoe, M., & Henderson, L. (1991). Feedback to remote areas: Technology, the influencing factor. Educating Australia, 15, 12-16.

Putt, I., & Henderson, L. (1993, July). Culture, IMM and Mathematics Education. Invited paper presented at the 11th biennial conference of the Mathematics Education Lecturers' Association, Brisbane, Australia.

Putt, I., Henderson, L., & Stillman, G. (1995). Teaching mathematics education at a distance via interactive multimedia, video and projects. In Proceedings of the ICMI Regional Collaboration in Mathematics Education, Monash University, Melbourne, Australia.

Rhen, G., & Towers, S. (1994). Audiographic technology: The Cinderella of interactive multimedia.

Willett, F. J. Report to the Minister for Education on the Queensland open learning project. Brisbane, 26 March, 1991.

A FRAMEWORK FOR DEVELOPING THE INTEGRATION OF COMPUTER TECHNOLOGY AND THE MATHEMATICS CURRICULUM

Ted Redden and Garry Clark
University of New England
Armidale, 2351
Australia.
tredden@metz.une.edu.au
gclark@metz.une.edu.au

Abstract

In this paper a separateness between the computer microworlds and the traditional mathematics curriculum is identified as a result of a survey of high schools in rural New South Wales, Australia. A number of dimensions of this separateness are described and compared with reports of Hoyle and Papert. These dimensions are reported as six constraints that have hindered progress in integrating computer technology, available software and the mathematics curriculum.

To address these issues the Maths-Science Learning Cycle has been adapted to identify how computer technology can contribute to the traditional curriculum and how the computer based activities must be supported to provide an effective set of learning experiences. The focus of this model is not just on providing a 'unit of work', but it is also on assisting in teacher development.

Finally a simple example of a Learning Cycle is presented which attempts to make explicit the way the Learning Cycle can be used.

Introduction

> I love using computers with the kids. They really seem to get a lot out of some programs. But we really don't have time to do much of it. The maths syllabus is very full and it uses up all of our class time.
>
> A 'computer literate' mathematics teacher.

This quote encapsulates a separateness that exists in the schools in our region of rural New South Wales, Australia. This separateness is between computer microworlds on one hand, and the traditional mathematics curriculum on the other hand. This paper reports on a situational analysis that identifies some issues underlying this separateness and develops a framework to facilitate the integration of modern computer applications and the traditional mathematics curriculum. This framework is adapted from the Maths-Science Learning Cycle, (Hill 1986) and it provides an opportunity to make explicit what role a particular piece of software might play in developing mathematics concepts within the learning cycle. In constructing our approach the need for teacher development is recognised as well as consideration given to the learning process in children. The final part of this paper presents some activities, based on Geometer's Sketchpad, that are integrated into the Learning Cycle and which address specific curriculum content. Early indications are that this integrated approach is accessible to teachers, who have previously seen the curriculum and computer microworlds as separate.

Situation Analysis

In spite of computers having been available for some 15 years, they seem to have made little headway in changing the way mathematics is taught in schools. This lack of development has been apparent to the writers, who in a survey of high schools in their region (rural N.S.W. Australia) have found a consistent pattern emerging. This pattern involved extensive use of computers for administration services, the existence of one or two computer laboratories, each consisting of, typically, 20 computers. These computer laboratories were predominantly used for computer literacy and computer studies courses. Computer literacy courses were to a greater or lesser extent, a compulsory component of the schools curriculum. These courses were characterised by developing keyboard skills and familiarity with a variety of software applications such as wordprocessors, databases and spreadsheets. Lesser attention was given to the social implications of a rapidly developing technology. Computer studies courses, characterised by a study of the functioning and applications of computers and the many social issues that result from the introduction of technology, are offered as electives in most schools and studied by some students in both of senior and junior years. These courses are increasingly making demands on the schools computer laboratories.

What was particularly noteworthy, was the separateness of mathematics teachers' interest in computers and their mathematics teaching. Mathematics teachers were often actively involved in using computers in an administrative capacity, or closely related to the computer literacy and studies courses and were often reported to be the Computer Coordinator within their school. There was almost no use of computer software in the teaching of mathematics. The software that was being used was predominantly behaviourist in its orientation and was used spasmodically as a text book replacement. Some teaching of BASIC programming was being undertaken as part of optional lobes within the standard curriculum. There were no teachers developing and using microworlds as advocated by Papert (1981).

If the overall use of computers in the mathematics classroom were limited, then the use of computers to develop concepts and understandings were even more restricted. Two teachers reported attempts to teach introductory algebra within a spreadsheet context. Support was equivocal and it seemed computer access was perceived to be a major inhibitor to further experimentation, along with "spending too much time teaching how to use the spreadsheet" rather than developing algebraic concepts. Another school had borrowed a set of graphic calculators from the local university and had trailed their use in a senior calculus class. The reports were enthusiastic but no commitment had been made to acquire the necessary equipment in the immediate future to ensure the ongoing development of the initiative. Further probing revealed that the students had very quickly become competent users of the calculators and had "discovered" some interesting mathematics. However, pressures of time forced the teacher to focus on the formal curriculum. It seemed that the teacher was unable to integrate the graphics calculator experience with the curriculum expectations of the state syllabus.

Most schools reported having some software that could be seen as having some potential to assist in the teaching of mathematics. Software commonly reported was Logo, Geometer's Sketchpad, A.N.U. Graph and spreadsheets. In spite of owning the software and having some expertise with computers, the use of this software was not being integrated with the curriculum in a systematic manner, nor was it being used in a manner consistent with constructivists' view of teaching/learning. It's use was characterised as 'incidental.' That is, the occasional mathematics period was spent in what was euphemistically called 'exploring the software.' However, that use was not directly related to the formal curriculum, nor was the students' exposure to the software prolonged.

These findings in a rural Australian setting seem to be largely consistent with the situation reported by Papert (1993) and Hoyle (1994) who have both expressed concern at the lack of realisation of the potential of new ways of teaching and learning, based on computer environments.

Papert identified two major reasons for this lack of progress. The first, was the relatively few computers in schools (students out number computers by 50 to 1). This is worse than our findings of approximately 40 computers for student use in high schools of six to eight hundred students. However, these computers were dominated by the computer studies classes and only restricted time was available for use in mathematics classes. This use of computers is indicative of Papert's second reason for lack of progress. Instead of improving the curriculum by making it a tool for exploration, the computer was used to reinforce traditional school values and practice.

> What had started as a subversive instrument of change was neutralised by the system and converted into an instrument of consolidation.
>
> (Papert 1993, p. 39)

The problem is not one of lack of teacher skills in the use of computers in the technical sense, nor one of the unavailability of software (although this needs constant attention and development). The problem is more one of inertia within the school system which rejects revolutionary change. One dimension of this "defence against foreign bodies" is the existence of a curriculum defined in terms of outcomes, delivered by transmission and reinforced by external examination. These formal constraints allow teachers to feel comfortable, allow supervisors to control and politicians to react to accusations of falling standards. Hoyle's somewhat cynical view of the manner in which Logo was included in the UK national mathematics curriculum supports this view. In a process that Hoyle calls "incorporating" (Hoyle 1994, p. 178) an innovative vehicle for exploring mathematical ideas was reduced to some relatively trivial behavioural objectives. Hoyle also reported processes of "compartmentalizing" (p. 177) and "neutralizing" (p. 179) as components of the transformation of curriculum innovations. Compartmentalizing is a coping strategy in which the innovation is marginalised and kept out of the mainstream, and neutalizing is a process of reducing the perceived value of an innovation by failing to recognise its full potential e.g., computers are "just a tool".

Hoyle (1994, p. 180) argues that attempts to overcome these problems by producing curriculum packages that bypass the teacher have failed, since it ".... leads to the separation of the bottom up spontaneous mathematizing of pupils from the top down specifications of the curriculum." She calls for a different vision of teacher development that recognises change as a "process not an event." This reflects a counterpoint of evolutionary change to the revolution referred to above.

Papert suggests that since we espouse a constructivist curriculum for children, we ought to consider the same developmental approach to school practice.

> Like good developmental teachers, researchers can contribute best if they understand change in School development, and support this by transferring the ideas that were successful for understanding change in children.
>
> (Papert 1993, p. 41)

The challenge is to identify a structure that will facilitate the incorporation of exploratory software into the existing mathematics curriculum. Such a structure must recognise a number of constraints that have been recognised above:

a) The existence of a specified curriculum and time constraints on its implementation.
b) The teacher's need to be involved in the learning process in a variety of ways.
c) That teacher change is incremental in nature.
d) Teachers need to be given opportunities to construct for themselves the purpose and value of computer microworlds and their subsets.
e) That computers can play a number of roles in teaching and learning-they can be more than just a tool for drill and practice. They can be a medium for problem solving and for the construction of new concepts.
f) That computer facilities will not always be available to the mathematics teacher, and hence their use must be specifically defined and capable of being planned.

A structure for contributing to these issues is the Learning Cycle which is discussed in general terms prior to describing how some computer based activities can be incorporated into the model.

The Learning Cycle

The learning cycle is a means of synthesising the diverse range of information about "content, children's learning, teaching method and the culture of the discipline" (Boylan, Francis & Hill 1990, p. 6). The concept of a learning cycle has evolved over many years. Howard (1927) advocated the structure of a phased development cycle, and more recently it has been used in a diversity of settings and adapted to reflect research reports on constructivism, meta-learning and the organisation of knowledge, and reports on using the learning cycle itself (Boylan, Francis and Hill 1990). This adaptation is called the Maths-Science Learning Cycle (Hill et al. 1986). This model is based on the earlier work of Wilson (1976) and the Science Curriculum Improvement Study (1977) and parallels to this approach can be found in the work of Van Hiele (1986) and Dienes (1971) (Pegg 1995).

The Maths-Science Learning Cycle is described briefly here in Table 1 which has been adapted from the presentation of the originators of the model (Hill et al. 1986). This manifestation of the Learning Cycle has been chosen for a number of reasons.

a) There is some research support for its effectiveness (Boylan 1988; Redden 1987).
b) The language of its phases has been adapted to be accessible for teachers.
c) The specific roles of phases, learners and teachers have been identified.
d) It has been designed for use across a number of educational settings.
e) It provides guidance as to the appropriate learning activities and explicitly includes exploratory and open ended inquiry activities.
f) It specifically attempts to incorporate both traditional and inquiry methods of teaching and learning.

The Learning Cycle therefore advances the "traditional-teaching versus inquiry-learning debate beyond a simple for or against" (Boylan 1990, p. 17) and in doing so facilitates the gradual development of new ideas, teaching practice, and the technologies into the more traditional curriculum.

In addition, the Maths-Science Learning Cycle addresses the issues raised in this paper and the problems identified in the situation analysis above, by identifying where computers can be incorporated into the teaching sequence, the role of computer activity and hence in the selection of software. (It may also provide useful indicators as to areas in which software needs to be designed). It assists in identifying the additional activities and functions of the teachers needed to supplement and complement the computer based activity. It provides opportunities for teachers to explore the potential of computer microworlds with the support of some traditional teaching activities. In doing so it enables the change to evolve in a manner that reflects the teacher's individual values and style (Hoyle 1994).

While the work in this project is at an early stage the hypothesis underpinning the study suggests that the role of the computer activities will expand as the teacher becomes more familiar and confident with the environment. Additionally the amount of direction given to children in the early exploration phases will decrease as teachers develop a greater openness in the exploration of ideas and children become used to learning in a more open environment.

An Example of a Learning Cycle

This section outlines some activities that have been designed on the readily available Geometer's Sketchpad (Jackiw, 1991). The way these relate to the Maths-Science Learning Cycle is also made explicit. These activities are seen as suitable for the earliest stages of teacher development and hence are quite highly structured. They assume little familiarity with Geometer's Sketchpad, in fact approximately 10 minutes familiarisation with the software seems adequate preparation for the initial activities.

Central to the use of Geometer's Sketchpad below is the provision of an 'indestructible' figure. Each sketch is constructed in such a way that the defining properties of the figure

Table 1. Details of the Mathematics-Science Learning Cycle
Adapted from Hill, Redden, Francis, Baker &Spikler (1986)

Phases and Purpose	Learning Activities	Children's Role	Teacher's Role	Computing Implications
Finding out about the learner's knowledge, language and attitudes to the central Idea	Students respond to situations that involve the idea	To use past experiences to reflect on the idea. They should respond openly by talking, taking risks and predicting .	Provide situations embodying the concept and an environment which encourages and accepts students' responses	Make explicit the computer skills necessary. Develop some familiarity with software. Perform necessary prerequisite computer skills.
Exploring the Idea Learners encounter instances of the new idea, record and discuss the experiences in their own language	Students explore a range of situations which embody the idea, eg. conducting experiments, modelling, group discussions, role playing	Accept role of investigator by exhibiting such behaviours as observing, measuring, modelling, recording, discussing, predicting, analysing and questioning	Provide situations to explore as well as means to investigate them. Listens to and values students' responses, asks questions such as what if..., provides encouragement and guidance.	Computers can provide many dynamic environments for exploring new ideas. Geometers' Sketchpad, Logo and ANU Graph are examples of software that facilitates exploration.
Getting the Idea involves making the new idea explicit and introducing new language	Discuss outcomes of investigations, Testing inferences and hypotheses, Expressing ideas using own language. Refining statement of idea using appropriate language	To commit themselves to acquiring the idea by generating statements or representations of the idea, incorporating new language into statements and structuring the idea for themselves	Teacher plays direct role in establishing and clarifying the new idea by focusing on common elements of the investigation, summarising responses, introducing the new language monitoring the development of the idea	Results of computer investigations are shared. Other students' findings can be tested and checked.
Organising the Idea Develop a degree of permanence and confidence with the new idea, relating it to previous knowledge	Activities involving communicating the idea through posters, models, explaining to others etc. Practicing the idea through games, drills worksheets , assignments. Analysing situations for examples and non examples of the idea. Interchanging representations of the idea	To be aware of their own knowledge and the way they are thinking about it by feeling confident in using the idea, knowing when and how to use the idea, being able to explain it to others, being able to link it with previous ideas.	To provide situations and time for learners to reflect on their understanding of the idea. Activities should facilitate links between new and old ideas. Encouraging students to practice and communicate the idea.	Connections to previously investigated computer investigations are made and ideas tested. Drill and practice software packages could be used. Appropriate computer games used to consolidate the idea.
Applying the Idea Apply and identify the idea in new and novel situations	Situations which require the learner to recognise the potential use of the idea, problem solving, Games, Using the idea, with others, to generate new insights	To use the idea in a variety of situations including, identifying instances when the idea might be used or applied in conjunction with other ideas and comparing and evaluating contexts in which the idea might be applied.	Teacher creates situations which require the learner to identify where the idea might be used. Provides encouragement and guidance in its use.	More sophisticated problems using the software can be solved. Use software to review ideas if difficulties are encountered. Use computer environment to solve problems involving flexibility and reversibility

under investigation remain invariant. The points and sides of the figure can be moved to generate an infinite set of such figures. Manipulation of these figures provides an environment in which the learner can generate conjectures, test hypotheses and generalise results.

It is assumed that children have developed some familiarity with the properties of general quadrilaterals and trapeziums, and the associated geometric language in a general unit on quadrilaterals prior to attempting a learning cycle on the properties of parallelograms. The unit could (ideally should) commence with manipulating a quadrilateral in Geometer's Sketchpad and be followed by a sketch of a trapezium. The sketches would also provide the opportunity to develop skills with the software that are necessary for investigating other figures to be used in the unit. Such skills include selecting and dragging points and segments. In this way new skills with the software are gradually introduced as each cycle is undertaken.

Initially each sketch could be supported with on screen measurements and a worksheet, but as the learners progress through the unit this support should be reduced and even eliminated as familiarity with the software increases, the nature of the activities are appreciated, and investigative skills developed. For example, when the parallelogram sketch is introduced, on screen measurements could be given (see Figure 1) together with a worksheet to provide direction to the activity. If the learners are more familiar with open investigations, only a sketch of an indestructable figure (see Figure 2) should be provided, with learners asked to use their software skills and past experiences with other quadrilateral sketches to investigate its properties. In this way the learners are gradually being introduced to more open environments where they can hypothesise, record, generalise and report verbally - all important skills that learners need to develop. Teachers also are gradually moving toward providing more open investigative environments.

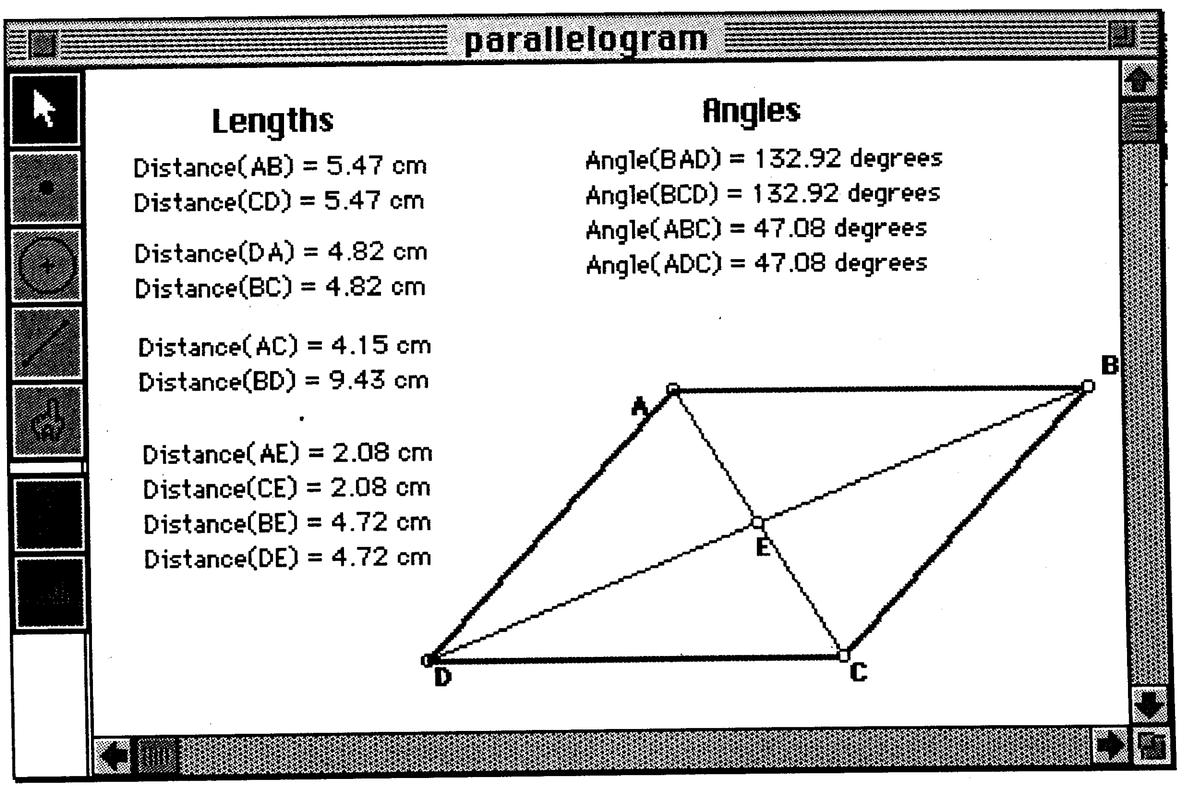


Figure 1. Parallelogram sketch with measurements

A variety of experiences are required to complete the Learning Cycle framework. An overview of these experiences are presented as Figure 3. Consider the activity of finding properties of parallelograms. In the first phase, **finding out about the learner**, the teacher should ask questions to determine any informal definitions of parallelograms the learners may

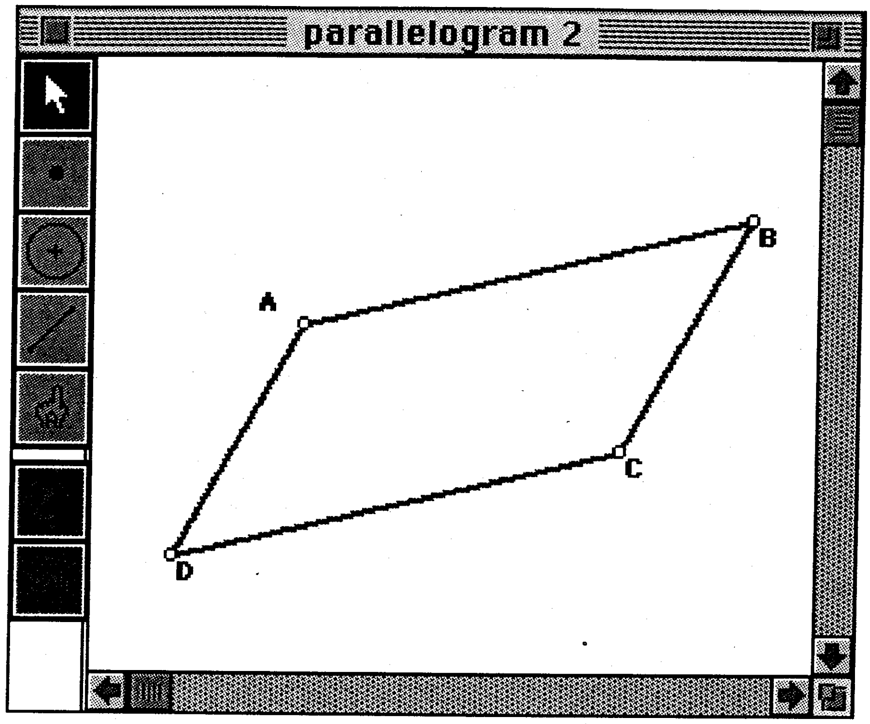


Figure 2. Parallelogram sketch with measurements

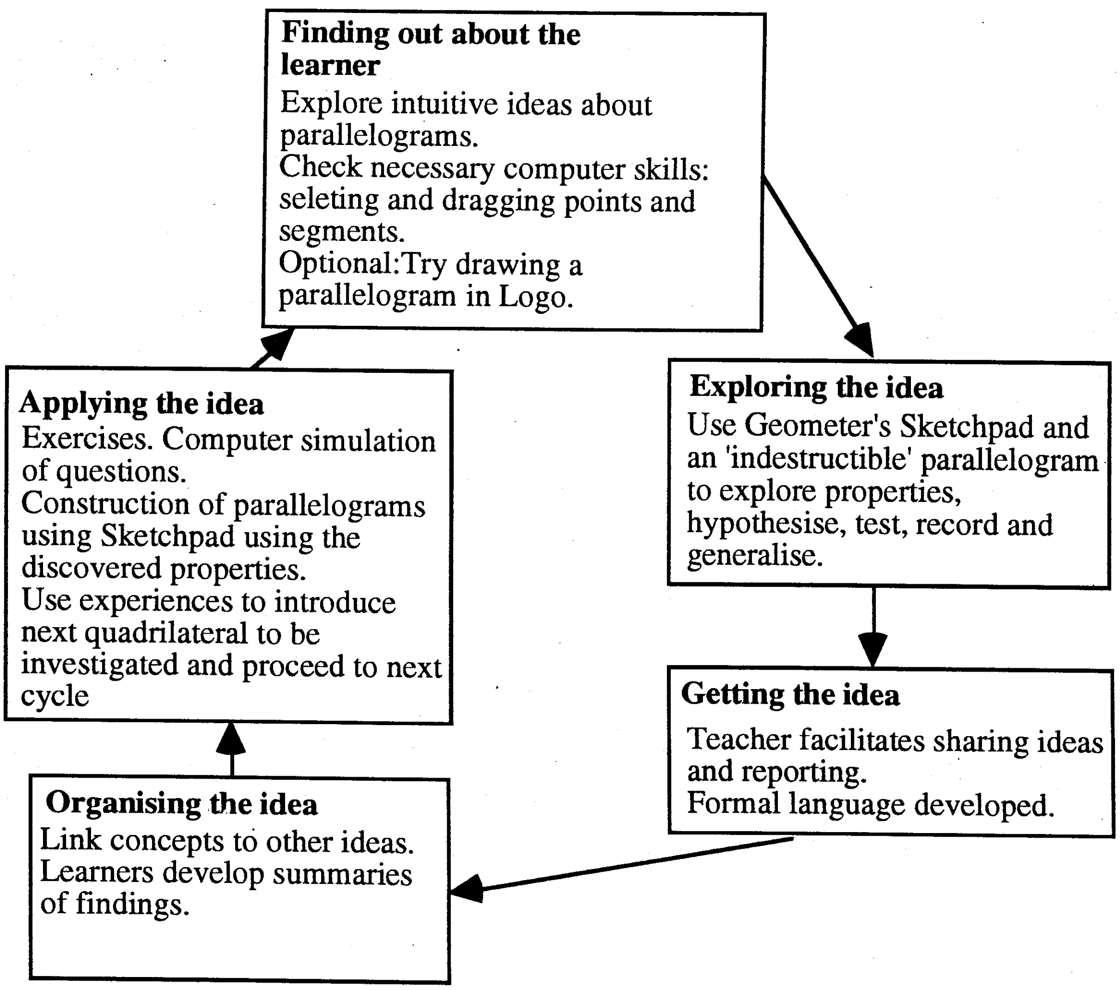


Figure 3. Overview of an M-S Learning Cycle for properties of a parallelogram.

have, and to determine any intuitive ideas about properties of parallelograms. Geometer's Sketchpad skills such as selecting and moving points and segments should be checked. New skills such as constructing and measuring segments, constructing points of intersection and measuring angles should be introduced or revised if Figure 2 is to be provided. Writing a Logo procedure to draw a parallelogram is an appropriate activity in this phase if the learners and teacher have the necessary skills.

It is during the second phase of **exploring the idea** that the 'indestructible' parallelogram sketch, Figure 1 or Figure 2, is introduced. The sketch is constructed in such a way that a large number of parallelograms can be investigated with a minimum of skill in using the software. Moving any vertex or side of the parallelogram generates a new figure, but at all times the figure remains a parallelogram.

Students are asked to investigate properties of the parallelogram. The focus of such inquiry includes the investigation of properties such as opposite sides are equal, opposite angles are equal, adjacent angles are supplementary, and diagonals bisect each other. Using the new skills and being supported by a directed worksheet with questions such as,

> Measure angle(ABC) and angle(ADC). Now select and drag points and segments of the parallelogram. What can you conclude about these two angles?

and

> Construct the diagonals AC and BD and their point of intersection. Measure the diagonals and any other segments of interest. Now select and drag points and segments of the parallelogram. What can you report about the diagonals?

students are be able to report and generalise their findings.

The computer does not have a major focus in the third phase of **getting the idea**. Here the teacher would facilitate the sharing of ideas and develop the more formal language necessary for communicating ideas accurately. It may be possible that specific ideas reported by learners could be demonstrated on the computer and discussed or counter examples may be demonstrated using the computer. This is a time for reporting and summarising ideas, and developing language.

Having established the various properties of a parallelogram it is time to **organise the ideas**. Questions aimed at making connections between any of these properties and previous work on parallel lines cut by a transversal, or questions seeking comparisons with previously investigated figures, are asked in this phase. The trapezium sketch could be revisited to test whether trapeziums also have any of the properties of a parallelogram. Other activities undertaken in this phase could be written reports on findings, designing a poster comparing and contrasting the parallelogram with other quadrilaterals previously investigated.

The last phase of **applying the idea** includes using the properties to solve standard numerical textbook exercises. If difficulties are experienced the learner could return to the parallelogram activity on the computer and simulate the question. Now that it is known that opposite sides and angles are equal, Logo procedures for a general parallelogram could be revisited. Other activities that could be undertaken in this phase include constructing a parallelogram using geometrical instruments, constructing a design using parallelograms and constructing a parallelogram using Geometer's Sketchpad.

Problems such as

> A figure has one pair of opposite angles equal. Is it a parallelogram? Why?
>
> If the ends of two segments bisecting each other are joined, will a parallelogram be formed? Why?

and

How many different ways can you construct a parallelogram?

are able to be considered. In addition investigating the steering mechanism on a Lego car could prove worthwhile. Where possible, making explicit how ideas are used in our world is seen as essential in this phase.

Later activities on the kite, rhombus, rectangle and square should not need support measurements on the screen and as learners become familiar with this type of activity they should not require the directed worksheet. Only the 'indestructible' figure need be supplied with the instruction to investigate properties of the figure. Students develop their own methods of searching, reporting and generalising their findings.

Conclusion

The Geometer's Sketchpad activities developed provide dynamic diagrams that are infinitely variable but have certain invariant properties. They allow extensive exploration, hypothesising, hypothesis testing and generalising in an empirical environment rather than a formal deductive environment. However, the opportunity to explore concepts are provided in a context that allows the teacher to gradually build confidence in both the software and approach since the curriculum remain central (at least initially) to the teaching/learning process. Additionally, the precise role of the software and the computer environment are specified and the additional teacher support activity is made explicit.

It is expected that as teachers' confidence builds, and as they become familiar with both the software and its potential, resulting in teaching style changes, they will be in a position to move more easily to the innovative microworlds advocated by Papert and Hoyle.

References

Boylan, C. (1988). Enhancing learning in science. *Research in Science and Technological Education*, 6, 205-217.

Boylan, C., Francis, R., & Hill, D. (1990). A Case of Curriculum Synergy. In C. Boylan (Eds.), *Learning Cycles in Primary Classrooms* (pp. 5-26). Wagga Wagga: Charles Sturt University-Riverina.

Dienes, Z., & Golding, E. (1971). *Approach to Modern Mathematics*. New York: Herder & Herder.

Hill, D. M., Redden, M., Francis, R., Baker, S., & Spikler, T. (1986). Science and mathematics: Developing a common approach to instruction. *Journal of Science and Mathematics Education in South East Asia*, 9(2), 38-50.

Howard, R. (1927). The unit method of instruction as applied to the teaching of physics. *School Science and Mathematics*(27).

Hoyles, C. (1994). Computer-Based microworlds: A radical vision or a trojan mouse? In D. F. Robitaille, D. H. Wheeler, & C. Kieran (Eds.), *Selected Lecture from the 7th International Congress on Mathematical Education* (pp. 171-182). Sainte-Foy: Les Presses De l'Universite Laval.

Jackiw, N. (1991). The Geometer's Sketchpad. In Berkeley: Key Curriculum Press.

Karplus, R. (1977). Science teaching and development of reasoning. *Journal of Research in Science Teaching*, 14(2), 169-175.

Pappert, S. (1993). *The Children's Machine*. Basic Books.

Pegg, J. (1995). Learning and teaching geometry. In L. Grimison & J. Pegg (Eds.), *Teaching Secondary Mathematics: Theory and Practice* (pp. 87-103). Sydney: Harcourt Brace.

Redden, M. (1987). Using the learning cycle: An implementation study. In *Mathematics Education Research Group of Australasia*, . Townsville:

Van Hiele, P. (1986). *Structure and Insight: A Theory of Mathematics Education*. New York: Academic Press.

Wilson, J. (1976). *Diagnisis and Treatment in Arithmetic: Belief, Guiding Models and Procedures*. Maryland: University of Maryland.

The relation in the learning process of mathematics for engineering Students at the Institute of Technology Sepuluh Nopember (ITS).

(Pre research for improving teaching mathematics)

Susanti Linuwih
(Institute of Technology Sepuluh Nopember, Surabaya Indonesia)

ABSTRACT

Teaching mathematics has been given since kindergarten, primary school,high school up to University degree in the field of engineering and commerce.

The number of Students who like mathematics increase in accordance with the increasing of the educational level from primary to high school.

To overcome the difficulty in Studying mathematics in primary school the children are guided by their parents, where in secondary and high school they have a study club.

Mathematics subject in primary school up to high school are related with mathematics in the University.

Presented at the ATCM, Singapore 18 - 21 December 1995

The relation in the learning process of mathematics for engineering Students at the Institute of Technology Sepuluh Nopember (ITS).

(Pre research for improving teaching mathematics)

Introduction

Basic and simple mathematics is given in Kindergarten and it is continued to the primary school and high school with a general objective as follows :

1. To prepare the Students for the changes of situation in their life as well as the changing world, through exercises of thinking logically, rationally, critically accurately, objectively, creatively and effectively.
2. To prepare the Student in order to be able to use mathematics properly in their everyday life and in Studying Science (GBPP 1986)

Mathematics at the primary School is given from the first year up to the sixth year, with a rate of 6 hours/week.
In the first and second year, one teaching hour is equal to 30 minutes. In the third year one teaching hour is equal to 35 minutes, while in the fourth, fifth and sixth year one teaching hour is equal to 40 minutes.

Mathematics in the secondary school is given from the first year up to the third year with a rate of 6 hours/week for un even semester and 4 hours/week for even Semester. One teaching hour is equal to 45 minutes.

Mathematics at the high school is given from the first year up to the third year with a difference in total hours. In terms of the core subjects, general high school is divided into four programmes namely; Physics (A_1); Biology (A_2); Social (A_3) and cultural Subjects (A_4). These programmes are determined after the Students have passed their first year.

Note : In the new curriculum 1994 these programmes are determined after the students have passed their second year. They are : Science and Mathematics; language and Social.

To take the programmes of A_1 or A_2, the students have to fulfill the following conditions :

- The average score for all subjects should higher than 6 (maximum score is 10)
- The average score of mathematics, physics, biology and chemistry must be more than 6

The Table 1 below shows the total teaching hours, at high school.

Table 1 : Total teaching hours

Semester	Total teaching hour							
	First Year	Second Year				Third Year		
		A_1	A_2	A_3	A_4	A_1	A_2	A_3
I	72	108	72	48	36	144	108	72
II	64	95	64	56	32	48	48	24

Students who have passed the final high school exams from A_1, A_2 and vocational school for engineering can enter the faculty of engineering.

Mathematics at the faculty of engineering is given from the first Semester up to the fourth semester with a rate of 4 credits for the first and second semester; 3 credits for the third and fourth semester. Mathematics is a basic sporting subject at the faculty of engineering

ITS is one of the engineering Institute in East Java.
The curriculum of ITS becomes the standard curriculum of all private Universities in East Java.

To make the function of mathematics more stable in primary school up to University, we need to know the interest of students from primary school up to University and the difficulty in Studying mathematics I,II,III and IV at the faculty of engineering.

This research has not donated directly for the development of technology, but it is a pre research as a purpose for improving teaching mathematics at the faculty of engineering.

According to the above problems, the question that we need to answer in this research is as follows

1. What is the relation between the attitude towards mathematics and the level of education of the ITS student
2. Which part of the mathematics they think difficult and how to overcome it.
3. Is there any relation between mathematics at primary school up to high school and faculty of engineering ITS.

In accordance with the research questions, the purpose of this research is

1. To obtain data about the student's towards mathematics when they were in primary up to high school and to overcome the difficulty in studying mathematics when they were in primary up to high school.
2. To obtain the student's difficulties when they were in primary school up to high school and University
3. To obtain the student's opinion about the relationship between mathematics in high school with in engineering faculty.

Discussion

1. Methodology

The Population in this survey is all of the ITS Students who are in the seventh session upwards from Faculty of Design and Civil engineering department of Civil, Faculty of industrial technology department Electrical, Mechanical, Chemistry, Industry and Computer, Faculty of Oceanology department of ship building. Using sample random sampling we took 20 students as a sample from each department. They are chosen from each department. Out of 140 students involved in this survey only 118 returned the questionnaire.

2.1. Data Analysis

We discuss the result of the questionnaire using statistical descriptive and inference by using cross tabulation for students when they were in primary school up to high school and university degree.

2.1.1. When the Student were in primary school up to senior high school

a. From the questionnaire data was obtained about the student interest in mathematics from primary school up to high school

Table 2. Percentage of student interest in mathematics beginning from primary school up to high school

Level	Normal	dislike	like
Primary School	35,6%	6,8%	56,8%
Secondary School	26,3%	11,0%	61,9%
High School	26,3%	5,1%	67,8%

The data in table 2 showed that students interest in mathematics increased from primary school up to high school. Using the statistical analysis cross tabulation, we can see the correlation between the students' interest in mathematics when they were in primary school with secondary school, primary school with high school and secondary school with high school. The correlation between the students' interest in mathematics in primary with secondary school, primary with high school and secondary with high school

H_0 : there is no correlation between the students' interest in mathematics in primary with in secondary school

H_1 : there is a correlation between the students' interest in mathematics in primary with in secondary school

X^2_{cal} = 77,151 and $X^2_{(4;0,05)}$ = 9,488

Conclusion Ho is rejected, that means there is a correlation between the students' interest in mathematics in primary with in secondary school

H_O : there is correlation between the students' interest in mathematics in primary with in high school .

H_1 : there is a correlation between the students' interest in mathematics in primary with in high school

X^2_{cal} = 18,72 and $X^2_{(4;0,05)}$ = 9,488.

Conclusion : Ho is rejected, that means there is a correlation between the students' interest in mathematics in primary with in high school

H_O : there is no correlation between the students' interest in mathematics in secondary with in high school

H_1 : there is a correlation between the students' interest in mathematics in secondary with in high school

X^2_{cal} = 41,79 and $X^2_{(4;0,05)}$ = 9,488

Conclusion : Ho is rejected, that means there is a correlation between student's interest in mathematics in secondary with in high school

b. From the questionnaires we find out the data about how to overcome the difficulty in learning mathematics in primary up to high school

Table 3 : To overcome the difficulty in learning mathematics

In	Extra lesson		Study club		Guided by	
	Yes	No	Yes	No	Parent	Brother /Sister
Primary School	19,5%	60,2%	28,8%	44,1%	37,3%	22%
Secondary School	11,9%	61,9%	32,2%	50,0%	14,4%	16,9%
High School	24,6%	48,3%	39,8%	39%	5,1%	14,4%

The data in the table 3 shows that overcoming the difficulty in learning mathematics with extra lesson and study club increases as well as the level of education, where as the parents guidance decreases at either secondary school or high school.

C. From the questionnaires we find out the data about the material of mathematics which is difficult when they were in primary, secondary and high school.

Table 4 : The difficulty of mathematics material in primary school up to high school

Material	primary	secondary school	high school
Fraction	9,3%	0,8%	
Stories problem	22,0%	3,4%	0,8%
Inequality	-	2,5%	-
Real number	1,7%	0,8%	0,8%
Solid geometry	3,4%	17,8%	0,8%
Plane geometry	-	1,7%	-
Statistics	-	2,5%	3,4%
Probability	1,7%	1,7%	1,7%
Differential	-	0,8%	4,2%
Integral	-	1,7%	0,8%
Linear programming	-	0,8%	1,7%
Matrix	-	0,8%	-
Vector	-	1,7%	2,5%
Quadratic equations	-	0,8%	0,8%
Logarithms	-	-	0,8%

2.1.2. Student at the University degree

a. From the questionnaires, we find out that 78,8% of students explained that the mathematics material in primary school up to high school support the mathematics in University.

b. From the questionnaires, we find out that 84% of students explained that they do not use computer as a tool to solve problem in mathematics or to draw a graph, but 76,3% of students explained that they know that to solve the problem in mathematics they can use computer.

c. From the questionnaires, we find out that 4,2% of students felt difficult in double integral; 1,7% of students felt difficult in Fourier Series, Gamma and Beta function; Bessel-Legendre, Boundary values and variation of Calculus.

Conclusion :

1. The students' interest in mathematics in primary school is 56%; secondary school is 61,9% and high school is 67,8%, it means that the students' interest in mathematics increases as well as the level of education.
2. To overcome the difficulties in school studying mathematics :
   -in primary school 37,3% of students were guided by their parents;
   -in secondary school 33,3% of students joined a study group
   -in high school 39,8% of student joined a study group.
3. 22% of students explained that they felt difficult in the story problem in mathematics when they were in primary school;

17,8% of students explained that they felt difficult in solid geometry when they were in secondary school; 3,4% of students explained that they felt difficult in statistics when they were in high school and 4,2% of students explained that they felt difficult in double integral in the University degree.

4. There is a correlation between mathematics material in high school and University

Suggestion :

1. We hope there will be a computer package which will be interesting for students when they study mathematics in primary school up to high school.
2. This research should be done further to find out another variable which will support the technology, such as the relation of material with technology.

REFERENCE

Garis-garis Besar Program Pengajaran (GBPP) bidang study Matematika Departemen Pendidikan dan Kebudayaan, 1986.

The Tentative Ideas and Practice of Multi-media Instruction Software

Tang Rui-fen, East China Normal University.

I . Background and Origin

Respond to to the call of the Education Committee: "Turn the direction to the future, to the world and to the modernization" and "Popularizing the computer education should begin with the children". Since the middle of 1980s, computers entered into Chinese schools and increased rapidly. Many teachers were involved in the computer assisted instruction, the education reform was promoted by combining their rich experience with the powerful tool-computer. Originally, the software was designed for those simple machine, most of them was only used for drill and tutoring, can not changed flexibly, thus limited the advanced development of technology in mathematics education.

Until 1990s, the multi-media technique is developed rapidly, which provides favourable condition for making multi-media instruction software with visualizable and vivid representation. On the other hand, in some provinces and cities of China (especially some large cities in the coast), several schools were equiped with many Personal Computers, meanwhile,more and more families buy microcomputers which is for family use only. We would like to take advantage of this opportunity, and make a break through in making multi-media instruction software for family use. To stimulate the motivation for learning, to enhance the initiative of students, to encourage the exploration with conjecture, and to embody the mathematics application and modelling, and finally, we expect the education ideas would be changed, and "education for examination" would be turned into "education for quality" in the support of modern technology.

II. Suggestion and Attempt

1. Strengthen the realism of mathematics

Mathematics originate from real world, and closely related with everyday life. It is the best way to show something through presenting the real situation, which will explain problems clearly, and will promote further understanding of students. We select certain passages of some videotapes or films, and also put a series of pictures in our softwares; particularly, while the mathematical concept need to be introduced or some actual problems need to be solved, we fully utilize these lively scene (include some cartoon) to reflect the mathematical content.

For example, in geometry, the congruent figures are introduced by several pairs of real things-the hands, the gloves, the socks, and so on; then give the concept of congruent figures, which means that two figures can coincide completely if you put them together. Other example, in algebra, the word problems about linear equation with one unknown are discussed, we display a video of a factory, which shows the process that the flour is how produced from the wheat, then it is easy to get the equivalent relation as follows:

the weight of the flour + the weight of the wheat bran

= the weight of the wheat

in general,

the weight of the wheat × 85%

= the weight of the flour

If the question is : How many tons of wheat are needed when we need 15 tons of flour? As a result , to suppose the unknown, to establish the equation, and to solve the problem, those are self-evident.

Just like above treatment, we attempt thorough the multi-media software to arrive a goal, that is, in the course of teaching, learning and applying mathematics, the fact, theory and rules of mathematics have to integrate with the real situation. Consequently, students can get useful mathematical knowledge which conform to reality.

2. Urge students to become the main body of study

In the design of the multi-media software we stress the participation and control of students, we designed a varied menu divided into several units in different levels, which reflect the structure and connection of eah unit. The units are arranged in linear order, but independent each other. Students can learned on the basis of order, or selected different content according to themselves. Students not only participate in "teaching", but also control the learning process, and become the "master of this special classroom", to arrange the pace, order and content based on their own needs. According to the distinguishing feature of content and the distinct forms of treatment, the situation of students may be considered into two side, one is participating and the other is controlling:

(1) Students' participation have various stages and forms, which can divided into: simulating participation, controlling participation and responding participation.

Simulating participation simulates the communication between teacher and students in the real classroom, the role of student still keep in receiving. Controlling participation can change the learning process temporally or longer through some factors of control in accordance their special needs, this is a kind of participation that students can work as the master of the "classroom". For instance: they may use the function of dictionary to look for the theorems or formulae that are unfamiliar to them. And responding participation is the participation that students can select the problems to solve or answer the questions.

(2) The control of students can be classified in two patterns according to the position in the link and the nature: choosing control and corresponding control. There will be different ways to solve same problem from different respect, students choose one from these ways according to their own understanding, then the representation and interpretation of the contents will along this way, this is the choosing control. Figure 1 is the block diagram of this kind. Corresponding control adopts different ways of control according to the different reaction of students to same question. This kind of control always appears in the process of the explanation of examples and the handling of questions. Figure 2 is the block diagram of this kind.

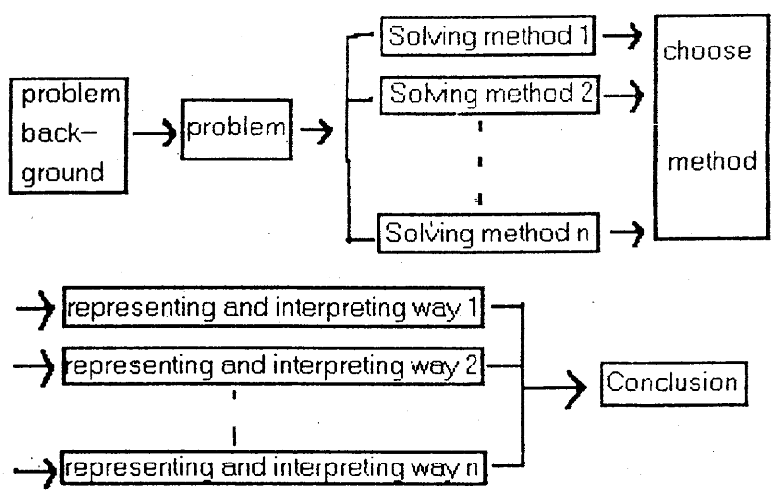


Figure 1 Choosing Control

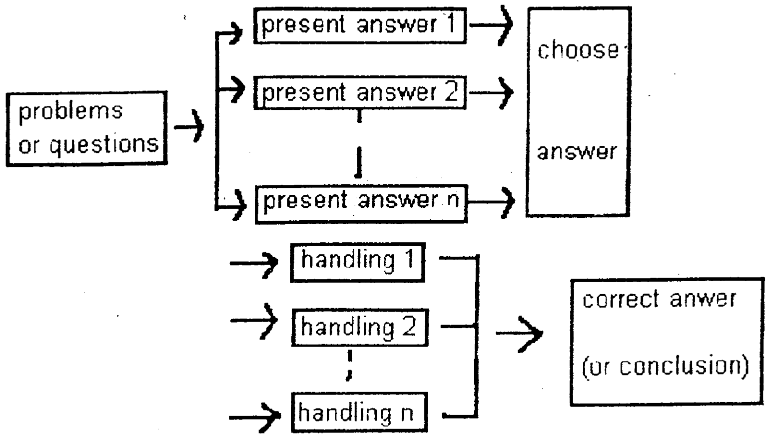
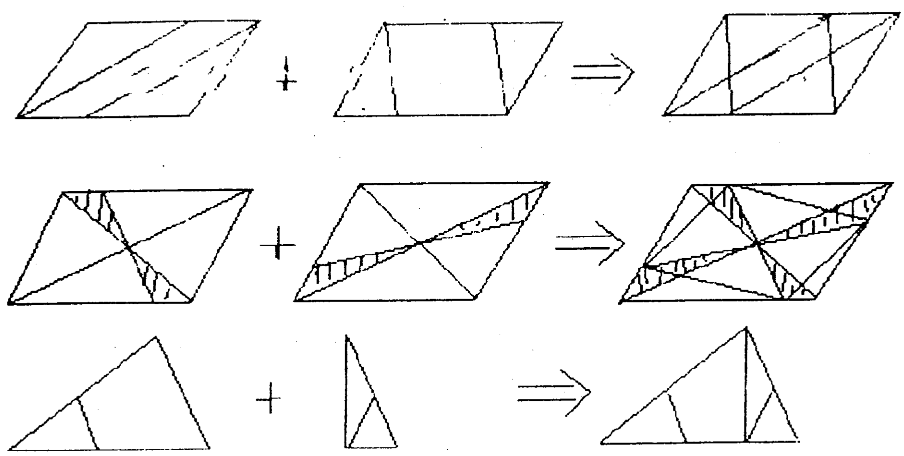


Figure 2 Corresponding Control

3. Represent mathematical process dynamically and desalinate formal reasoning.

Using the animated function of computer to represent some mathematical process dynamically, and using audio visual form that is easy to be accepted to express those abstract mathematics, in order to desalinate strict deductive reasoning or the abstract notion properly. Such as in plane geometry, the combination of elementary figures:



Another example: During the course of discussing the volume of triangular pyramid, the software demonstrate the animated procedure which shows that a triangular prism can be separated into three equal-volume triangular pyramid. Then it is easy for students to get the formula:

$V_{\text{triangular pyramid}}=(1/3)\ V_{\text{triangular prism}}=(1/3)Sh$

When studying the relationship of function graphs, we use animation to demonstrate dynamically the process of how to obtain a function graph through the translation, rotation or other transformations of elementary function graphs.

In the learning of the conception of limit, we use the limit of mumber sequence {1/n}, combining with number axis, to represent the conception of limit by dividing into the following steps.

(i) To observe the trend of {1/n} as n increasing.

Let students input different value of n through keyboard, then flash the point 1/n on the number axis according to the value of n. Thus the trend that the points corresponding to 1/n more and more close to the origin is represented visually.

(ii) Which range is n in, such that |1/n-0|<0.01?

Let students input the value of n which is increasing. Then they can see that all the points on the number axis corresponding to 1/n fall in the interval (-0.01,0.01), after the value of n increases to a certain number (n>100).

(iii) To get any smaller value (e=0.0001), students can also find a value N of n. When the value of n is bigger than N, all the 1/n satisfy |1/n-1|<0.0001. Like this, students will have a visual comprehension about the conception of limit.

4. Build the mathematics laboratory

Mathematics laboratory can provide an arena in which students can explore and discover mathematical rules, verify and prove mathematical conclusions, use the rules and conclusions practically.

It is not random to set up mathematics laboratory. First, select suitable content, then, elaborately design materials and rules of experiment. According to the aim of the experiment, the materials may be the concrete things in reality, or abstract mathematical elements (e.g. point, line, plane, and angle. etc.) or elementary theorems, operation laws and so on. However, the rules of experiment are some stipulation of manipulation.

Based on the diffrent aims when they are set up, mathematics laboratories can be classified into: the laboratory of exploration and discovery, of verification and proving and of problem solving. The laboratory of explorationor discovery is to let students observe and analyse their manipulation results, then find the conclusion. For example, the experiment of "the relationship of three sides in triangle ", the material is "triangle", the rule is "to straighten up two sides of the triangle, put them at the same line, then compare with the third side", and students will discover the conclusion: "the sum of two arbitrary sides is bigger than the third one in triangle". The laboratory of vevification or proving is to verify the conclusion through students operation. For example, the experiment of "in equicrural triangle the three line segments coincide". The laboratory of problem solving is the laboratory in which they solve problems using the propositon, the theorem, or the law what they have learned.

5. Pay attention to the interest.

Construct a happy and relaxed atmosphere for students utilizing the function of multi-media technique sufficiently, in order to stimulate their interest, arouse their curiosity about mathematics. Such as: set up the background music, evaluate and encourage to the answer in time; design mathematical games craftily, set up date store for collecting abundant relative materials (e.g. the history of mathematics, the anecdotes etc.), thus, students can choose those interesting content to further their learning.

III. Reflection and Prospect

1. Existing Problem

Because of following reasons, there are some difference between the developed software and the original thinking.

First, making use of multi-media technique to produce instruction software is at a fumbling stage in our country, there are not so many methods and skills to use. Especially, it is poor in theory, so the quality of the software is influenced to a certain extent.

Secondly, due to the shortage of realistic mathematics materials, the research about teaching process and the psychology of learning are not enough, so the design and display of the software are not satisfied.

Thirdly, it is not so concerted between the designer and the maker of the software, some technical problem weakened the features of original thinking.

2. Prospect

Although there are some questions, it still have a bright future to combine the multi-media technique with the mathematics education.

(1) The modern ideas of mathematics education can be permeated through multi-media software.

There arevarious models of numbers and figures in the real world, which provide realistic context for mathematics. To sort out and classify these materials using multi-media technique, we can set up a rich mathematical data store. It can give the situation of problem solving, and engage in meaningful learning and exploring teaching. It is very useful for students to learn mathematization and to applying mathematics knowledge to reality.

Moreover, as a means of individual education, it cooperates with school education, provides environment and data that are complementary with the learning in classroom. For example, we can design some topic softwares, such as history of mathematcs, stories of mathematicians, and following the footprints of mathematicians etc.

(2) Using the multi-media instruction software to develop the students' ability of creative thinking.

The course itself which students explore mathematics individually with the aid of multi-media environment, is the course training the ability of creative thinking. Such as, providing some mathematical materials to let students conjecture and explore in learning. Verifying its correctness by practising and impel to form and develop the mathematics thinking in the process of constantly attempt and improvement.

(3) Set up expert-system of mathematics teaching.

The expert-system of mathematics teaching, is set up by utilizing the achievement of artificial-intelligence. It has the following funtions:

a, the guide system of individual learning.

b, the identify and diagnose system of students' learnig level.

c, the system of explanation.

d, the laboratory of the knowlege growing.

e, the computer automatically organize mathematical material for learning.

f, the checking system of the applying ability etc.

In one word, the multi-media instruction software provides a valuable way of improving mathematical teaching and learning. It has a wide prospect in mathematics education.

East China Normal University
Tang Rui-fen

TEACHING ϵ - δ LIMIT OF FUNCTIONS OF TWO VARIABLES

Tran van Thuong
Missouri Southern State College
Joplin, Missouri 64801

In this paper, we will present two different approaches of teaching the ϵ - δ limit of functions of two variables in calculus 3 during two different semesters, then we will examine the feedback of each method.

We spent one hour to teach students computing the limit and showing the nonexistence of the limit, two hours to teach them understanding the concept of the ϵ - δ limit and writing the proof. We had 22 students in semester 1 and 28 students in semester 2; therefore the class size of two semesters is not quite different.

Our typical presentation was followed by two basic principles: formal definitions and procedures will be given from the investigation of practical problem, and every topic should be presented geometrically, numerically, and algebraically.

First we consider a real valued function f(x,y) and a point A(a,b,c) on its graph, we cut the graph by two parallel planes to the xy-plane by $z_1 = c + \epsilon$ and $z_2 = c - \epsilon$, to obtain a graph G between two planes. Assume that G has no hole (fig.1). We asked three questions:

1) Can we find a δ - square in xy-plane, center at point B(a,b,0) so that point (x,y,f(x,y)) is on the graph G whenever point (x,y) stays inside the δ - square?
2) What is the maximum value of δ that we may have?
3) Does δ depend on ϵ ?

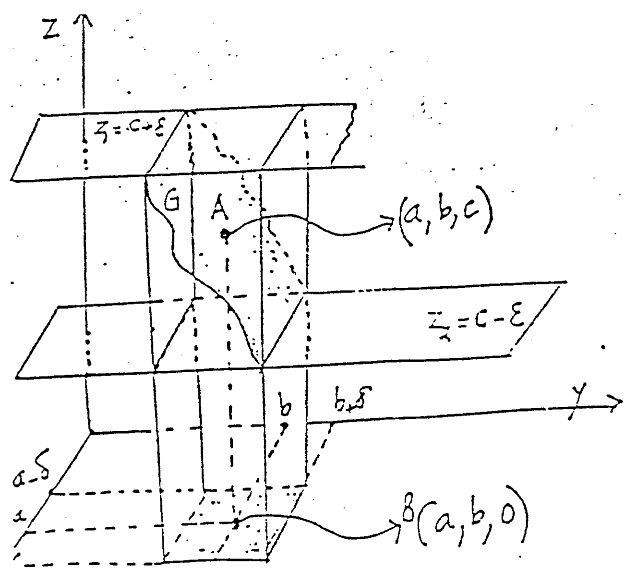
If (1) can be done then we say that $\lim_{(x,y)\to(a,b)} f(x,y) = c$

After a geometric investigation, a formal definition was given as follows:

<u>Formal</u> <u>definition</u>: [1] $\lim_{(x,y)\to(a,b)} f(x,y) = c$ means that f is defined for values of x and y arbitrarily close to a and b respectively, and if ϵ is a positive number, then there is a positive number δ such that $|f(x,y)-c| < \epsilon$ whenever (x,y) is in the domain of f, $|x-a| < \delta$ and $|y-b| < \delta$ and $|y-b|$ and $|x-a|$ are not both 0.

In fact, figure 1 explains the geometric meaning of the definition.

Figure 1: Geometric Definition of ϵ - δ Limit of Functions of Two Variables.



Two examples were given, then a few exercises were assigned to make sure students understand those examples, and to make them explore more different situations.

EXAMPLE 1: Prove that $\lim_{(x,y)\to(1,1)} f(x,y) = 3$ where $f(x,y) = x + 2y$.

Algebraic and numerical works:

$|x+2y-3| = |x-1+2(y-1)| \le |x-1|+2|y-1| < \delta+2\delta=3\delta<\epsilon$. So, *Choose* $\delta = 1/3\epsilon$.

PROOF: Given $\epsilon>0$, let $\delta = 1/3\epsilon$, then $|x+2y-3| < \epsilon$, whenever $0<|x-1| < \delta$ and $0< |y-1| <\delta$.

The difference of two approaches is in semester 1, we did not cover the geometric proof which impressed many students in semester 2.

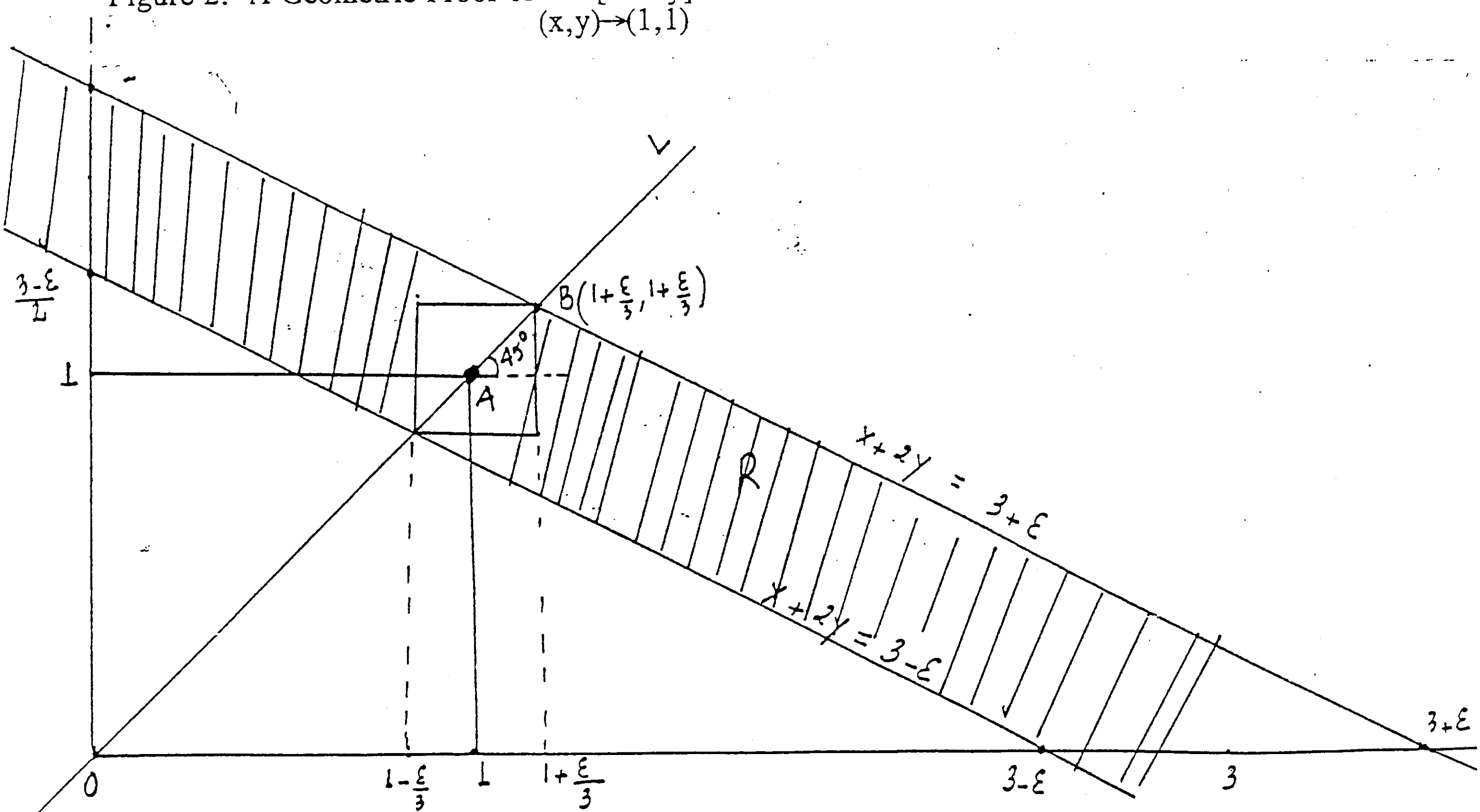
A GEOMETRIC PROOF (fig. 2). (only in semester 2). $|x+2y-3| < \epsilon$ means that $-\epsilon<x+2y-3 < \epsilon$ or $3-\epsilon < x+2y < 3+\epsilon$ (*). Now we follow the following steps:

(1) Find the region R for (*).
(2) Find a maximum square inside R with center at A(1,1), two sides are parallel to the x-axis and the y-axis respectively by:
  (a) Drawing a line L passing through A(1,1) with slope m=1.
  (b) Writing the equation of L: y-1 = 1(x-1) or y=x.
  (c) Finding the intersection B $(1+1/3\epsilon, 1+1/3\epsilon)$ by solving the system:

$$\begin{aligned} x+2y &= 3+\epsilon \\ y &= x \end{aligned}$$

  (d) By a geometric observation, $\delta = 1/3\epsilon$
(3) In fact, the maximum value of $\delta = 1/3\epsilon$

Figure 2: A Geometric Proof of $\lim_{(x,y)\to(1,1)} [x+2y] = 3$.



EXAMPLE 2: Prove that $\lim_{(x,y)\to(1,1)} [x^2 + y^2] = 2$

Algebraic and numerical works:

$|x^2 + y^2 - 2| = |x^2 - 1 + y^2 - 1| \le |x^2 - 1| + |y^2 - 1| = |x-1|\,|x+1| + |y-1|\,|y+1| <$
$\delta\,|x+1| + \delta\,|y+1| < 3\delta + 3\delta$ (**) $= 6\delta = \epsilon$. *So choose* $\delta = 1/6\epsilon$.

(**) We may assume that $\delta < 1$, then $|x - 1| < 1$, meaning that $-1 < x-1 < 1$ or $1 < x+1 < 3$. It follows that $|x+1| < 3$; similarly, we have $|y+1| < 3$.

PROOF: Given $\epsilon > 0$, let $\delta = \min\{1, 1/6\epsilon\}$, then $|x^2+y^2-2| < \epsilon$ whenever $0 < |x-1| < \delta$ and $0 < |y-1| < \delta$.

A GEOMETRIC PROOF (fig. 3). (Only in semester 2). $|x^2+y^2-2| < \epsilon$ means that $-\epsilon < x^2+y^2-2 < \epsilon$ or $2-\epsilon < x^2+y^2 < 2+\epsilon$ (***). Now we follow the following steps:

(1) Find the region R for (***).
(2) Find a maximum square inside R with center at A(1,1), two sides are parallel to the x-axis and the y-axis respectively by:
(a) Drawing a line L passing through A(1,1) with slope m=1.

(b) Writing the equation of L: $y-1 = 1(x-1)$ or $y=x$

(c) Finding the intersection point $B(\sqrt{1+1/2\epsilon}, \sqrt{1+1/2\epsilon})$ of the line L and the circle

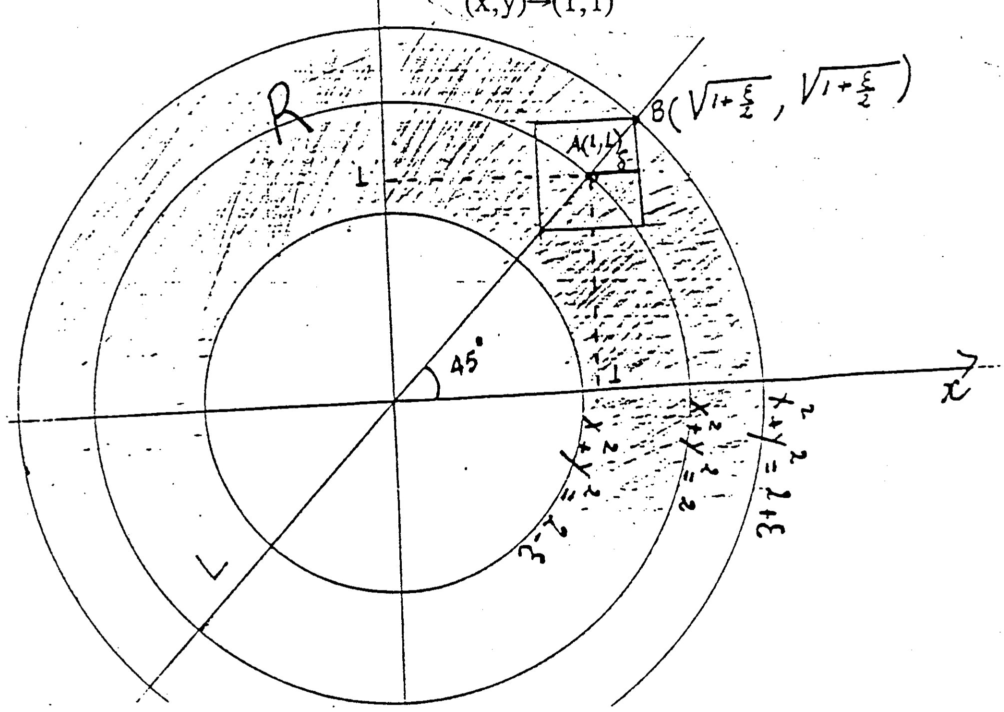
$x^2+y^2 = 2 + \epsilon$ by solving the system:

$$x^2+y^2 = 2 + \epsilon$$
$$y = x$$

(d) By a geometric observation, we see that $\delta = \sqrt{1+1/2\epsilon} - 1$

(e) The maximum value of δ is $\sqrt{1+1/2\epsilon} - 1$

Figure 3: A Geometric Proof of $\lim_{(x,y)\to(1,1)} [x^2 + y^2] = 2$



HOMEWORK: *1) Show that* $1/6\epsilon < \sqrt{1+1/2\epsilon} - 1$ for $\epsilon < 6$.

2) Using Algebraic and numerical method to show that the maximum value of δ *is* $\sqrt{1+1/2\epsilon} - 1$.

3) Prove that $\lim_{(x,y)\to(1,2)} [3x + 2y] = 7$.

Compute $f(1.1, 2.1)$, $f(1.001, 2.0001)$, $f(1.00002, 2.000001)$. $f(0.9999, 1.99999)$.

Compare all those values with 7.

4) Prove that $\lim_{(x,y)\to(1,2)} [x^2+y^2] = 5$

5) (Only in Semester 2) The disadvantage of the geometric method in general, it is very difficult to graph many other functions. In this case we will try to approximate by another function which is easy to be plotted.

Prove that $$\lim_{(x,y)\to(0,0)} \left[\frac{xy(y^2-x^2)}{x^2+y^2}\right] = 0$$

We asked students 3 questions in semester 1 and 4 questions in semester 2, and here is the feedback:

1) Do you understand fairly well the concept of ϵ - δ limit?

	Class 1	Class 2
Yes	10 *(45%)*	22 *(79%)*
No	12 *(55%)*	6 *(21%)*

2) Do your understand very well the concept of ϵ - δ limit ?

	Class 1	Class 2
Yes	4 *(18%)*	14 *(50%)*
No	18 *(82%)*	14 *(50%)*

3) Do you need more discussion times? YES: all.

4) Does the geometric approach help you to understand the concept of limit? (only for Class 2)

	Class 2
Yes	22 *(79%)*
No	6 *(21%)*

Based on the results of two semesters, we may conclude that the geometric approach helps students in understanding the concept of ϵ - δ limit, we may also ask three open questions:

1) Do we need to spend more time for class discussion?
2) Do we need to use computer to teach? If yes, then how can we use it effectively?
3) Do we have another more effective approach to teaching the concept of ϵ - δ limit?

REFERENCES

1. D. F. Riddle, "Calculus and Analytic Geometry", Fourth Edition, Wadsworth Publishing Company, Belmont, California 94002.

MATHEMATICS SOFTWARE IN SERVICE TEACHING

Cristina Varsavsky
Mathematics Department, Monash University
900 Dandenong Road - Caulfield East VIC 3145 - Australia
e-mail: c.varsavsky@monash.edu.au

Abstract

The Department of Mathematics at Monash University provides service teaching of mathematics subjects to the faculties of Engineering, Computing, and Arts. Many of these subjects incorporate a computing component to familiarise students with the current technologies available to the relevant areas, including computer algebra programs, spreadsheets, and statistical packages. The computer program DERIVE plays since 1990 a particularly significant role in the engineering and computing subjects. Through the Mathematics Learning Centre, the Department also provides access to computerised material for self-paced learning. This paper describes the aims for the introduction of mathematical software in service subjects, the different ways it has been implemented, and the positive and negative aspects related to it.

1. Introduction

Mathematics is no longer only a science or engineering subject; it is becoming more and more important in other disciplines like computing, business, social sciences, and health. A broader range of students need to master mathematical fundamentals to serve their professions in the present technologically based society. Students must be trained to apply the mathematical concepts to their relevant disciplines and make use of the up-to-date technology to illustrate, analyse, and solve the problems they will have to tackle in their courses and their professional lives.

The Mathematics Department of Monash University, Caulfield Campus, provides service teaching to several non-science faculties. The provision of these services is mainly done through the teaching and management of the subjects students undertake as part of their degrees and also through the Mathematics Learning Centre which students use as a support service (see [V, 1995]).

The faculties serviced by the Mathematics Department are:

- Engineering: first, second, and third year mathematics for the Bachelor of Engineering (Applied) and Bachelor of Technology are compulsory subjects.

- Computing: mathematics subjects are not compulsory for students undertaking the Bachelor of Computing degree; nevertheless a significant proportion of these students take a stream of mathematics subjects to complete a minor or a major.
- Business: although the Mathematics Department is not responsible for the structured delivery of mathematics for business courses, it directly caters for these students through the support services provided by the Mathematics Learning Centre.
- Arts: a significant number of arts students choose to do statistics as a minor or major for their arts degrees.

It is usually the responsibility of the Mathematics Department to design and implement these subjects, and therefore to decide how to make use of the available technology in their service subjects.

2. Mathematics Software

The availability of sophisticated computer software has opened up a great range of possibilities in the teaching and learning of mathematics. A description of how they have been incorporated in the various mathematics service courses at Monash (Caulfield) follows.

2.1 Computer Algebra Systems

Computer algebra packages are certainly the ones who had a greater impact in the mathematics curricula. They can perform symbolic and algebraic computations, solve equations, find derivatives and integrals, plot graphs, and many other tasks in a very short time. Mathematician's find themselves in the dilemma whether it still makes sense to teach students what a computer would do at the touch of a button. If used properly, these packages can enhance students' intelectual activities, helping them to become active learners.

Two are the computer algebra packages used at Monash (Caulfield):

- DERIVE has been incorporated to the relevant Monash (Caulfield) courses six years ago. The main reasons for opting for this particular package were its user friendliness, the low cost and the rather modest hardware requirements to run it. Students have access to them through the university network; they can also acquire their own copy for an affordable price.

  In first year subjects, DERIVE has not yet substantially changed the way mathematics is being presented to students but it is rather supplementing the traditional teaching. In first year mathematical methods for engineering, technology and computing, students are expected to learn how to use DERIVE software in their own time. Students needing extra help during this process always have the additional support provided by the Mathematics Learning Centre. An introductory DERIVE Tutorial (see [V, 1995]), in which the different commands are illustrated through examples, was written for this purpose. This tutorial has been now used for four years, and has proven to be successful in giving students a sound introduction to be able to tackle their DERIVE assignments on their own. Mathematical methods topics are still being delivered in the traditional lecture-tutorial setting. DERIVE supplements these lectures through assignments which students complete without supervision. The aim of these assignments is twofold: firstly, for students to become acquainted with the software, to learn about their capabilities and limitations, so they can use it in the rest of their discipline work; secondly, for students to take an exploratory approach to solve mathematical problems, many of which would not have been possible to attempt without this powerful graphical and computational tool. In these assignments some emphasis is placed on the interpretation of the answer output on the screen. These assignments usually count from 10 to 20 percent of their final mark.

In second year engineering, DERIVE is mainly used in the Fourier module. Tutorials are held in a computer lab where students work through their Fourier series and transforms problems. The role of DERIVE here is mainly to eliminate the drudgery of integral calculations. In addition, the plotting power of the package is very useful to illustrate the convergence of a Fourier Series. The examination of these topics is also held in a computer lab.

There is evidence that students use DERIVE to solve problems for other subjects in which this package is not a formal part of it. This is particularly noticeable with engineering students, they don't only use what they learnt during their mathematical methods units but also seek to extend their knowledge in the usage of the software.

- MAPLE is at moment used only in one subject for computing students, namely Symbolic Computation. This is a second year elective which aims to expose students to its capabilities and limitations, and to use a computer algebra package to solve mathematical problems. All classes are held in a computer lab in which students complete weekly assignments; the final examination is also held in a computer lab. The subject places a strong emphasis on the different ways a mathematical problem can be solved with MAPLE as well as on the interpretation and verification of the results given on the screen. Half of the subject is dedicated to programming. This subject has proven to be highly successful and highly demanded by computing students.

  As up until 1994 MAPLE was available to students only in a VMS environment, lecturers felt this package was not user friendly enough for engineering students. With the WINDOWS version now available at university labs, its introduction for second year engineering and technology students is being considered, but it is still intended to use DERIVE for first year students.

2.2 Spreadsheets

Spreadsheets are a very useful tool to analyse iterative mathematical problems. They had an important impact on mathematics for business as students can now work with real business data to apply their newly acquired knowledge, making their learning more meaningful.

Business students have a weekly session (one quarter of their contact time) in a computer lab in which they analyse and solve financial and statistical problems. LOTUS 123 was used until 1994; at present spreadsheet work is being done with EXCEL in a WINDOWS environment.

Spreadsheets are also being preferred to programming languages in numerical methods units which are offered to engineering and computing students.

2.3 Other

Statistics packages have significantly changed the way statistics subjects for engineering, computing and arts students are being taught. Less time is being spent on calculating the statistical parameters and more emphasis is being placed on the analysis of real data and interpretation of the computer output. The packages range from the simple program MICROSTAT, through the WINDOWS based MINITAB and SPSS.

Various other public domain mathematical software are available to students through the university network. Although they are not incorporated to the relevant subjects, students are encouraged to use them, especially through the Mathematic Learning Centre. The most used programs are VENN, CAPGRAPH, GNUPLOT, MATRIX, and others.

3. CBL and CBT

Although strictly speaking they are not mathematical software, the accessibility to Computer Based Learning (CBL) and Computer Based Testing (CBT) also had an impact on the learning of mathematics. CBT and CBL are being developed through the Mathematics Learning Centre, using the authoring package AUTHOR. Two are the major areas which have been covered to the present:

- A self paced sequence of statistics modules (see [A, 1992]). They were designed to supplement lecture material or to be used as revision for the final examination. These are interactive modules: students are presented with a question and if they get it wrong, the program leads them through the different stages to the solution, giving them another question of a similar type at the end. If a student gets a wrong answer the second time, she/he is referred to read more about the topic in a booklet which supplements the CBL.
- A sequence of weekly assignments for technology students. They consist of multiple choice and short answer questions. These were introduced to encourage students to work throughout the semester rather than concentrating their work only around the exams time. This is the first year they are being used, and although there is no documented evidence yet, it appears to be that the aims are being fulfilled.

4. Problems

Lecturers are being continuously challenged to design mathematics service courses which incorporate the currently available technology, making them meaningful to the relevant disciplines. Their learning took place in a completely different environment making it very difficult for them to decide what is the best way to prepare students for the future. Therefore, results do not always turn out to be as expected. This is particularly true for computer algebra packages. Students fall, very easily, in the naive approach of believing everything the computer tells them. More emphasis must be put on the interpretation and verification of the answer provided by the software.

In mathematical methods subjects too much time is still being spent on tasks which mainly consist in imitating the computer. The main focus should be placed on the concepts, and then use those concepts to model the problem, leaving for the computer the drudgery of lengthy calculations. It is certainly much more challenging to teach interpretation of a computer output than to teach a computational algorithm. At the moment the available computer resources would not allow to have a good proportion of the classes held in a computer lab.

In a recent survey (see [V, 1994]) to the engineering faculty it has been made evident that lecturers in the non-mathematical subjects do not make use of the computing skills students acquire in their mathematics subjects. Inter-faculty collaboration must be improved to design comprehensive and coherent courses.

Many, if not all of the above listed problems, are due to budget restrictions. Not only computing resources need to be expanded but also educators should have more time available for training and for development and evaluation of new teaching material.

5. Conclusion

Computers are already part of our everyday life and they are here to stay. Their introduction created a new technological environment in which new skills are needed and some needed a while ago are now obsolete. Universities catering for students in the area of engineering, computing, health, business, social sciences, etc. have the responsibility to give them the best realistically achievable preparation for the technological world in which they will serve their profession. Therefore more emphasis must be placed on the ongoing review of the relevant curricula, continually assessing the needs of current students and the different disciplines, keeping up with the new technological tools, and investigating the potential of those tools in the enhancement of the teaching-learning processes.

6. References

[Adlem, 1992] Adlem, R G W Computer aided review in introductory statistics Proceedings ASCILITE 92 , pp 12-22, 1992.

[Varsavsky, Norton, 1995] Varsavsky, C and Norton, P, Building up the self-confidence of mathematics students in non-science courses, *Int J Math Educ Sci Technol,* to be published.

[Varsavsky, 1994] Varsavsky, C, Mathematics in engineering, *Proceedings of Contemporary Approaches to Research in Mathematics and Science Education,* 1994.

[Varsavsky, 1995] Varsavsky C, *DERIVE Tutorial*, Monash Mathematics Learning Centre, 1995.

TEACHING SIMULATION, VISUALIZATION, AND MODELING USING *MATHEMATICA*

Paul C. Abbott
Department of Physics, University of Western Australia, WA 6907, AUSTRALIA
`paul@physics.uwa.edu.au`

Abstract

This paper describes an introductory course on Simulation, Visualization, and Modeling given at the University of Western Australia since 1992. The course level is introductory and directed at third year physics majors, though a significant number of honours and post-graduate students from physics, engineering, and mathematics have taken the course.

The course objectives are: (1) to use computers as an aid to understanding real physical systems; and (2) learn efficient methods for the analysis of these systems. Course delivery is via nine three-hour laboratory sessions using *Mathematica* to provide the presentation environment and computational framework for the course. The course has been well received by the students and their achievement and understanding is generally of a high standard. To demonstrate the flavor of the course and the level of student achievement, selected student exam solutions are presented.

Introduction

Conventionally, simulation and modelling are taught by stressing numerical techniques. Students are introduced to a programming languages such as fortran, C, pascal, or basic and expected to learn, by example, enough programming to study a number of physical problems (see, *e.g.*, [Koonin and Meredith] and [De Jong]). An alternative approach – the one taken here – is to use the computer algebra system *Mathematica*, as the computational medium.

Mathematica has extensive numerical capabilities. One particular strength is arbitrary precision numerics, which can be useful for examining numerical methods. See, for example, the recent text by [Skeel and Keiper]. Another advantage is that the majority of the special functions of science and engineering are immediately available.

The symbolic capabilities of *Mathematica* assist the analytical study of techniques such as series expansion, Fourier analysis, differential equations, and variation of parameters so that the student is free to explore such topics in more detail.

Mathematica also has excellent graphics, including animation, which enables powerful visualization without requiring additional tools or laborious programming. *Mathematica* documents – called Notebooks – are portable and combine text, commands, graphics, and sound in a single medium. Taken together, this provides an excellent teaching and learning environment as well as a good platform for developing courseware.

Exam

To demonstrate the flavor of the course and the level of student achievement, selected student exam solutions (edited only for length) are presented below. Bonus marks are given for elegant or unconventional solutions. The students are given two weeks to complete the exam.

In these *Mathematica* Notebooks, *Mathematica* input (output) is indicated using **`Bold`** (`Plain`) Courier font and graphics are directly generated in place.

Chaos Game

The Chaos Game is an example of an iterated function system. Define three vertices of an equilateral triangle in the plane and an arbitrary starting point both located within [0,1] x [0,1]. At each step, select one of the vertices at random and define the next point in the iterative sequence by taking the midpoint between these two points. Iterating this operation (you might want to use **`NestList`**) to obtain a sequence of points. Visualize this sequence using graphics and animation. Describe what you see. Does any structure appear? Are you surprised by this considering the randomness involved?

Solution by David Saxey

Firstly, define the vertices of the equilateral triangle:

```
tripoints = {{0,0},{1,0},{0.5,N[Sqrt[3/4]]}};
```

The numerical value of the square root is used to speed up the calculations. (*Mathematica* doesn't have to evaluate exact expressions.)

Then, in preparation for the use of **`NestList`**, define the function **`trans`** which takes a point and calculates a new point at a position half way to one of the three vertices chosen at random.

```
trans[{x_,y_}] := 1/2 (tripoints[[Random[Integer,{1,3}]]]+{x,y})
```

Seed the random number generator. (an attempt to ensure that the 'random' numbers are as random as possible):

```
SeedRandom[];
```

Select an arbitrary starting point within the square {0,0} to {1,1}.

```
{x,y} = {Random[],Random[]};
```

Apply the **`trans`** function to the starting point five thousand times, thus generating five thousand points to form the structure.

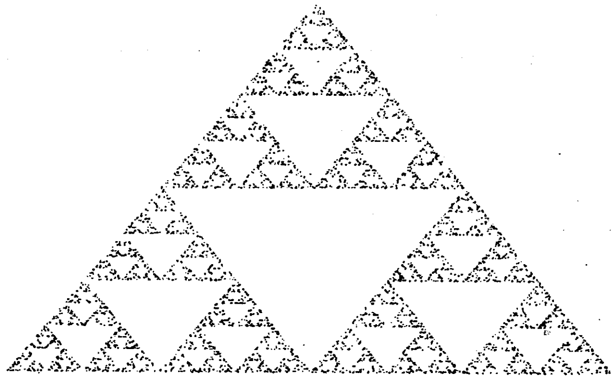
```
NestList[trans,{x,y},5000];
```

Convert the data into graphics points:

```
Map[Point, %];
```

Display the structure:

```
Show[Graphics[{PointSize[0.0005],%}]];
```



The structure that appears is very intricate. However it appears that it is fractal-like which is understandable considering the nature of the iterating function. For example taking the midpoint between any two vertices gives a point which may be considered as one of the vertices of one of the secondary triangles, in fact any application of **trans** will give a point that may be thought of as a vertex of a sub-triangle. In this way the self-similarity of the structure is easily understood.

Air Resistance

Find the maximum range and the angle of launch of a golfball which has a drag coefficient of 0.5 and an initial velocity of 75 ms^{-1}. Compare this to the case when there is no air friction. Find the maximum range and the associated launch angle. Plot the energy loss as a function of time for a trajectory of 30 Degrees. The drag force is given by

$$F_D = kv^2, \quad k = \frac{1}{2} C\rho A,$$

and the equation of motion is

$$m\mathbf{a} = m\mathbf{g} - \hat{\mathbf{v}} k \mathbf{v} \bullet \mathbf{v}.$$

Solution by Matthew Hollingworth

Introduction

To estimate **rho** I used the fact that the average molecular mass of air is 30 a.m.u.

```
rho = 6.022 10^23 30 1.661 10^-27 / (22.4*(0.1)^3)
1.33963
```

For a sphere the size of a golfball (about 20 mm radius), and mass

```
m = 0.1;
```

I will take surface area to be the cross-sectional surface area

```
A = N[Pi (0.02)^2]
0.00125664
```

Now compute **k**

```
c = 1/2;
```

```
k = 1/2 c rho A
```

```
0.000420856
```

Define the earth's surface gravitational acceleration and call it **g**. (We are using a positive-up co-ordinate system, so g is a negative constant).

```
g = - 9.8;
```

Define the velocity:

```
v[t_] = Sqrt[x'[t]^2 + y'[t]^2];
```

General Trajectory Plot

I defined a general function which takes **theta**, **v[0]**, and **x** and **y** plotting ranges as parameters. The function executes a list of commands, and returns the results of the last one (i.e. the plot) as the result. **tf** is an approximation of the total time of flight, taken from the ideal case of no friction. In reality, the friction will reduce the time of flight, so this estimate will do (the equation if tf = - 2 y'[0] /g).

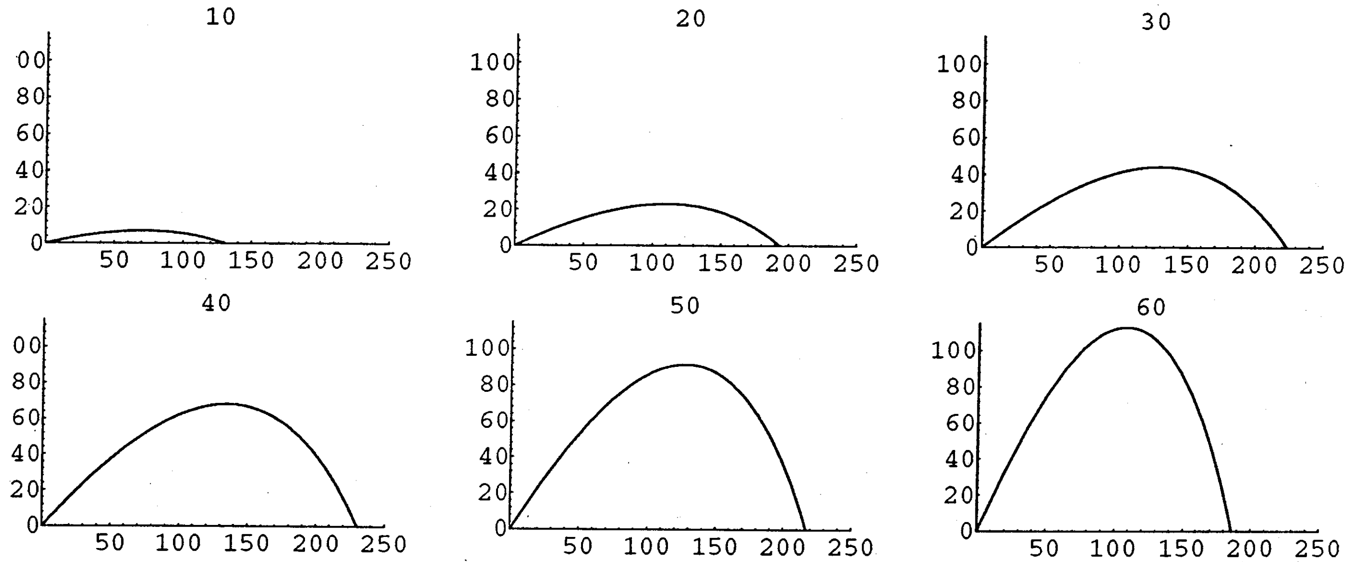
NDSolve is used to solve the equations, and the results are assigned to **x[t]** and **Y[t]**, and plotted, with the range given by the supplied parameters.

```
degree = N[Degree];

Trajectory[theta_,Vo_,Xlo_,Xhi_,Y_] :=

Last[{

     Clear[x,y];

     eqx = x''[t] ==   -k/m x'[t] v[t];
     eqy = y''[t] == g -k/m y'[t] v[t];

     eqx0 = x[0] == 0;
     eqy0 = y[0] == 0;

     eqvx0 = x'[0] == Vo Cos[theta*degree];
     eqvy0 = y'[0] == Vo Sin[theta*degree];

     tf = -2 Vo Sin[theta*degree] / g;

     s = Flatten[NDSolve[
          {eqx,eqx0,eqvx0,eqy,eqy0,eqvy0},{x,y},{t,0,tf}]];

     {x[t_],y[t_]} = ({x[t],y[t]} /. s);

     ParametricPlot[{x[t],y[t]}, {t, 0, tf},
          PlotRange->{{Xlo,Xhi},{0,Y}},
          PlotLabel->theta]
}]
```

Plots for various angles

The Trajectory function defined above is used to plot the flight paths for a range of theta values, in 10° increments.

```
Table[Trajectory[n,75,0,250,115],{n,10,60,10}];

Show[GraphicsArray[Partition[%,3]]];
```



Maximum Range

The problem could I suppose be solved by analytic methods but I decided to just refine the solution graphically. From the above table of plots, the approximate angle of greatest range was around about 40°. A couple more refinements (not shown here), showed that theta = 38.2 is the angle of maximum range (to 1 d.p.).

The graph in which the maximum range occurred was found, and the range for this angle is about 230.41 metres. This is about as much accuracy as could really be justified, given other real effects not accounted for (e.g. wind, inaccuracy of the friction model, etc).

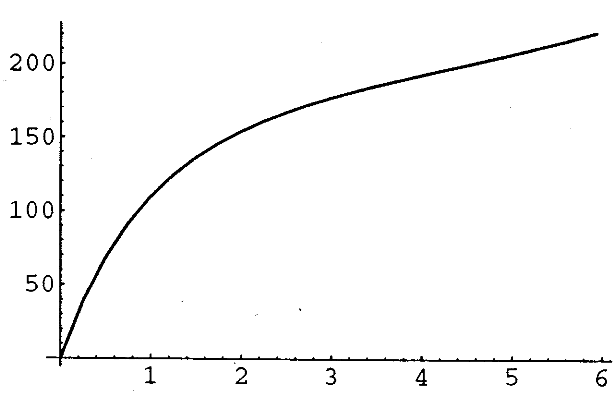
Energy Loss

The frictional energy loss as a function of time is given by the formula below:

```
Eloss[t_]:=0.5 m(75^2 - x'[t]^2 - y'[t]^2) + m g y[t]
```

This is derivable by subtracting the **total** energy at any instant from the initial kinetic energy(initial potential energy being zero). The total energy at time t is given by $^1/_2$ m v[t]^2 - m g y[t]. Subtracting this from the initial energy, $^1/_2$ m v[0]^2, gives the above formula.

```
Plot[Eloss[t],{t,0,6},PlotRange->All];
```



Newton's Method

Newton's method for finding the root of an equation, $f(x)=0$, relies on iterating the sequence:

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}$$

from a suitable starting value $\mathbf{x}_0$.

Use Newton's method to find the square root of 3 to 30 decimal places. The equation $z^3 = 1$ has three roots (2 are **complex** values). Investigate the number of iterations required to achieve convergence for the equation $z^3 = 1$ for **complex** values of z. Produce a plot of the result of iterating Newton's method in the complex plane starting with an arbitrary complex value `z = x + I y` for **x** and **y** values in the range [-1.2, 1.2] x [-1.2, 1.2]. Iterate initial points until it is clear to which of the three cubic roots the result is closest (its attractor). Identify the initial point according to its destination. Visualize your results.

Solution by Matthew Hollingworth

Real-Variable Case

Define a function **next**, which takes a point x_n and returns x_{n+1}, to accuracy specified by **acc**:

```
next[x_,acc_] := N[x-f[x]/f'[x],acc+1]
```

The **root** function takes an initial estimate and an accuracy, and uses recursion to refine the solution to the given accuracy. The terminating case is when the new and old values differ by less than the specified accuracy. Since our function is well behaved, we don't need a maximum recursion limit. To avoid calculating the **next** function twice, we use a temporary variable.

```
root[x_,acc_] := (temp = next[x,acc];
                  temp /; Abs[temp - x] < 10^-(acc+1))

root[x_,acc_] := root[next[x,acc],acc]

f[x_] := x^2 - 3
```

Getting the root to thirty d.p, and squaring to check that it is correct:

```
root[2,30]

1.73205080756887729352744634150 6

%^2

3.
```

Complex Roots

We need to define a function which classifies a complex number as to which root it represents. The following is enough for this case (alternatively, the **Arg** of the number might be taken).

```
class[x_] := 3 /; Re[x]>0
class[x_] := 4 /; Re[x]<=0 && Im[x]>0
class[x_] := 5 /; Re[x]<=0 && Im[x]<=0
```

I decided to use a compiled function instead to increase the speed.

```
Newton =
  Compile[{{z, _Complex},{acc,_Real}, {m, _Integer}},
    Module[{w = z, n = 0},
        While[Abs[w^3-1] > 10^-(acc+1) && n < m,
            w = 2w/3 + 1/(3 w^2);
            n++
        ];
        w
    ]
  ];
```

It takes any point on the complex plane and iterates it according to Newton's method. The iteration terminates when either the maximum allowed number of iterations has been reached, or when the cube of the number differs from one by less than a specified accuracy (i.e. the convergence condition). Checking that it works:

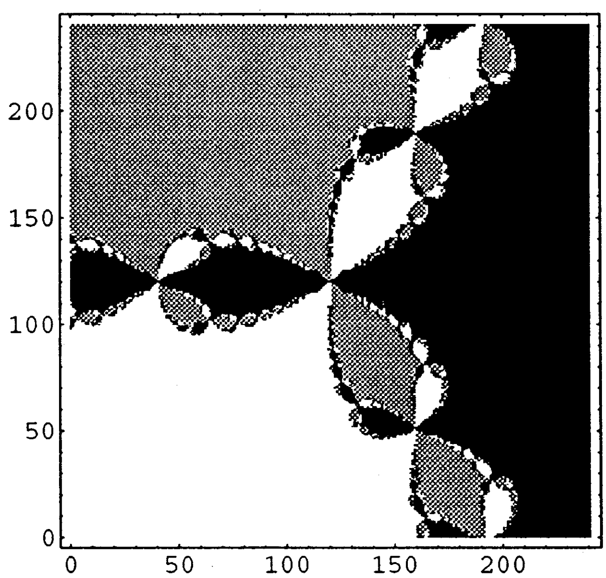
```
Newton[1/2+2I,10,100]
```

```
-0.5 + 0.866025 I
```

which is one of the known roots. The function `class` is used to classifiy the converged values according to which root they represent. To visualise the process, a matrix of the region, of whatever resolution is required, is filled with the value returned by the `Newton` for the point.

```
Table[class[Newton[x + y I,5,10]],
    {y,-1.2,1.2,0.01},{x,-1.2,1.2,0.01}];
```

```
ListDensityPlot[%, Mesh->False];
```



Each shade represents a different root (e.g. points in black converge to the root z=1).

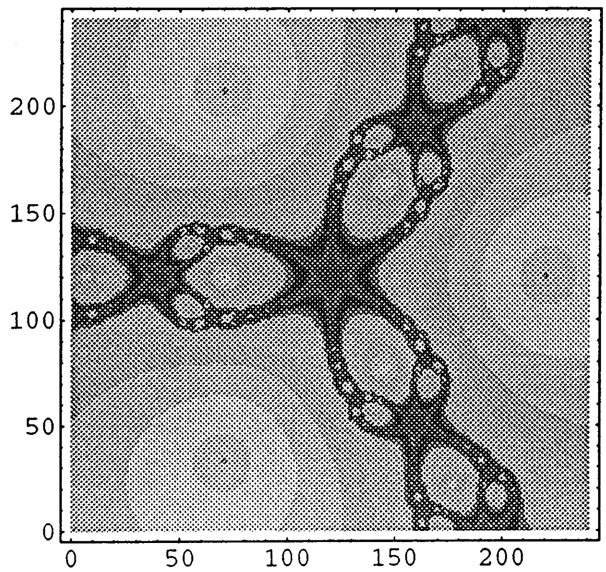
Convergence Time

If we plot the number of iterations taken to converge instead, we get another pretty pattern:

```
NewtonNumber =
  Compile[{{z, _Complex},{acc,_Real}, {m, _Integer}},
    Module[{w = z, n = 0},
        While[Abs[w^3-1] > 10^-(acc+1) && n < m,
            w = 2w/3 + 1/(3 w^2);
            n++
        ];
        n
    ]
  ];
```

```
Table[NewtonNumber[x + y I,2,50],
    {y,-1.2,1.2,0.01},{x,-1.2,1.2,0.01}];
```

```
ListDensityPlot[%, Mesh->False, ColorFunction->Hue];
```



Gravity

Show numerically (and graphically) using a differential equation solver that the effect of the Moon on a tightly orbiting Earth sattelite is to cause the sattelite's orbit to elongate in a particular direction with respect to the earth moon line. Assume that the Earth and Moon are stationary and that the initial trajectory of the sattelite is a circle in the same plane as the Earth-Moon line. Thus the coordinates `{x[t], y[t]}` of the sattelite completely specify the motion.

Solution by Daniel Oi

Stable Orbit

The Earth is approximately 100 times the mass of the Moon. The Earth is placed at (1,0) and the Moon at (-4,0). The form of the gravitational potential is

```
U:= -mm ((x[t]+4)^2+y[t]^2)^-0.5 - ((x[t]-1)^2+y[t]^2)^-0.5;
```

`mm` is the ratio of the Moon's mass to the Earth's. Currently, we will have it not contribute anything at the moment so we can get a stable orbit first.

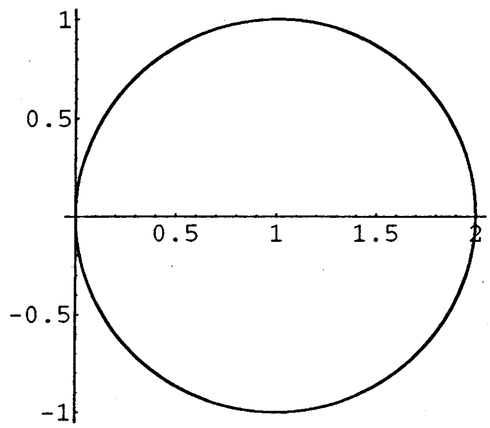
```
mm=0;
```

```
eqx = x''[t]==-D[U,x[t]]; eqy = y''[t]==-D[U,y[t]];
```

The force on an object is the negative derivative of the potential.

```
Motion=NDSolve[ {eqx,eqy, x[0]==0, y[0]==0, x'[0]==0, y'[0]==1},
     {x,y},    {t,0,65}];
```

```
ParametricPlot[Evaluate[{x[t], y[t]} /.Motion],
     {t,0,65}, AspectRatio->1];
```



This is a stable orbit without the Moon's influence.

Moon's Effect

```
mm=0.01;
```

```
eqx=x''[t]==-D[U,x[t]]; eqy=y''[t]==-D[U,y[t]];
```

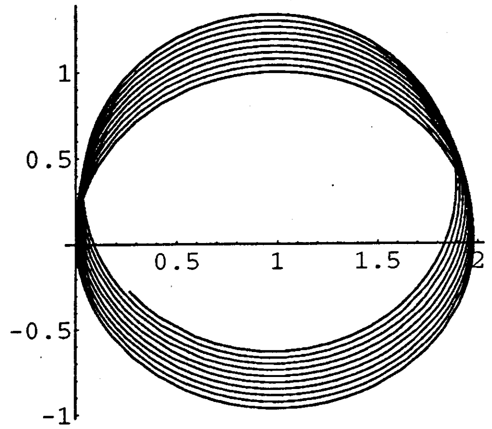
To illustrate the effect more clearly, we increase the mass of the Moon to one tenth the mass of the Earth.

```
mm=0.1;
```

```
eqx=x''[t]==-D[U,x[t]]; eqy=y''[t]==-D[U,y[t]];
```

```
Motion=NDSolve[ {eqx,eqy, x[0]==0, y[0]==0, x'[0]==0, y'[0]==1},
     {x,y},    {t,0,65}];
```

```
ParametricPlot[Evaluate[{x[t], y[t]} /.Motion],
     {t,0,61}, AspectRatio->1];
```



We see that the orbit is clearly being compressed in the Earth-Moon line and extended in the line perpendicular to the Earth-Moon line. The orbit also drifts due to the perturbation of the Moon.

Conclusions

In my opinion, it is more important to show the level of student achievement in a course rather than just describing the course itself and I hope that this is clearly demonstrated through the examples above. Of course, the student solutions chosen here were the best ones but the majority of the class made a good attempt at all the problems. I personally enjoy reading the student solutions because I feel that it gives me an insight into their thought processes.

The use of *Mathematica* Notebooks provides the students with a medium in which they can document their solution along side the computation itself. In fact, this paper itself was printed directly from the lightly edited student solutions.

The full course materials are available on-line at the URL
`http://www.pd.uwa.edu.au/Physics/Courses/Third_Year/`

Acknowledgements

It is a pleasure to acknowledge the contribution of all the students who have undertaken this course over the last four years; their input has led to many course refinements. In addition, two former students – Alistair Rowe and Peter Thomson – have assisted me greatly in marking the assignments and helping out as laboratory demonstrators.

References

[De Jong] De Jong M L *Introduction to Computational Physics*, Addison-Wesley, 1991

[Koonin and Meredith] Koonin S E and Meredith D C *Computational Physics*, Addison-Wesley, 1990

[Skeel and Keiper] Skeel R D and Keiper J B *Elementary Numerical Computing with Mathematica*, McGraw-Hill, 1993

DIFFERENT PERSPECTIVES ON USING TECHNOLOGY IN MATHEMATICS TEACHING: THE CASE OF CUBED DIGITS PROBLEM

Sergei Abramovich & Tingyao Zheng
Department of Mathematics Education, University of Georgia
105 Aderhold Hall, Athens, GA 30602-7124, U.S.A.
<sergey@moe.coe.uga.edu>, <tzengh@moe.coe.uga.edu>

Abstract: This paper focuses on three different perspectives of teaching in a technology-rich environment which allow for varying degrees of computer use in mathematical problem solving. The first one limits the use of technology to computation only. The second perspective views the use of technology as an encouragement of mathematical explorations and discovery learning. The third is concerned with issues of mathematical visualization, role of multimedia, computing complexity, and extension of original problem solving situation, which affect curriculum development. These perspectives are illustrated in the context of a cubed digits problem using a spreadsheet and a relation grapher. The use of a spreadsheet is based on its capacity to model mathematical problems formulated in a language of discrete processes depending on two integral variables. This allows to discover on an interactive spreadsheet template intriguing dynamic behaviors of the sums of cubed digits such as convergence and oscillation. Technology makes it possible to visualize what numbers attract these sums in dependence on arithmetical property of a starting value and to discover the existence of cycles. The role of relation grapher is shown to be not only an enhancement in creating visual images, but as an independent tool for number theory explorations through mathematical visualization.

Introduction

The rapid expansion of technology in mathematics education affects both teaching and learning (Kaput, 1992). As computer software becomes more sophisticated, it proves to be well-adapted to particular educational purposes, capturing more and more parts of the curriculum. Students of all ages and abilities are intrigued by computing devices, irrespective of the particular assignment given. The educational task, then, is to find ways to use appropriate technology in response to students' readiness to learn mathematics in a new environment. This, however, is not a simple enterprise. Although using more technology is called for in the teaching and learning of mathematics (National Council of Teachers of Mathematics, 1989), the effectiveness of infusing computing technology in an actual classroom depends largely upon the teacher's attitude towards technology, familiarity with software, mathematical knowledge about specific subject matters, and the environment of the classroom (Becker, 1994). This paper discusses the interplay among possible uses of technology in mathematics classroom and different perspectives and goals taken by teachers. Two generic computer applications — a spreadsheet and relation grapher (Hoffer, 1990) — are chosen as an environment. In this environment, the cubed digits problem is explored.

As far as the authors know, the existence of integers for which the sum of their cubed digits equals the integer itself was first mentioned by Hoppenot (1937). For example, 153 is such number because

$$1^3+5^3+3^3=153 .$$

After Becker (1984) we shall call the problem of exploring integers with respect to this interesting property the cubed digits problem. Carmony, McGlinn, Becker, & Millman (1984) report that this problem was used for many years in the work with inservice teachers. Recently, in a conference paper, Leitzel (1994) introduced the same problem, and he suggested that a number is a *cube* *attractor* if the sum of the cubes of its digits is the number itself. From this, the following question arises: How can one find all cube attractors?

The authors have found that a spreadsheet is an extremely appropriate tool for exploring the cubed digits problem. Particularly, its use allows the discovery of properties of integers not mentioned earlier in the literature. However, how this can be achieved depends on teachers' attitudes toward the use of technology in mathematics teaching and the following pages are just about that.

A straightforward use of computational capability of a spreadsheet

One way to use technology in mathematics instruction is employing it for computations only. This view still prevails in many classrooms. In the spirit of the metaphor of ocean waves by Kaput & Thompson (1994), this use of technology lies on the surface level of waves' activity. It is almost intrinsic. With this perspective, the user of technology is concerned with generating the answer alone, no matter what additional information emerges as a result of instructional computing and how this information could affect the classroom discourse and change its objectives. He may take advantages of different software, but he believes that technology is a means for computation rather than a learning environment. The teachers who share this view usually ignore the potential that technology can bring into mathematics instruction. They are inclined to a traditional discourse in a classroom, where their lecturing dominates and the students pursue the answers of the problems assigned by the teachers. Therefore, *the uses of technology are occasional and impulsive in the nature.* Issues that go beyond the input and output of the problem are taken as objects in a black box (Figure 1). Students are not encouraged to poke and pry the box. Computing efficiency as an issue may be considered, but not out of the context of solving an original problem.

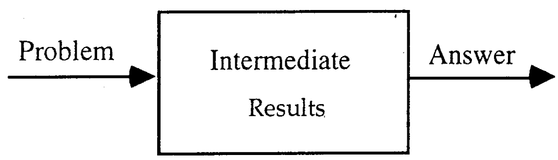


Figure 1. Model of the first perspective

There are at least two different ways one can choose to solve cubed digits problem with this stand. The most natural approach is just to check by definition which natural numbers are cube attractors. One can use a spreadsheet for that as follows. Suppose that number *N* has at most three digits: *a*, *b*, and *c*, i.e.,

$$N=100a+10b+c.$$

It follows

$$a=INT(N/100); \qquad 10a+b=INT(N/10).$$

Therefore,

$$b=INT(N/10)-10*INT(N/100); \qquad c=N-10*INT(N/10).$$

Let us define natural numbers N in column **A** so that cell **A1** is entered with 1, cell **A2** is entered with the spreadsheet function **=A1+1** which then is replicated down column **A**. In cell **B1** the spreadsheet function

=(INT(A1/100))^3+(INT(A1/10)-10*INT(A1/100))^3+(A1-10*INT(A1/10))^3 (1)

is defined and replicated down column **B**. This function computes the sums of cubes of the digits for all integers from 0 to 999. A fraction of this spreadsheet is shown in Figure 2. Here and thereafter *C* denotes the operator that transforms any integer *a* into the sum of cubes of its digits C(a).

Comparing numbers in columns **A** and **B** results in the finding of six cube attractors:

0, 1, 153, 370, 371, 407.

	A	B
140	140	65
141	141	66
142	142	73
143	143	92
144	144	129
145	145	190
146	146	281
147	147	408
148	148	577
149	149	794
150	150	126
151	151	127
152	152	134
153	153	153
154	154	190
155	155	251
156	156	342

Figure 2. A fraction of the search done by function (1)

One may notice, however, that this process is somewhat cumbersome because 1000 comparisons should be done. The improvement may be in the following slight modification of spreadsheet function (1):

=IF(A1=(INT(A1/100))^3+(INT(A1/10)-10*INT(A1/100))^3 +(A1–10*INT(A1/10))^3), A1, " ") (2)

Function (2), when replicated in column **B**, displays cube attractors only. It alleviates the burden of a routine work, because the spreadsheet substitutes the user for doing 1000 comparisons.

The second approach to find cube attractors is to solve the following equation: $a^3+b^3+c^3=100a+10b+c$ in non-negative single digits a, b, and c , or

$$c(c^2-1)=a(100-a^2)+b(10-b^2) \quad (3)$$

In order to handle an equation in three unknowns two spreadsheet templates are used (Figures 3, 4). The 10 numbers in column **B** of Figure 3 are values of $c(c^2-1)$ for c=0 to 9. In Figure 4 the 10 numbers in column **B** are values of $a(100-a^2)$ for a=0 to 9 and the 10 numbers in row 2 are $b(10-b^2)$ for b=0 to 9; the 100 other entries are their sums for different pairs of a and b.

The comparison of the entries of Figures 3 and 4 allows the finding of cube attractors. Indeed, one can find, for example, that $c(c^2-1)$ 336 for c=7. Because 336 is also an entry in Figure 4 when a = 4 and b=0, number 407 is a cube attractor. In much the same way five other cube attractors—0, 1, 153, 370, and 371—can be found. This confirms the search done before. One advantage of the second approach is in the possibility of using a relation grapher for solving the problem. This will make it possible to consider the issue of using computer-based multirepresentational strategies (Abramovich, Fujii, & Wilson, in press) with respect to the cubed digits problem. We shall discuss the use of a relation grapher later in the paper.

c	$c(c^2-1$
0	0
1	0
2	6
3	24
4	60
5	120
6	210
7	336
8	504
9	720

Figure 3. Modeling of the left-hand side of (3)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1			b	0	1	2	3	4	5	6	7	8	9
2				0	9	12	3	-24	-75	-156	-273	-432	-639
3	a												
4	0	0		**0**	9	12	3	-24	-75	-156	-273	-432	-639
5	1	99		99	108	111	102	75	**24**	-57	-174	-333	-540
6	2	192		192	201	204	195	168	117	36	-81	-240	-447
7	3	273		273	282	285	276	249	198	117	**0**	-159	-366
8	4	336		**336**	345	348	339	312	261	180	63	-96	-303
9	5	375		375	384	387	378	351	300	219	102	-57	-264
10	6	384		384	393	396	387	360	309	228	111	-48	-255
11	7	357		357	366	369	360	333	282	201	84	-75	-282
12	8	288		288	297	300	291	264	213	132	15	-144	-351
13	9	171		171	180	183	174	147	96	15	-102	-261	-468

Figure 4. Modeling of the right-hand side of (3)

Free Exploration

Another trend in using technology is to allow sufficient autonomy of the students by encouraging them to use software as a learning environment. The focus is not necessarily on the solving of an original problem. Instead, solving a problem serves only as a means to ignite students' curiosity and willingness to do mathematics. Patterns occurring as a result of computations are explored immediately in the classroom by students. In other words, the right answer is not the first priority. A rich environment, an interesting topic, and cooperative learning are the features of this kind of a class. Questions to be discussed are usually improvisational. For instance, as shown by Abramovich (1995), students using a computer, can find intriguing patterns among Pythagorean triples and then, exploring these patterns in depth, touch occasionally upon very profound properties of integers. In that setting, the teacher does not assist the students in finding the answer to the original problem; he or she just observes how technology transforms mathematics classroom into a lab that promotes "important social aspects of doing mathematics" (Steen, 1991, p. 19). Figure 5 may illustrate the mode.

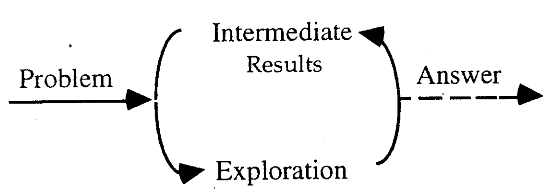


Figure 5. Model of the second perspective

The dotted line indicates that the answer may not be the sole purpose of the problem solving activity, because the objective of looking for answers may be sacrificed for improvisational learning.

Although it is difficult to predict what kind of actions the students will take for the particular problem involving cube attractors, some questions that are likely to be encountered in an open interactive learning environment include the following:

- How does the extension of the range affect the number of cube attractors?
- Is there a largest cube attractor? Why or why not?
- What are the results of applying repeatedly the operator *C* to natural numbers?

Assume that the students continue the search of attractors through modifying the spreadsheet function in such a way that it would handle integers with more than three digits (see

Figure 2). Then, after a while, the students will notice that they cannot find any new cube attractor that is greater than *407*. Although the *fill down* feature of a spreadsheet would enable students to check numbers that are significantly larger than 1000 (Figure 6), the answer is never certain. (One can also change a first entry for *a*. The spreadsheet will immediately recalculate contents of all other cells in one template according to the change of this "seed.") The search ends at students' will.

a	Ca
1467	624
1468	793
1469	1010
1470	408
1471	409
1472	416
1473	435
1474	472
1475	533
1476	624

Figure 6. Modeling beyond 1000

To answer the third question, one can use the *fill right* feature of a spreadsheet. Figure 7 is a snapshot of this action, where $C^n(a) = C(C^{n-1}(a))$ and $C^1(a) = C(a)$.

a	Ca	C^2a	C^3a	C^4a	C^5a	C^6a	C^7a	C^8a
89	1241	74	407	407	407	407	407	407
90	729	1080	513	153	153	153	153	153
91	730	370	370	370	370	370	370	370
92	737	713	371	371	371	371	371	371
93	756	684	792	1080	513	153	153	153
94	793	1099	1459	919	1459	919	1459	919
95	854	701	344	155	251	134	92	737
96	945	918	1242	81	513	153	153	153
97	1072	352	160	217	352	160	217	352
98	1241	74	407	407	407	407	407	407
99	1458	702	351	153	153	153	153	153
100	1	1	1	1	1	1	1	1
101	2	8	512	134	92	737	713	371
102	9	729	1080	513	153	153	153	153
103	28	520	133	55	250	133	55	250
104	65	341	92	737	713	371	371	371
105	126	225	141	66	432	99	1458	702
106	217	352	160	217	352	160	217	352
107	344	155	251	134	92	737	713	371
108	513	153	153	153	153	153	153	153
109	730	370	370	370	370	370	370	370
110	2	8	512	134	92	737	713	371
111	3	27	351	153	153	153	153	153
112	10	1	1	1	1	1	1	1

Figure 7. Modeling of $C^n(a)$

The result shown in Figure 7 may surprise the students because after executing the operation *C* several times, cube attractors 1, 371, 153, 370, and 407 can be seen at many cells of the template. In other words, these numbers are not only cube attractors themselves but are also the *images* of many other numbers under the operation C. Another interesting phenomenon is that for some numbers, cycles occur when the operator C is repeatedly executed. For instance, 106 is such a number.

Students may also discover that numbers having image 153 are of the form 3n, numbers having images 370 or 1 are of the form 3n + 1, and numbers having images 371 or 407 are of the form 3n + 2.

It is possible that some students may extend this problem by looking for *square attractors* or attractors with powers greater than three. These are definitely promising directions for explorations because the students can think about the generalization of approaches and extension of problems they have already explored. But again, there is no guarantee that these problems would be solved.

Bridging the use of technology and new curriculum goals

An interplay between technology and conceptual understanding.

A third way of using technology is between the two perspectives on the spectrum that we have discussed. It consists of directing students' involvement in the environment by guiding them to certain new curriculum and subject goals and helping them to gain problem solving skills. For example, new technology-rich curriculum for all students can focus on the development of self-confidence and the ability of mathematics reasoning and making connections (Lindquist, M.M., Harvey, J., & Hirsch, C., 1991). It also accommodates mathematical problems derived from the use of technology. This component should be part of a new curriculum in responding to the tremendous development of today's technology. The new curriculum emphasizes conceptual understanding—problems selected for instruction serve only as means to reach this goal. There is a definite curriculum goal set beforehand.

Although this use of technology takes into account the importance of instructional practices that are aimed at students' active involvement in the constructing of their own knowledge about mathematics, the big difference between it and free exploration is that discourse in the classroom is more organized. Furthermore, the teacher would let the students understand that the use of technology can not totally replace a mathematical mind. For instance, in the case of cubed digits problem the use of formula (1) is not very helpful in answering questions like: When should the search of cube attractors be terminated? What is the largest cube attractor? Formula (1) provides a means for computing only. Even when it seems hopeless to find a new attractor beyond 1000, it is not completely apparent why one should terminate the search unless other instructions added. These kind of issues leaves a room for more thinking.

With this perspective, a teaching plan is worked out before the class. Students' curiosity and energy of exploration is directed by the teacher toward the goal of the lesson. At times, the teacher may be willing to cope with students' discovery and needs in terms of a probably promising and valuable learning that leads to the overall objectives. The problems that the teacher poses may or may not have immediate answers. If they are short of temporary solutions, it is expected that the students' exploration of intermediate processes is worthwhile. The idea is illustrated by Figure 8.

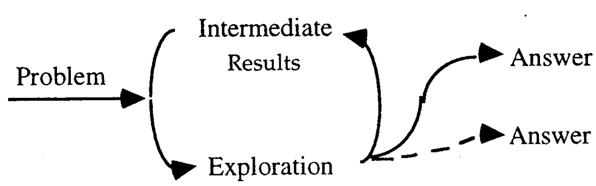


Figure 8. Model of the third perspective

As an example, consider the original problem. Exploring attractors when the number is getting bigger can prompt different subject matters and problem solving activities. Making decision about the termination of searching cube attractors can be directed to inequality. Here a discussion should lead to the explicit question: Is there exist a number $k>0$, such that for any integer $a>k$, $a\neq C(a)$? This is a first step of modeling. A stretch of this question is to ask: At least how many digits the number *a* should have to guarantee $a>C(a)$?

Suppose *a* has *n* digits. Then $a\geq 10^{n-1}$. Because $C(a)\leq n\cdot 9^3$, *a* cannot be a cube attractor, provided $10^{n-1}> n\cdot 9^3$. Taking the logarithm of both sides of the last inequality yields

$$n-1>6\lg 3+\lg n.$$

Modeling this inequality either numerically on a spreadsheet or through computer graphics results in $n>4$. Therefore, there is no cube attractor greater than 9999. This conclusion is not bad. However, before the students start to exhaust the natural numbers between 0 and 9999, the teacher can pose the following questions:

- Should one check the numbers all the way to 9999?
- Could one shrink the range of search further? If so, how can this be done?

In fact, for any $a\leq 9999$, $C(a)\leq C(9999)=2916$, no cube attractors exist beyond 2916. Therefore, the students can reduce their search significantly. This allows the students to experience the role of conceptual understanding for the effective use of technology in a problem solving setting.

As we have pointed out, new mathematics curriculum should address mathematical problems derived from using technology. When our students face technology tomorrow, it is very likely that this could happen frequently. On the other hand, problems from technology have been a source of mathematics exploration. Students should learn to deal with real-life situations. The cubed digits problem and many other problems alike that use computers and calculators bring about the issue of *computing complexity* .

The role of mathematical reasoning for solving problems on computing complexity.

Let us compare and contrast the two approaches in the first section of the paper. Because there are only 110 total entries in two templates of Figures 3 and 4, it seems that this is a substantial improvement compared to the first searching strategy. But comparing each of the ten entries in Figure 3 with the 100 entries in Figure 4 is equivalent to a total of 1000 comparisons! Therefore, one may argue that the second approach is not a real improvement at all. One actually pays the price of using his or her eyes and brain to do a job that a computer could easily handle. Why do the authors advocate this approach?

In our opinion, the teacher should help the students understand that although reducing the number of entries here is not practically significant, the issue itself is meaningful in evaluating an algorithm for computing complexity especially for large scale problems in the real world. The improvement of complexity can be emphasized as a merit of a good algorithm, in that it results in a great savings of resources. More important for the students is that *the improvement of an algorithm almost always builds a better understanding of the problem and involves mathematical reasoning and making sense of connections.* Therefore, seeking to improve a strategy, even if the reward is little, may be viewed as an important problem solving heuristic, something that should be fostered.

Within this context, there is room to further reduce the entries. For example, because $c(c^2-1)$ is always divisible by 6, one needs only those entries in Figure 4 that are multiples of 6. There exists another way to reduce the number of entries in Figure 4. Indeed, because $c(c^2-1)$ is always an even number, $a(100-a^2)$, and $b(10-b^2)$ must be both even or odd at the same time. This is equivalent to saying that *a* and *b* are both either even or odd. This fact induces the following two 5x5 templates (Figure 9 & 10), instead of a 10x10 template as in Figure 4, the entries of which are to be compared with those of Figure 3 in order to find the cube attractors. The teacher can also show how an advanced use of the spreadsheet makes it possible to improve the strategy. In fact, by using logical functions **IF**, **OR**, **AND** with the information that Figure 3 displays, all numbers in Figure 4 that are not cube attractors can be eliminated. The corresponding spreadsheet function (defined in cell **B2**) has the form

```
=IF(OR($A2*(100-$A2^2)+B$1*(10-B$1^2)=0,$A2*(100-$A2^2)
+B$1*(10-B$1^2)=6,$A2*(100-$A2^2)+B$1*(10-B$1^2)=24,
$A2*(100-$A2^2)+B$1*(10-B$1^2)=60,$A2*(100-$A2^2)
+B$1*(10-B$1^2)=120,$A2*(100-$A2^2)+B$1*(10-B$1^2)=210,
$A2*(100-$A2^2)+B$1*(10-B$1^2)=336,$A2*(100-$A2^2)+B$1*(10-B$1^2)=504,
$A2*(100-$A2^2)+B$1*(10-B$1^2)=720),$A2*(100-$A2^2)+B$1*(10-B$1^2)," ")      (4)
```

The result is shown in Figure 11.

		b	0	2	4	6	8
			0	12	-24	-156	-432
a							
0	0		**0**	12	-24	-156	-432
2	192		192	204	168	36	-240
4	336		**336**	348	312	180	-96
6	384		384	396	360	228	-48
8	288		288	300	264	132	-144

Figure 9. Modeling of the right-hand of (3) for even values of *a* and *b*

		b	1	3	5	7	9
			9	3	-75	-273	-639
a							
1	99		**108**	102	24	-174	-540
3	273		282	276	198	0	-366
5	375		**384**	378	300	102	-264
7	357		366	360	282	84	-282
9	171		180	174	96	-102	-468

Figure 10. Modeling of the right-hand of (3) for odd values of *a* and *b*

	0	1	2	3	4	5	6	7	8	9
0	0									
1						24				
2										
3								0		
4	336									
5										
6										
7										
8										
9										

Figure 11. The result of using function (4)

Some problems that the students find in their exploration may trigger useful discussions, serve to enhance the curriculum goals. For example, exploring how to narrow down one's searching range is a rich problem for learning combinatorics; studying why there are cycles and why the attractors are always the images of operator C would target number theory.

We want to emphasize that this use of technology could involve more than one software format in a learning environment. The reason for using technology is not for use itself but to meet the need of teaching and learning certain subject matters. It should not be improvisational and/or impulsive. It is a part of the curriculum chain. In this matter and for the cubed digits problem, the teacher might take into consideration that because the problem is discrete in nature, the issue of visualization is of great importance. For this reason, graphing software may be helpful. In turn, visualization can provide insight of how the problem could be solved. One of the applications useful for the cubed digits problem is Algebra Xpresser, a dynamic application for Macintosh computers. The advantage of Algebra Xpresser over other drawing applications is its ability to graph relations from any two variable equations (Abramovich, 1994). This provides an opportunity to graph an equation in two variables without the need to convert it into a form suitable for "function grapher" software.

As has been discussed, the problem can be shaped as solving equation (3) in non-negative integers *a* and *b* less than 10, given values of $c(c^2-1)$ in Figure 3. Therefore, by setting a=x, b=y, one can graph equation (3) for nine different values of its right-hand side on the *xy* plane of Algebra Xpresser (Figure 12). The solution is any integral point within the square [0, 9] X [0, 9] through which the curve passes. In this way, students are able to see how a discrete mathematical problem can be transformed into or linked to a continuous problem.

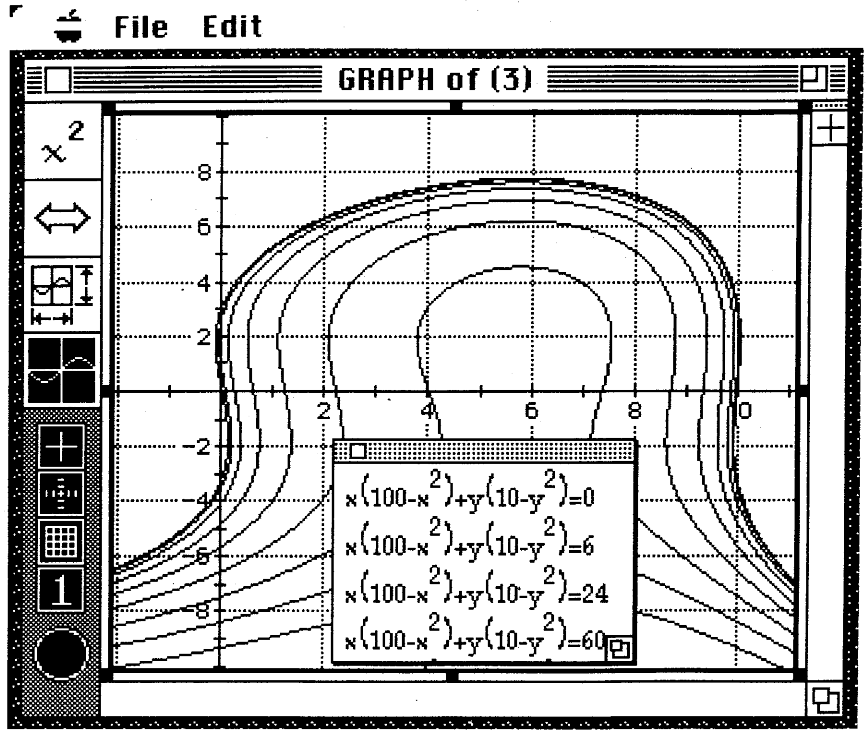


Figure 12. Graphs of equation (3) on the *ab*-plane for different *c*

Conclusion

In general, because teachers having different perspectives about using technology, they will take different approaches to solving mathematics problems when technology is available. Classroom discourse that reflects those beliefs will be also differ. The first perspective that we discussed relates to a traditional classroom. The teacher has more control, and uses technology as a simple computing tool. The second perspective suggests a more open and exploratory style of learning. The discovery made by the students themselves in their exploration in that environment is encouraged and valued. The teacher acts as a promoter of knowledge and coordinator of learning and imposes no boundaries on students' explorations. This perspective does not necessarily represent a view of a majority of teachers. The third perspective tries to employ the potential of technology in dealing with certain content knowledge and takes it as catalyst. The scope of the original problem may thus be expanded. In such a classroom, the teacher has some control of what should be involved and what should be abandoned. The classroom discourse here is more flexible than that in the first case but more teacher-directed than the second. It demands a wider view by the teacher of the relationship between technology and an innovated curriculum.

References

Abramovich, S., Fujii, T., & Wilson, J. (in press). Multiple-application medium for the study of polygonal numbers. Journal of Computers in Mathematics and Science Teaching.

Abramovich, S. Pythagorean triples and technology. Paper presented at the National Council of Teachers of Mathematics' Southern Regional Meeting, Birmingham, AL, U.S.A., 1995 (February).

Abramovich, S. Technology and teacher education: relation grapher as a tool for solving equations. Proceedings of the EMT: 30 Years of Excellence Conference (pp. 1-5). Athens, GA: The University of Georgia, 1994.

Becker, H. J. Analysis and Trends of Schools Use of New Information Technologies. Prepared for the Office of Technology Assessment, U.S. Congress (under contract: #K3-0666.0). Department of Education, University of California, Irvine, 1994 (March).

Becker, J. P. Integrating problem solving into mathematics teaching. In T. Kawaguchi (Ed.), Proceedings of ICMI—JSME Regional Conference on Mathematical Education (pp. 79-94). Tokyo: Japan Society of Mathematics Education, 1984.

Carmony, L., McGlinn, R., Becker, J.P., & Millman, A. Problem Solving in Apple Pascal. Rockland, MD: Computer Science Press, 1984.

Hoffer, A. Algebra Xpresser. Macintosh, Version 1.1. System 6.0.3 or above. MA: William K. Bradford Publishing Co., 1990.

Hoppenot, F. Courrier du "Sphinx". Sphinx 5 . Bruxelles, p.72, 1937.

Kaput, J. J. Technology and mathematics education. In D. A. Grouws (Ed.), Handbook of Research on Mathematics Teaching and Learning, (pp. 515-556). New York: Macmillan, 1992.

Kaput, J. J., & Thompson, P. W. Technology in mathematics education research: The first 25 years in the *JRME*. Journal for Research in Mathematics Education, 25(6), pp. 676-684, 1994.

Leitzel, J. R. C. The lure of number theory: More questions than answers. Paper presented at the National Council of Teachers of Mathematics' 72nd Annual Meeting, Indianapolis, IN, U.S.A., 1994 (April).

Lindquist, M.M., Harvey, J., & Hirsch, C. How mathematics content courses must change in light of technology and the NCTM curriculum standards. In F.Demana, B.K.Waits, & J.Harvey, (Eds.). Proceedings of the Second Annual Conference on Technology in Collegiate Mathematics: teaching and learning with technology. (pp.20-27). Addison-Wesley Pub. Co., 1991.

National Council of Teachers of Mathematics. Curriculum and evaluation standards for school mathematics. Reston, Va.: The Council, 1989.

Steen, A.L. Twenty questions for computer reformers. In F. Demana, B.K. Waits, & J. Harvey, (Eds.). Proceedings of the Second Annual Conference on Technology in Collegiate Mathematics: teaching and learning with technology (pp. 16-19). Addison-Wesley Pub. Co., 1991.

A GLANCE AT ALGORITHMIC MATHEMATICS

Rajendra Akerkar
Research Institute for Symbolic Computation
J. Kepler University
4040 Linz, Austria
e-mail: rakerkar@risc.uni-linz.ac.at

Abstract

Mathematical software systems are based on advances in algorithmic mathematics.Main goal of the researchers in mathematical software community has been to develop algorithms that are efficient in practice. In this paper, we list some characteristic features of algorithmic algebra. Then we present some areas of application in various subareas of computer science and engineering and give some problem areas in algorithmic algebra. We conclude our discussion with presenting some of the limits of algorithmic algebra.

I. Areas of application for algorithmic algebra

Mathematical software systems are based on advances in algorithmic mathematics. Algorithmic mathematics, which allows only algorithmic functors for the solution of the problems. In this paper, we are going to discuss about algorithmic algebra, which is a branch of algorithmic mathematics. It is really next to impossible to give a definition of algorithmic algebra that would satisfy most of the researchers in this area of research. Nevertheless, we can quote the definition of Loos [Loo 83] :

Algorithmic algebra is that subfield of scientific computation, which develops, analyses, implements, and applies algebraic algorithms.

As a consequence of these characteristics, decision algorithms can be built on computer algebra methods.

In order to distinguish computer algebra from other research areas in scientific computation, let us list some characteristic features of computer algebra.

(1) computation in algebraic structures

- there is a variety of algebraic structures for which computer algebra provides algorithms. Some of them are:

  basic domains $\mathbb{Z}, \mathbb{Q}, \mathbb{Z}_p, \mathbb{Q}_p$ (p–adic numbers), ...

  algebraic extensions $\mathbb{Q}(\alpha)$, $p(\alpha) = 0$

  polynomials $R[x_1, \ldots, x_n]$

  rational functions $K(t_1, \ldots, t_n)$

  matrices

  differential fields $(K,')$, difference fields (K, ∇)

  finitely presented groups $\{a, b \mid aba = 1, bab = 1\}$

  ...

(2) manipulation of formulas

typically, in computer algebra we want to compute

$$\int \frac{x}{x^2-1} dx = \frac{1}{2} \ln |x^2 - 1|$$

instead of

$$\int_0^{\frac{1}{2}} \frac{x}{x^2-1} dx = 0.1438...$$

i.e. in general both the input and the output of algorithms and programs are mathematical expressions or formulas rather than numbers.

(3) exact computation

exact computation rather than approximate computation is the goal of computer algebra.

So, typically, we want to compute

$$(\frac{\sqrt{3}}{2}, -1/2)$$

instead of

$$(0.86602..., -0.5)$$

as the solution of the system of equations

$$x^4 + 2x^2y^2 + 3x^2y + y^4 - y^3 = 0$$
$$x^2 + y^2 - 1 = 0$$

In this way, the problem is reduced to the question:

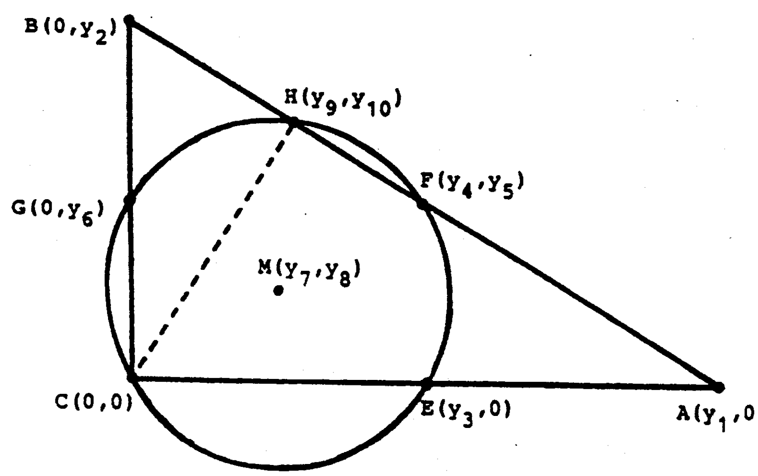
"Can two points P_1, P_2 in a semi-algebraic set L be connected by a path, i.e. are they in the same component of L?"

This question can be decided, for instance, by Collins' cylindrical algebraic decomposition method [ACM 84].

<u>theorem proving in elementary geometry</u>

The issue is to automatically prove geometric statements that can be described by polynomial equations. An example of such a theorem is:

"the altitude pedal of the hypothenuse of a right-angled triangle and the midpoints of the three sides of the triangle lie on a circle"



hypotheses:

$$h_1 \equiv 2y_3 - y_1 = 0 \qquad (\text{E is midpoint of } \overline{\text{AC}})$$

$$h_2 \equiv (y_7 - y_3)^2 + y_8^2 - (y_7 - y_4)^2 - (y_8 - y_5)^2 = 0$$

$$(\overline{\text{EM}} \text{ and } \overline{\text{FM}} \text{ are equally long})$$

$$\vdots$$

$$h_m$$

conclusion:

$$c \equiv (y_7 - y_3)^2 + y_8^2 - (y_7 - y_9)^2 - (y_8 - y_{10})^2 = 0$$

$$(\overline{\text{EM}} \text{ and } \overline{\text{HM}} \text{ are equally long})$$

The geometric problem is reduced to the algebraic problem:

"show that $c \in \text{radical}(h_1, \ldots, h_m)$*"*

This algebraic problem can be decided by the method of Gröbner bases [Buc 89].

modelling in science and engineering

In science and engineering, it is common to express a problem in terms of integrals or differential equations with boundary conditions. Numerical integra[illegible] to approximations of the values of the solution functions. But, as R.W. Ham[illegible] written, "the purpose of computing is insight, not numbers." So instead of computing tables of values it would be much more gratifying to derive formulas for the solution functions. Computer algebra algorithms can do just that for certain classes of integration and differential equation problems.

Consider, for example, the system of differential equations

$$-6\frac{\partial q}{\partial x}(x) + \frac{\partial^2 p}{\partial x^2}(x) - 6\sin(x) = 0,$$
$$6\frac{\partial^2 q}{\partial x^2}(x) + a^2\frac{\partial p}{\partial x}(x) - 6\cos(x) = 0$$

subject to the boundary conditions $p(0) = 0, q(0) = 1, p'(0) = 0, q'(0) = 0$. Given this information as input, any of the major computer algebra systems will derive the formal solution

$$p(x) = -\frac{12\sin(ax)}{a(a^2-1)} - \frac{6\cos(ax)}{a^2} + \frac{12\sin(x)}{a^2-1} + \frac{6}{a^2},$$
$$q(x) = \frac{\sin(ax)}{a} - \frac{2\cos(ax)}{a^2-1} + \frac{(a^2+1)\cos(x)}{a^2-1}$$

for $a \notin \{-1, 0, 1\}$.

III. Some problem areas in computer algebra

Solution of algebraic equations

If $f(x), g(x)$ are polynomials in $R[x]$, R a commutative ring with 1, the resultant of f and g, $\text{res}_x(f,g)$, is a constant in R, the determinant of the Sylvester matrix of f and g.

Fact: $\text{res}_x(f,g) = 0$ *if and only if f and g have a common factor, or, in other words, there is a solution to the system of equations* $f(x) = g(x) = 0$.

Let us consider the following example:

$$f_1 = xz - xy^2 - 4x^2 - \frac{1}{4} = 0$$
$$f_2 = y^2 z + 2x + \frac{1}{2} = 0$$
$$f_3 = x^2 z + y^2 + \frac{1}{2}x = 0$$

First we eliminate the variable z:

$$g_1(x,y) = \mathrm{res}_z(f1, f2) = y^4x + 4x^2y^2 + \frac{1}{4}y^2 + 2x^2 + \frac{1}{2}x$$

$$g_1(x,y) = \mathrm{res}_z(f_2, f_3) = y^4 - 2x^3 + \frac{1}{2}xy^2 - \frac{1}{2}x^2$$

Now we eliminate the variable y:

$$h = \mathrm{res}_y(g_1, g_2) =$$
$$\frac{1}{1024} \cdot x^4 \cdot (4x+1)^2 \cdot (32x^5 - 216x^4 + 64x^3 - 42x^2 + 32x + 5)^2$$

Every solution of the system has to have an x-coordinate which is a root of $h(x)$. Unfortunately, not every root of $h(x)$ can be extended to a solution of the original system. There are certain *extraneous factors* in the resultant. In our example, only the last factor gives the correct x-coordinates.

We could also use the method of Gröbner bases for solving this (or in fact any other) system of algebraic equations. If

$$F = \{f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)\}$$

is given (as the basis for a polynomial ideal, or in other words all the linear combinations of $f_1, \ldots, f_m$, or in other words all the consequences of relating these polynomials to 0), then the Gröbner basis of F (w.r.t. some ordering of the power products)

$$GB(F) = \{g_1(x_1, \ldots, x_n), \ldots, g_r(x_1, \ldots, x_n)\}$$

is another basis for this same ideal. $GB(F)$ has certain nice properties.

Fact: *The Gröbner basis of F w.r.t. a lexicographic ordering is a triangu*[illegible] *system of polynomials having the same solutions as the original system F. All the solutions of the univariate polynomial in the Gröbner basis of a zero–dimensional ideal can be extended to solutions of the whole system of polynomial equations.*

So in order to solve the above system of algebraic equations, we compute a Gröbner basis for $F = \{f_1, f_2, f_3\}$. This results in the new (and equivalent) system

$$65z + 64x^4 - 432x^3 + 168x^2 - 354x + 104 = 0$$
$$26y^2 - 16x^4 + 108x^3 - 16x^2 + 17x = 0$$
$$32x^5 - 216x^4 + 64x^3 - 42x^2 + 32x + 5 = 0$$

The univariate polynomial $g_3(x)$ in the Gröbner basis is irreducible. There are 5 solutions of this univariate polynomial, all of which can be continued to solutions of the full system. Thus we get exactly 10 solutions of the whole system (counting multiple solutions).

Now we could either stay in an algebraic mode of computation, extend the rational numbers by a root of g_3, and solve the system in such a way. Or we could easily approximate the roots of the system numerically. An approximation of one of the roots is $(-0.128475, 0.321145, -2.356718)$.

Indefinite summation

Given a sequence

$$a_1, a_2, \ldots$$

(i.e. an expression generating this sequence), we want to find an expression $S(j)$, in which the summation sign has been eliminated and such that

$$S(j) = \sum_{i=1}^{j} a_i.$$

This is a discrete analogue to indefinite integration. $S(j)$ may be thought of as an 'anti-derivative'. Having obtained $S(j)$ we then have

$$\sum_{i=m}^{n} a_i = S(n) - S(m-1)$$

where $S(0) = 0$. For instance,

$$S(n) = \frac{3n^2 + 5n}{4(n^2 + 3n + 2)} \quad \text{for} \quad \sum_{i=1}^{\infty} \frac{1}{i(i+2)}.$$

Gosper's algorithm [Gos 78] can produce indefinite summation for hypergeometric series, i.e. for series

$$\sum_{i=1}^{\infty} a_i$$

such that a_n/a_{n-1} is a rational function in n. For instance,

$$\sum_{i=1}^{n} i \cdot x^i = \frac{n \cdot x^{n+2} - (n+1) \cdot x^{n+1} + x}{(x-1)^2}.$$

Every major computer algebra system contains an implementation of Gosper's summation algorithm.

Indefinite integration

Let a (real or complex) function

$$f(x)$$

be given by an expression. We want to compute another expression describing a function

$$F(x)$$

such that

$$\int f(x)dx = F(x).$$

Computer algebra is able to do just that for a certain wide class of functional expressions (*elementary functions*), which can be roughly described as

- start with rational functions $p(x)/q(x)$,
- add logarithms and exponentials of expressions, and therefore trigonometric functions,
- add algebraic elements, i.e. roots of polynomials.

This class of functional expressions contains most of the functions treated in integral tables.

Of course computer algebra systems can integrate any rational function, e.g. for

$$r(x) = \frac{3x^{11} - 2x^{10} + 7x^9 + 2x^7 + 23x^6 - 10x^5 + 18x^4 - 9x^3 + 8x^2 - 3x + 1}{3x^9 - 2x^8 + 7x^7 - 4x^6 + 5x^5 - 2x^4 + x^3}$$

we obtain

$$\int r(x)\,dx = \frac{2\log(3x^2 - 2x + 1)}{3} - \frac{\log(x^2+1)}{2} + \frac{4\arctan((6x-2)/2\sqrt{2})}{3\sqrt{2}} + \log(x) + \arctan(x) + \frac{4x^3 - 4x^2 + 2x - 1}{2x^4 + 2x^2} + \frac{x^3}{3}.$$

Transcendental elementary functions can be integrated by Risch's algorithm [Ris 69]. It takes a transcendental elementary function $f(x)$, decides whether its integral can be expressed as an elementary function $g(x)$, and if so computes $g(x)$.

For instance, for

$$f(x) = 2x \cdot \exp(x^2) \cdot \log(x) + \frac{\exp(x^2)}{x} + \frac{\log(x) - 2}{(\log^2(x) + x)^2} + \frac{(2/x) \cdot \log(x) + 1/x + 1}{\log^2(x) + x}$$

we get

$$\int f(x) = \exp(x^2) \cdot \log(x) - \frac{\log(x)}{\log^2(x) + x} + \log(\log^2(x) + x).$$

Risch's algorithm is implemented in many computer algebra systems.

<u>Greatest common divisors of polynomials</u>

The computation of greatest common divisors (gcd) of polynomials is ubiquitous in computer algebra. Therefore we need very fast algorithms for gcd computation. One of the problems in gcd computation is the explosion of the coefficients. Over the last two decades sophisticated algorithms have been developed for tackling this problem.

For instance, consider the two bivariate polynomials

$$f(x,y) = y^6 + xy^5 + x^3y - xy + x^4 - x^2,$$
$$g(x,y) = xy^5 - 2y^5 + x^2y^4 - 2xy^4 + xy^2 + x^2y$$

with integral coefficients. Let y be the main variable, so that the coefficients of powers of y are polynomials in x.

Euclid's algorithm yields the polynomial remainder sequence

$$\begin{aligned}
r_0 =& f,\\
r_1 =& g,\\
r_2 =& (2x - x^2)y^3 + (2x^2 - x^3)y^2 + (x^5 - 4x^4 + 3x^3 + 4x^2 - 4x)y + x^6 - 4x^5 + 3x^4 +\\
& 4x^3 - 4x^2,\\
r_3 =& (-x^7 + 6x^6 - 12x^5 + 8x^4)y^2 + (-x^{13} + 12x^{12} - 58x^{11} + 136x^{10} - 121x^9 - 117x^8 +\\
& 362x^7 - 236x^6 - 104x^5 + 192x^4 - 64x^3)y - x^{14} + 12x^{13} - 58x^{12} + 136x^{11} - 121x^{10}\\
& - 116x^9 + 356x^8 - 224x^7 - 112x^6 + 192x^5 - 64x^4,\\
r_4 =& (-x^{28} + 26x^{27} - 308x^{26} + 2184x^{25} - 10198x^{24} + 32188x^{23} - 65932x^{22} + 68536x^{21} +\\
& 42431x^{20} - 274533x^{19} + 411512x^{18} - 149025x^{17} - 431200x^{16} + 729296x^{15} -\\
& 337472x^{14} - 318304x^{13} + 523264x^{12} - 225280x^{11} - 78848x^{10} + 126720x^9 -\\
& 53248x^8 + 8192x^7)y\\
& - x^{29} + 26x^{28} - 308x^{27} + 2184x^{26} - 10198x^{25} + 32188x^{24} - 65932x^{23} +\\
& 68536x^{22} + 42431x^{21} - 274533x^{20} + 411512x^{19} - 149025x^{18} - 431200x^{17} +\\
& 729296x^{16} - 337472x^{15} - 318304x^{14} + 523264x^{13} - 225280x^{12} - 78848x^{11} +\\
& 126720x^{10} - 53248x^9 + 8192x^8.
\end{aligned}$$

The greatest common divisor of f and g is obtained by eliminating common univariate factors $p(x)$ in r_4. So actually the gcd turns out to be

$$\gcd(f,g) = y + x.$$

Although the inputs and the output are small, the intermediate expressions get very big. The biggest polynomial in this computation happens to occur in the pseudo-division of r_3 by r_4. The intermediate polynomial has degree 70 in x.

Fortunately, there are several ways of reducing this enormous growth of coefficients. The most efficient algorithm for computing gcd's of multivariate polynomials is a modular algorithm. The basic idea is to apply homomorphisms to the coefficients, compute the gcd's of the evaluated polynomials, and use the *Chinese remainder algorithm* to reconstruct the actual coefficients in the gcd.

If the input polynomials are univariate, we can take homomorphisms H_p, mapping an integer a to $a \bmod p$. If the input polynomials are multivariate, we can take evaluation homomorphisms of the form $H_{x_1=r_1}$ for reducing the number of variables.

In our example we get

$$\gcd(H_{x=2}(f), H_{x=2}(g)) = y + 2,$$
$$\gcd(H_{x=3}(f), H_{x=3}(g)) = y + 3.$$

So the gcd is $y + x$. Never during this algorithm did we have to consider big coefficients.

Factorization of polynomials

Like gcd computation, factorization of polynomials is a basic building block of computer algebra algorithms. For factoring a univariate polynomial

$$f(x)$$

with integer coefficients, we proceed in the following way:

- factor $f(x)$ modulo a prime p,
- determine a bound B for the coefficients of factors of f and an integer k s.t. $p^k > B$,
- lift the factorization of $f \bmod p$ to a factorization of $f \bmod p^k$,
- combine factors in order to get the factors over the integers.

For instance, consider the polynomial

$$f(x) = 6x^7 + 7x^6 + 4x^5 + x^4 + 6x^3 + 7x^2 + 4x + 1.$$

Taking $p = 5$, we get the factorization

$$f \equiv (x-2)(x^2-2)(x^2+2)(x^2-x+2) \pmod{5}.$$

Lifting this to a factorization modulo 5^2 we obtain

$$f \equiv (2x+1)(x^2-7)(x^2+7)(3x^2+2x+1) \pmod{25}.$$

Now we can combine the second and third factors to obtain the factorization over the integers

$$f = (2x+1)(x^4+1)(3x^2+2x+1).$$

References

[ACM 84] D.S.Arnon, G.E.Collins, S.McCallum, *Cylendrical Algebraic Decomposition I: The Basic Algorithm*, SIAM J. Computation 13/4, 865-877 (1984)

[Buc 89] *Applications of Gröbner Bases in non-Linear Computational Geometry*, Geometric Reasoning, D.Kapur and J.Mundy (eds.), 413-446, The MIT Press (1989)

[CF 76] B.F.Caviness,R.J.Fateman, *Simplification of Radical Expressions*, Proc. 1976 ACM Symp. on Symbolic and Algebraic Computation, R.D.Jenks (ed.), ACM, 329-338 (1976)

[Gos 78] R.W. Gosper, *Decision procedure for indefinite hypergeometric summation*, Proc. Natl. Acad. Sci. USA 75/1, 40-42 (1978)

[KB 70] D.E. Knuth, P.B.Bendix, *Simple word problems in Universal Algebras*, Computational Problems in Abstract Algebra, J.Leech (ed.), 263-297, Pergamon Press (1970)

[Loo 83] R.Loos, *Introduction*, Computer Algebra - Symbolic and Algebraic Computation (2nd edition), B.Buchberger, G.E.Collins, and R.Loos (eds.), 1-10, Springer-Verlag (1983)

[Ris 69] R.H.Risch, *The Problem of Integration in Finite Terms*, Trans. AMS 139, 167-189 (1969)

Graphical Approach to Solve Two Problems

David L Albig, James J. Corbet, and Wei-Chi Yang

Department of Math/Stat, Radford University,
Radford, VA 24142, U.S.A.

1 A probability problem

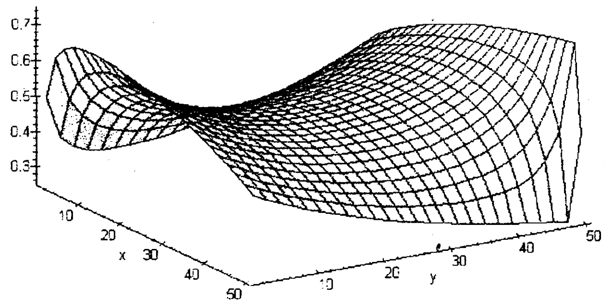
We shall use the numeric, graphic, and analytic approaches with the computer software Scientific Workplace and Maple V to maximize (or minimize) a simple probability function.

Example 1 *You are presented with two urns and an equal number of black and white balls. You may place the balls in the urns as you please. Then a friend will enter the room, select an urn at random and draw one ball from it. How should you place the balls in order to maximize the probability that your friend draws a white ball?*

First we call the urn which has the greater number of white balls urn 1. Then set

$$\begin{aligned} x &= \text{the number of white balls in urn 1} \\ y &= \text{the number of black balls in urn 1} \\ n &= \text{the total number of white balls} \\ &= \text{the total number of black balls.} \end{aligned}$$

We define the probability function$P(x,y,n) = \frac{1}{2}(\frac{x}{x+y} + \frac{n-x}{2n-(x+y)})$.
Let us investigate the maximum of P with $n = 50$. First we notice that $P(1,0,50) = \frac{74}{99}, P(0,1,50) = \frac{25}{99}, P(50,1,50) = \frac{25}{51}$. Next we graph the function $P(x,y,50)$ as follows:



Remark 1 *We observe from the graph of $P(x, y, 50)$ that the maximum probability of obtaining the white ball is near the left edge, since the surface is higher on both ends, which means that y is near 0, or near 50. But y near 50 is impossible, y must be close to 0. Let's see if our intuition is correct.*

Now we apply the partial derivative test to see if we can find the maximum of $P(x, y, 50)$.

$$\frac{\partial}{\partial x}P(x,y,50) = \frac{-125y^2 + 5000y + y^3 - 25x^2 - 150xy + x^2y + 2xy^2}{(x+y)^2(-100+x+y)^2}$$

$$\frac{\partial}{\partial y}P(x,y,50) = -\frac{5000x - 125x^2 - 150xy + x^3 + 2x^2y + xy^2 - 25y^2}{(x+y)^2(-100+x+y)^2}$$

By solving

$$\begin{cases} \frac{\partial}{\partial x}P(x,y,50) = 0 \\ \frac{\partial}{\partial y}P(x,y,50) = 0 \end{cases}.$$

the solution is

$$\{y = 25, x = 25\}$$

We find the second partial derivatives as follows:

$$D_{xx}P(x,y,50) = -\frac{1}{(x+y)^2} + \frac{x}{(x+y)^3} - \frac{1}{(100-x-y)^2} + \frac{50-x}{(100-x-y)^3}$$

$$D_{yy}P(x,y,50) = \frac{x}{(x+y)^3} + \frac{50-x}{(100-x-y)^3},$$

$$D_{xy}P(x,y,50) = -\frac{1}{2\,(x+y)^2} + \frac{x}{(x+y)^3} - \frac{1}{2\times(100-x-y)^2} + \frac{50-x}{(100-x-y)^3},$$

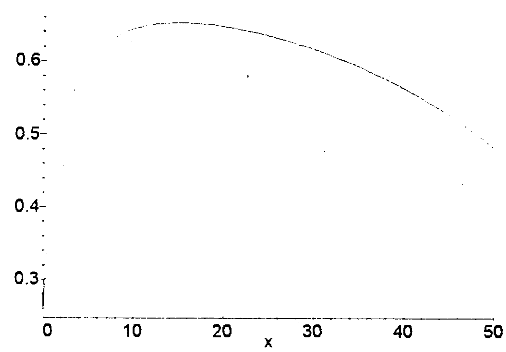
$$D_{yx}P(x,y,50) = D_{xy}P(x,y,50,)$$

and thus

$$D_{xx}P(25,25,50) \times D_{yy}P(25,25,50) - D_{xy}P(25,25,50)^2 = 0.$$

There is no conclusion due to the Second-Partials Test.

Next we shall investigate the probability if y is kept as a constant.
Case 1: Consider $P(x,2,50) = \frac{1}{2}\frac{x}{x+2} + \frac{1}{2}\frac{50-x}{98-x}$. The graph is shown below:
$P(x,2,50)$



By solving the equation

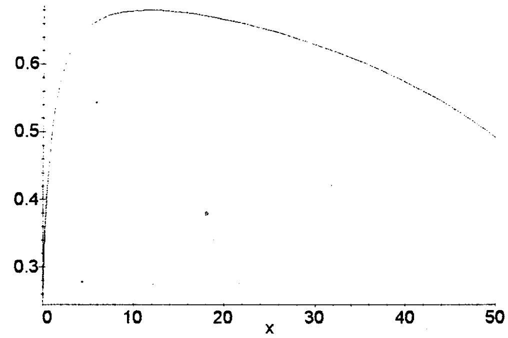
$$D_xP(x,2,50) = 0, \text{ and let } x \in [0,50],$$

the solution is

$$\{x = 14.95208\},$$

but $P(14,2,50) = .6517857$, and $P(15,2,50) = .6520198$. We see that the maximum of $P(x,2,50)$ is $P(15,2,50) = .6520198$.

Case 2: Consider $P(x,1,50) = \frac{1}{2}\frac{x}{x+1} + \frac{1}{2}\frac{50-x}{99-x}$. We plot this function as follows:
$P(x,1,50)$



We note that the minimum probability of obtaining the white ball with one black ball is

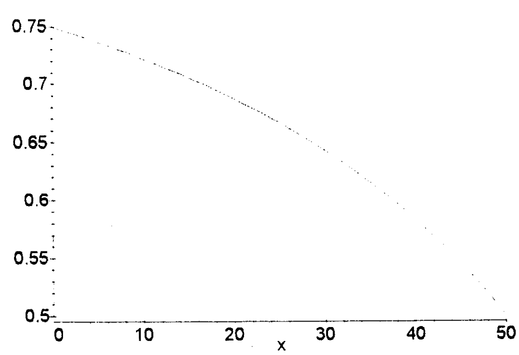
$$P(0,1,50) = \frac{25}{99}$$

Set

$$D_x P(x,1,50) = 0, \text{ and let } x \in [0,50],$$

we obtain the solution to be $\{x = 11.5\}$. Thus the maximum is either $P(11,1,50)$ or $P(12,1,50)$, but $P(11,1,50) = .6799242$ and $P(12,1,50) = .6799293$. Thus the maximum of $P(x,1,50)$ is $P(12,1,50) = .6799293$.

Case 3: Consider $P(x,0,50) = \frac{1}{2} + \frac{1}{2}\frac{50-x}{100-x}$. The graph is shown below: $P(x,0,50)$



We could find the extremum of $P(x,0,50)$ as we did in cases 1 and 2. But from the graph of $P(x,0,50)$, we predict the minimum to be $P(50,0,50) = .5$. Since $P(0,0,50)$ is undefined, the maximum has to be $P(1,0,50) = .7474747$.

Remark 2 *(1) As we expected, from the graph of* $P(x,y,50)$

$$P(1,0,50) > P(12,1,50) > P(15,2,50)$$

(2) We have used the computation capability of Scientific Workplace to find the derivatives, relative extremum, two dimensional plots and used Maple to do three dimensional plots.

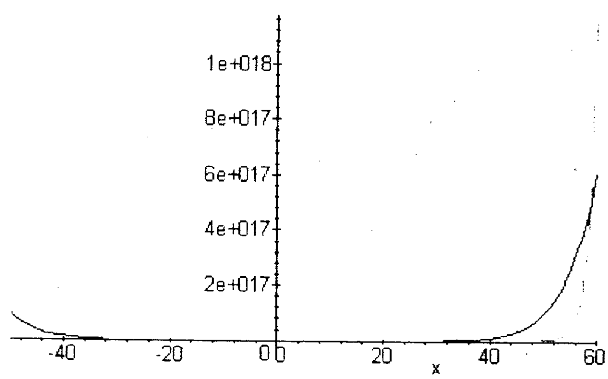
(3) It is shown algebraically that $P(1,0,50)$ indeed gives the maximum in a talk given by J.J. Corbet and D.L. Albig at the fall 1986 meeting of MD-DC-VA section of the MAA meeting, see [AC].

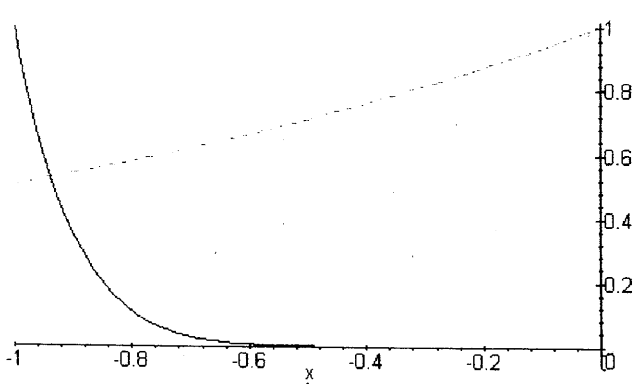
2 Fixed Point and Newton Iterations: For what n is $2^n > n^{10}$?

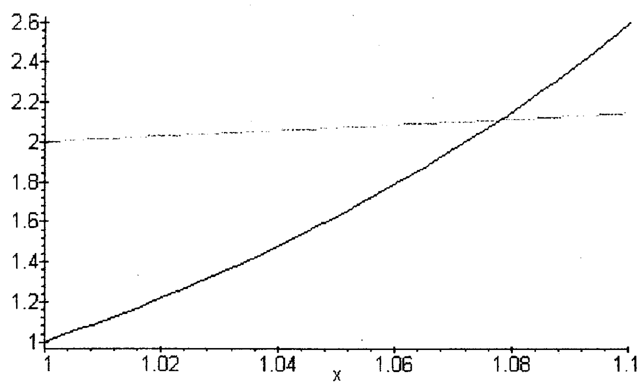
This problem originated from [T]. With available technology, we shall see how we solve this problem differently. We first graph the functions 2^x and x^{10} and see roughly where the possible solutions for

$$2^x = x^{10}$$

are?







We predict that there are three solutions for $2^x = x^{10}$. We shall use fixed-point and Newton's iterations to find these roots.

Step 1: We set $f(x) = \frac{10}{\ln 2} \ln x$, and use fixed-point iteration on f and with the help of Scientific Workplace, we obtain

$$\begin{array}{c} 50.0 \\ 56.43856 \\ 58.18609 \\ 58.62602 \\ 58.73469 \\ 58.76141 \\ 58.76797 \\ 58.76958 \\ 58.76998 \\ 58.77007 \\ 58.7701 \end{array}$$

We obtain the first solution $x_1 \approx 58.7701$.

Step 2: We set $g(x) = 2^{0.1x}$ and use fixed-point iteration on g and obtain

$$\begin{array}{c} -.5 \\ .9659363 \\ 1.069246 \\ 1.07693 \\ 1.077504 \\ 1.077547. \\ 1.07755 \\ 1.07755 \\ 1.07755 \\ 1.07755 \\ 1.07755 \end{array}$$

Thus $x_2 \approx 1.07755$.

Step 3: We shall apply Newton's method to obtain the third root We remark that the fixed-point algorithm fails on this one, see [R].

First we define $h(x) = 2^x - x^{10}$, and the Newton iteration function

$$N(x) = x - \frac{h(x)}{h'(x)}.$$

We apply the iteration on N (with the help from Scientific Workplace) and obtain the following:

$$\begin{array}{c} -2.0 \\ -1.800056 \\ -1.620213 \\ -1.458662 \\ -1.314134 \\ -1.186466 \\ -1.077959 \\ -.9956423 \\ -.9498938 \\ -.937813 \\ -.9371114 \end{array}$$

Thus $x_3 \approx -.9371114$. We conclude that $2^n > n^{10}$ if approximately

$$n \in [58.7701, \infty) \cup (-.9371114, 1.07755)$$

3 Conclusion

- Without the computer software, the preceding problem can not be solved easily unless we learn how to do the required computer programming. Thus, there is no doubt that the computer algebra systems made an impact on learning and teaching mathematics. Ten years ago these problems would have been much more difficult using our approach. One might speculate as to how difficult they may be ten years in the future. We are certain that with the help of powerful computer algebra systems, users can discover new methods of solving problems, make intuitive conjectures, which may lead to the discovery of new theorems.

- We feel that software developers should look into making their products accessible to a broader range of users. They should not assume the user knows the programming languages. The rule of thumb should be that if the user knows mathematics, he or she should be able to use the software.

References

[AC] D.L. Albig & J.J. Corbet. Three Especially Interesting Problems for Undergraduates. Problem Corner. Virginia Mathematics Teacher. Spring 1988.

[T] Travers, For what n is $2^n > n^{10}$? Mathematics Teaching. Harper and Row. 1977, page 146.

[R] Ralston, A First Course in Numerical Analysis. McGraw-Hill, 1965, page 384, 5(b).

fmat1117@selu.edu

FAMILIES OF PLANE CURVES BOUNDING A CONSTANT AREA

Tilak de Alwis
Department of Mathematics
Southeastern Louisiana University
Hammond, LA 70402
USA

1. Introduction In this paper, we will consider several one-parameter families of plane curves bounding a constant area in the first quadrant. At the beginning, we will consider one-parameter families of straight lines, parabolas and ellipses. The envelopes of these families are of significant interest. We will use the computer algebra system *Mathematica* to make certain conjectures regarding these envelopes. We will also show how to use *Mathematica* to create animations to illustrate the behavior of the above curves as the parameter changes. Finally, we consider more general types of one-parameter families of curves bounding a constant area.

The computer algebra systems have completely changed the way we teach and think about mathematics. This paper will illustrate the great impact of the computer algebra systems on the teaching and learning of mathematics.

2. The envelope of a family of plane curves

Intuitively, the envelope of a family of plane curves is a curve or a several set of curves which is tangent to each curve of the family, and at each point is tangent to some curve of the given family (see [1] and [4]). A more mathematically sound definition can be given as follows:

Definition 2.1 Consider the family of curves in $\mathbb{R}^2$ given by $f(x,y,t) = 0$, where t is a real parameter. Then its envelope consists of all points $(x,y) \in \mathbb{R}^2$ with the property that

$$f(x,y,t) = 0$$
$$\frac{\partial}{\partial t}f(x,y,t) = 0$$

for some $t \in \mathbb{R}$.

To see how this definition corresponds to the intuitive idea of envelope, the reader must refer to [3] or [4].

Example 2.1 Consider the family of curves given by $y = 2xt + 3t^2$ where $t \in \mathbb{R}$. To compute its envelope, we eliminate t from the following two equations, as given by definition 2.1.

$$y = 2xt + 3t^2$$
$$0 = 2x + 6t$$

The t-elimination between the above two equations yields the equation $y = -x^2/3$, which is the required envelope. Therefore, in this case, the envelope of the given family is a parabola.

The theory of envelopes is in fact a quite old subject. It has been discussed in very many old texts in geometry, coordinate geometry, and mathematical analysis (see [1], [2] and [5]). However, due to the general difficulty in computing and drawing them, the subject has lost some

fmat1117@VEGA SELU.EDU

algebra systems, such as *Mathematica*, has shed a new light on the subject. For some good references on Mathematica, see [6], [7] and [8]. Such computer algebra systems can generate graphs of families of curves with relative ease. The following *Mathematica* program graphs the family of curves in Example 2.1 and also its envelope $y = -x^2/3$.

Program 2.1

```
Show[{Plot[Evaluate[Table[2x*t+3t^2,{t,-2,2,0.05}]],{x,-10,10},
DisplayFunction->Identity], Plot[-x^2/3,{x,-10,10},PlotStyle->
{RGBColor[1,0,0],Thickness[1/80]}, DisplayFunction->Identity]},
DisplayFunction -> $DisplayFunction]
```

The output is given below.

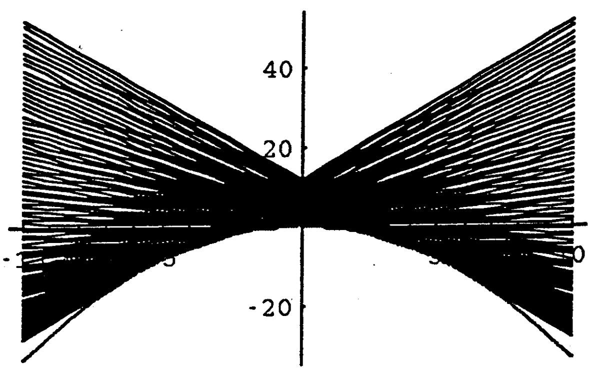


Fig. 2.1 The family in Example 2.1 and its envelope

This paper mainly deals with the envelopes of special families of plane curves, namely the ones bounding a fixed area in the first quadrant. We will investigate such families in the following section.

3. Envelopes of families of plane curves bounding a fixed area in the first quadrant

(a) Consider a family of straight lines bounding a fixed area k between the positive x and y-axes and the line itself.

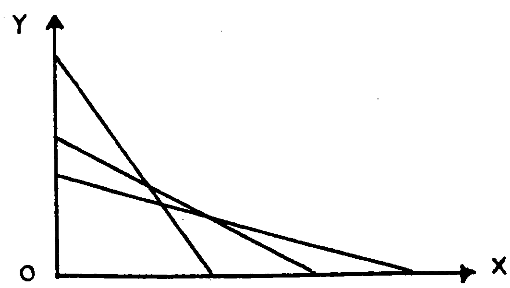


Fig. 3.1 A family of straight lines bounding a fixed area in the first quadrant

Let $A = (t,0)$ and $B = (0,s)$ be the x and y-intercepts of such a line where s and t are positive parameters. Let $O = (0,0)$ be the origin. Then since the area of the triangle OAB is equal to the constant k, one has the relationship $st = 2k$. Using this, one can easily see that this family of straight lines can be described by the single-parameter equation

$$y = -\frac{s^2}{2k}x + s \tag{3.1}$$

where s is a positive parameter and $x \geq 0$. To find the envelope of the family given by equation (3.1), one eliminates the s variable between the equation (3.1) and the following equation obtained by partially differentiating (3.1) with respect to s:

$$0 = -\frac{s}{k}x + 1 \tag{3.2}$$

In the process, one can also obtain a parametrization of the envelope

$$x = \frac{k}{s}, \quad y = \frac{s}{2} \tag{3.3}$$

The equation of the envelope is given by

$$y = \frac{k}{2}\cdot\frac{1}{x} \tag{3.4}$$

where $x \geq 0$. Therefore, the envelope of the family of curves given by equation (3.1) is a branch of a hyperbola in the first quadrant. One can write a *Mathematica* program similar to Program 2.1 to generate the following graph of the family (3.1) and its envelope for $k = 1$.

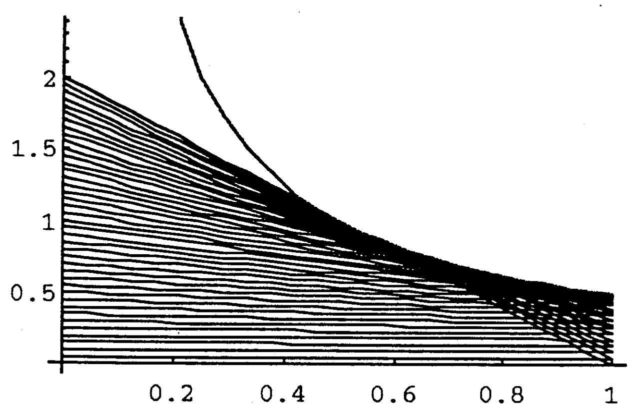


Fig. 3.2 The family (3.1) and its envelope

Another effective way to present the theory of envelopes to the beginning students is by means of computer animations. *Mathematica* is an ideal tool to create such animations. For instance, each frame of the following *Mathematica* animation consists of the graphs of the envelope and one member of the family (3.1) corresponding to a specific value of the parameter s. As the animation is run, the parameter s gradually changes, and one can see how each member of the family touches the envelope. One can effectively use these techniques to convey the intuitive meaning of the envelope of a given family of curves to the beginner.

Program 3.1

```
k=1; Do[Plot[{-s^2*x/(2k)+s,k/(2x)},{x,0,1},
PlotStyle->{{ }, {RGBColor[1,0,0], Thickness[1/95]}},
PlotRange->{0,2}],{s,0,2,0.05}]
```

A few frames of the animation are given below.

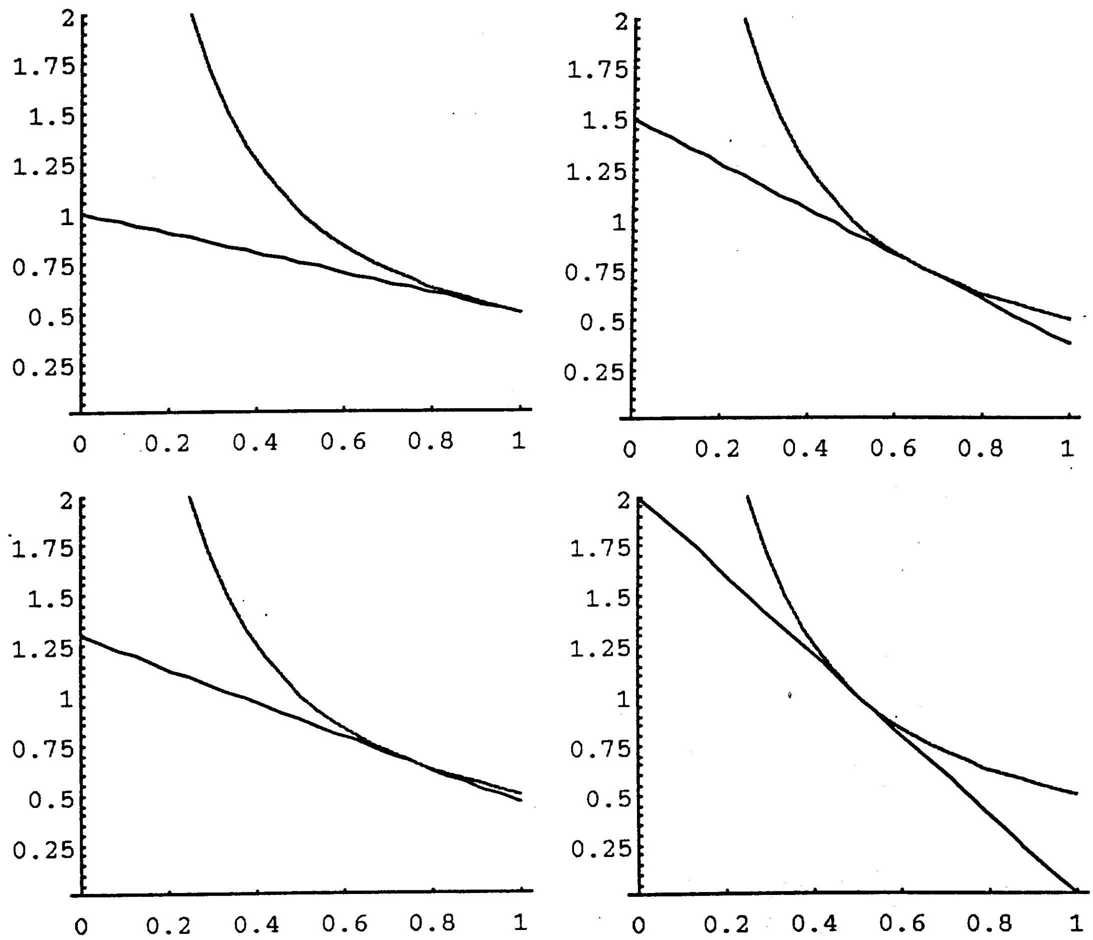


Fig. 3.3 An animation of the members of the family (3.1)

(b) Consider a family of parabolas opening downward with vertex on the y-axis, bounding a constant area in the first quadrant. If $A=(t,0)$ and $B=(0,s)$ are its x and y-intercepts respectively, where s and t are positive parameters, then it is not hard to see that the equation of the family is given by

$$y = -\frac{s}{t^2}x^2 + s \qquad (3.5)$$

However, since the area bounded in the first quadrant by the members of this family is equal to the constant k, one can obtain

$$k = \int_0^t \left(-\frac{s}{t^2}x^2 + s\right)dx = \frac{2st}{3} \tag{3.6}$$

Using equation (3.6), one can rewrite the equation of our family of parabolas (3.5) in terms of a single parameter:

$$y = -\frac{4s^3}{9k^2}x^2 + s \tag{3.7}$$

Partially differentiate equation (3.7) to obtain

$$0 = -\frac{4s^2x^2}{3k^2} + 1 \tag{3.8}$$

The above two equations imply that a parametrization of the envelope of the family (3.7) is given by

$$x = \frac{k\sqrt{3}}{2s}, \qquad y = \frac{2s}{3} \tag{3.9}$$

The envelope is given by

$$y = \frac{k}{\sqrt{3}}\cdot\frac{1}{x} \tag{3.10}$$

where $x \geq 0$, which is again a branch of a hyperbola in the first quadrant. A *Mathematica* generated output is given below for $k = 1$.

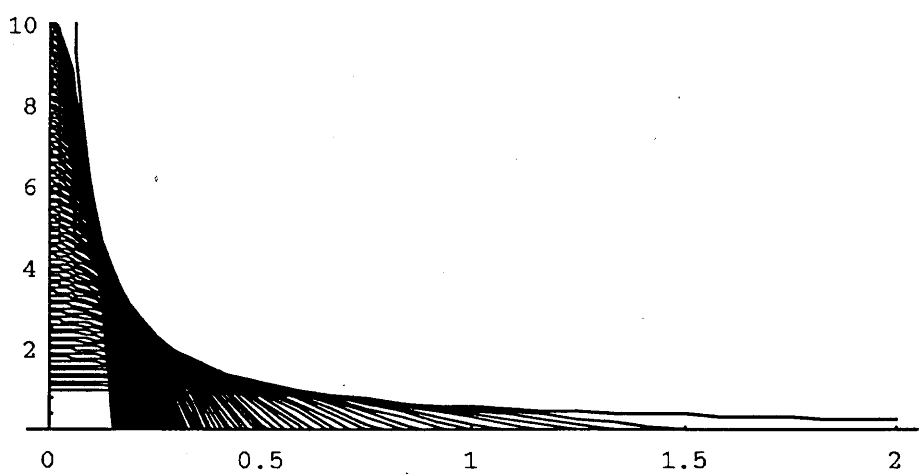


Fig. 3.4 The family (3.7) and its envelope

Judging from (a) and (b) one might wonder why the envelope was a branch of a hyperbola in both instances. One can use *Mathematica* to experiment with higher degree families to see their envelopes! In this way, we were able to obtain a conjecture, and then prove it to obtain the following theorem.

Theorem 3.1 Consider the family of plane curves given by $y = -sx^n/t^n + s$, where n is a fixed positive integer, s and t are positive parameters and $x \geq 0$. Suppose this family bounds a fixed area k in the first quadrant. Then the envelope of the family is a branch of a hyperbola.

Proof. Consider the family of plane curves

$$y = -\frac{s}{t^n} x^n + s \tag{3.11}$$

as given in the statement of the theorem. It is clear that its x and y-intercepts are respectively $(t,0)$ and $(0,s)$. Since the members of this family bound a constant area k in the first quadrant, one obtains

$$k = \int_0^t \left(-\frac{s}{t^n} x^n + s\right) dx = \frac{stn}{n+1} \tag{3.12}$$

Using the above equation, one can rewrite the family (3.11) in terms of the single parameter s.

$$y = -\frac{n^n s^{n+1}}{k^n (n+1)^n} x^n + s \tag{3.13}$$

Partially differentiate the above equation with respect to s to obtain

$$-\frac{n^n s^n}{k^n (n+1)^{n-1}} x^n + 1 = 0 \tag{3.14}$$

Then eliminate the s variable between the equations (3.13) and (3.14) to obtain the following parametrization of the envelope.

$$x = \frac{k}{ns} \sqrt[n]{(n+1)^{n-1}}$$
$$y = \frac{sn}{n+1} \tag{3.15}$$

Therefore, the equation of the envelope is given by

$$y = \frac{k \sqrt[n]{(n+1)^{n-1}}}{n+1} \cdot \frac{1}{x} \tag{3.16}$$

where $x \geq 0$. Hence the envelope is a branch of a hyperbola in the first quadrant as claimed. □

Remark. The above Theorem 3.1 generalizes the results of **(a)** and **(b)**.

Let us now consider a family of curves not fitting the description given by equation (3.11). How about ellipses?

(c) Consider a family of ellipses with major and minor axes along the coordinate axes, bounding a constant area k in the first quadrant. If $A = (t,0)$ and $B = (0,s)$ are its x and y-intercepts respectively, where s and t are positive parameters, then the equation of the family is given by

$$y = \frac{s}{t}\sqrt{t^2 - x^2} \tag{3.17}$$

where $0 \leq x \leq t$. However, since the members of this family bound a constant area k in the first quadrant, one obtains

$$k = \int_0^t \frac{s}{t}\sqrt{t^2 - x^2}\, dx = \frac{\pi s t}{4} \qquad (3.18)$$

Hence the family (3.17) can be written in terms of a single parameter t:

$$\pi y t^2 = 4k\sqrt{t^2 - x^2} \qquad (3.19)$$

To find the envelope of this family, differentiate equation (3.19) partially with respect to t and simplify to obtain

$$y = \frac{2k}{\pi\sqrt{t^2 - x^2}} \qquad (3.20)$$

Eliminate the t variable between the equations (3.19) and (3.20) to obtain the following equation of the envelope where $x \geq 0$:

$$y = \frac{2k}{\pi} \cdot \frac{1}{x} \qquad (3.21)$$

Therefore, the envelope of the family (3.19) is again a branch of a hyperbola. A *Mathematica* generated output is given below for $k = 1$.

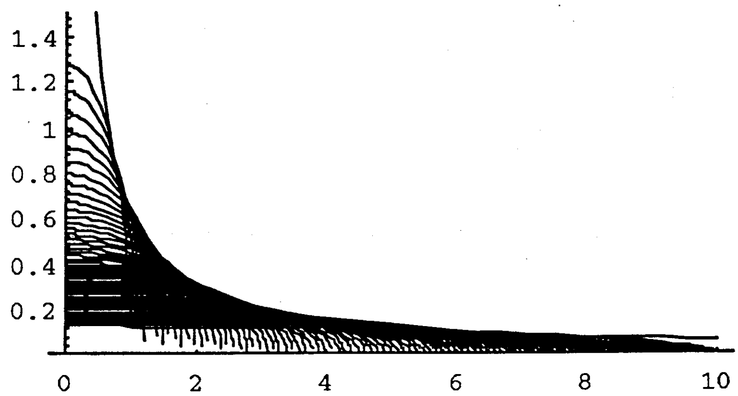


Fig. 3.5 The family (3.19) and its envelope

The curious student must be surprised at the outcome of the (a), (b) and (c). In each case, we were dealing with three different types of families, straight lines, parabolas and ellipses. However, their envelope was a branch of a hyperbola in each case. This is a good place to experiment with more families of plane curves bounding a constant area in the first quadrant.

(d) Consider the family of plane curves given by $y = s\left(1 - \dfrac{x^2 + x}{t^2 + t}\right)$ where s and t are positive parameters and $x \geq 0$. It is clear that $(t,0)$ and $(0,s)$ are the x and y-intercepts of each member of this family respectively. Suppose that this family bounds a constant area k in the first quadrant. This yields

$$k = \int_0^t s\left(1 - \frac{x^2 + x}{t^2 + t}\right)dx = \frac{s(4t^2 + 3t)}{6(t + 1)} \qquad (3.22)$$

Using this relationship between s and t, one can rewrite the equation of the family in terms of the single parameter t as follows:

$$y(4t^2+3t) = 6k(t+1) - \frac{6k(x^2+x)}{t} \qquad (3.23)$$

Differentiate equation (3.23) partially with respect to t to obtain

$$y(8t+3) = 6k + \frac{6k(x^2+x)}{t^2} \qquad (3.24)$$

The equations (3.23) and (3.24) imply that a parametrization for the envelope is given by

$$x = \frac{-3+\sqrt{12t^2+18t+9}}{6}$$
$$y = \frac{k}{t} \qquad (3.25)$$

One can use the *Mathematica* to graph the family given by equation (3.23) and its envelope given by (3.25) to obtain the following for $k=1$.

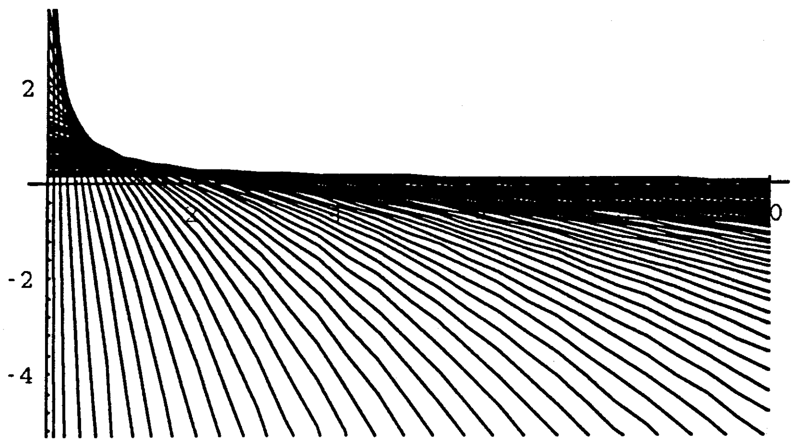


Fig. 3.6 The family (3.23) and its envelope

The envelope appears to be a branch of a hyperbola, though it is really not the case. One can eliminate the t variable from the two equations given in (3.25) to realize this. In the following, we will generalize the type of family considered in (d).

Theorem 3.2 Suppose ϕ is a one-to-one and a continuous function such that $\phi(x) \neq 0$ and $\phi'(x) \neq 0$ for all $x > 0$. Consider the family of plane curves given by $y = s(1-\phi(x)/\phi(t))$ where s and t are positive parameters and $x \geq 0$. Suppose further that this family bounds a constant area k in the first quadrant. Then the envelope of the family is given by the equation $y \int_0^{k/y} \phi(x)dx = k\phi(x)$.

Proof. Consider the family of plane curves

$$y = s\left(1 - \frac{\phi(x)}{\phi(t)}\right) \qquad (3.26)$$

where s and t are positive parameters and $x \geq 0$. It is quite clear that $(t,0)$ and $(0,s)$ are the x and y-intercepts of each member of this family respectively. Since each member of the family bounds a constant area k in and $(0,s)$ are the x and y-intercepts of each member of this family respectively. Since each member of the family bounds a constant area k in the first quadrant, one obtains

$$k = \int_0^t s\left(1 - \frac{\phi(x)}{\phi(t)}\right) dx = \frac{s(t\phi(t) - \int_0^t \phi(x)\,dx)}{\phi(t)} \qquad (3.27)$$

Using this relationship between s and t, one can rewrite the family (3.26) in terms of a single parameter t as follows.

$$y\,(\,t\phi(t) - \int_0^t \phi(x)dx\,) - k\phi(t) + k\phi(x) = 0 \qquad (3.28)$$

Differentiate equation (3.28) partially with respect to t to obtain

$$(yt - k)\phi'(t) = 0 \qquad (3.29)$$

Equations (3.28) and (3.29) imply that a parametrization for the envelope is given by

$$\begin{aligned} x &= \phi^{-1}\left(\int_0^t \phi(x)dx/\,t\right) \\ y &= k\,/\,t \end{aligned} \qquad (3.30)$$

The above equation implies that the equation of the envelope is given by

$$y\int_0^{k/y} \phi(x)dx = k\phi(x). \qquad (3.31)$$

Hence the theorem. □

We are now in a position to raise an important question. For what class of functions ϕ does the equation (3.31) represent a branch of a hyperbola? The author has more information on this question. However, due to lack of space it will not be included in this paper.

References:

[1] Baranenkov, G. et al, Problems in Mathematical Analysis, Mir Publishers, Moscow, 1964.
[2] Barnard, S. and Child, M., Elements of Geometry, Macmillan and Co., London, 1914.
[3] Bruce, J. and Giblin, P., Curves and Singularities, Cambridge University Press, Cambridge, 1984.
[4] Cox, D., Little, J., and O'Shea, D., Ideals, Varieties and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra, Springer-Verlag, New York, 1992.
[5] Loney, L., The Elements of Coordinate Geometry, Macmillan and Co., London, 1895.
[6] Maeder, R., Programming in Mathematica, Addison-Wesley Publishing Co., Redwood City, CA, 1991.
[7] Wagon, S., Mathematica in Action, W. H. Freeman and Co., New York, 1991.
[8] Wolfram, S., Mathematica: A System for Doing Mathematics by Computer, Addison-Wesley Publishing Co., Redwood City, CA, 1988.

GRAPHING CALCULATORS AND THE LEARNER

Judith Aronow
Independent Consultant
5590 Frost; Beaumont, TX 77706 U.S.A.
e-mail: aronowju@hal.lamar.edu

Calculators were used in the classroom as early as the 60's, and then with accelerating interest, despite claims that fundamental skills may be lost by the student. Now, calculators are required on the free response section and some problems on the multiple choice section of the Advanced Placement examination in mathematics. The College Board is allowing optional use of calculators on the Mathematics Level IIc of the Scholastic Assessment Test (SAT I). Curricular changes will follow. Which topics should be kept and which dropped? Does having a calculator which can do definite integration mean dropping the topics on techniques of integration? Does being able to use a calculator to find roots graphically to any desired degree of accuracy mean that learning synthetic division to apply the Factor Theorem is unnecessary?

Experience in giving workshops to teachers on the use of calculators in the classroom shows that the calculator can be used to teach many traditional topics in pre-calculus and calculus. Subsequent use of calculator algorithms and graphical representations at first can emphasize these concepts and gradually be used as a tool for later courses, in much the same way as a sub-routine is called by a program.

I <u>Quadratic formula and graphical representation</u>:

When teaching the quadratic formula, we introduce the notion of completing the square first, and then using A,B and C for the coefficients of the quadratic we wish to solve. We move terms around and eventually are able to solve for x in terms of those coefficients. Often students aren't too sure when and why to use the formula, because the chapter on roots and the chapter on parabolas are somewhere else and to follow at some time lapse. We sometimes have the student memorize the consequences of the discriminant D being equal, less than or greater than zero, but the student often is puzzled by the analysis.

The quadratic formula can be developed as always, with the graph brought in to support the analysis a slight abridgement of lessons as they are usually arranged in the algebra or pre-calculus textbook.

Using Quad2*[1] can be a game. As the student enters values for A, B, and H interactively, s/he can try to pick values which make the discriminant negative,

---

[1]All programs as marked by (*) may be found in the Appendix

for example; or the student can try to predict the roots.

An exercise inducing deep understanding of the quadratic formula, would be to have the student develop a program like Quad2* after demonstrating its use on the overhead, but before giving it to the student to use.

Notice that Quad2* uses Grquad* as a subroutine. Grquad* can be used alone to graph a quadratic (one must delete the command "Return") and Quad2* may be used to find the roots (one must add the command "Prompt A, B, H" at the beginning. Note too, that the demonstration emphasizes the importance of the graph in visualizing the roots of the quadratic as the x-intercepts of the parabola. This observation used to take many lessons years ago.

II <u>Roots of Continuous Functions and graphical representation</u>:

<u>Synthetic Division</u>:
The program Syn* does synthetic division. The student inputs the coefficients of a polynomial and the candidate rational root, and the program calculates the remainder, which if zero, then gives the coefficients of the reduced equation. One can illustrate the function graphically to gain insight into the roots, to begin the process. The theorems on bounds for roots can be illustrated by using the program Syn* and the graph of the polynomial.

<u>Bisection Method with Polynomial Functions:</u>
The program Syn* can be used to find $p(x_1)$ where x_1 is the candidate root. $p(x_i)$ is just the last number in the line indicated by the operation $p(x)/(x-x_1)$. In this method, we locate an interval $a \leq x_1 \leq b$ such that p(a) and p(b) have opposite signs. We pick x_1 to be halfway between a and b. Select the new interval to be either $[a,x_1]$ or $[x_1,b]$ so that $p(x_1)$ is opposite in sign to either p(a) or p(b). Select the midpoint of the selected interval as x_2 and continue subdividing intervals until the root is approximated to any prescribed number of decimal places.

The student can use Syn* as a subroutine for a program which achieves the Bisection Method. Creation of the program can be a distinct aid in understanding the algorithm.

<u>Finding Roots of Non-Polynomial Functions</u>:
The bisection method can be used to find the roots of any continuous function. Instead of using Syn* in creating a program, one can create a short program which evaluates $f(x_i)$, and call it in a program Bisection*. On some calculators (HP-48 and TI-85), evaluation of a function for a particular value, is a built-in capability. A meaningful way of doing the bisection method in a classroom setting, is to keep a written list of the left and right endpoints and midpoint of the interval containing functional values differing in sign, and do this for each step. The written list gives alot of insight into the process. ·

Newton's Method for Finding Roots:
A graphing programmable calculator, actually, in short, a computer, is well suited to using Newton's method for finding roots of a differentiable function, f(x). A short program calculates the next root $x_{n+1} = x_n - f(x_n)/f'(x_n)$, given the previous iterate, x_n and a starting value, x_0. The function and its derivative are stored as functions to be graphed . A look at the graph before beginning the iterative process gives one a good idea of where to begin the process which should be fairly close to the suspected root. One then invokes the interactive program using as input, one's initial guess. On the TI-85, for the program NEWTON*, the output appears on the Home Screen. Press [ENTER] to invoke the program again. When it requests the next input, press [2nd][(-)] or [ANS]; i.e. the answer from the previous iteration, and press [ENTER] again. The next iterate appears. Any number of decimal points needed, can be set by [MODE] ([2nd][MORE]). When the outputs no longer differ, the solution has been found. Using this method, allows one to see the outputs as they steadily march to the solution. Newton's method is the method used by many calculators for root finding because it converges so quickly with an initial good guess.

Use of Newton's Method has more advantages in the classroom than that of finding roots of differentiable functions . The method can be used to illustrate linearization of the function; use of the tangent line to approximate the function at a point. We can also let Newton's Method serve as a start for a discussion of fixed points, pointing out that Newton's Method is a particular method for finding a fixed point $f(x_i) = x_i$ of f(x).

III Differentiation:

Symmetric Difference Quotients:

The symmetric difference quotient is used on many graphing calculators as a numerical approximation to the derivative of f(x). The difference quotient

$$f'(x') \approx \frac{f(x+h) - f(x-h)}{2h}, \text{ h small}$$

can often provide an amazingly close approximation to the derivative, both graphically and numerically, particularly for very small h. (One must trade off speed with accuracy sometimes. The smaller h is , the longer it takes to calculate the difference.) An interesting activity is to graph the function with the
derivative, using the derivative capability of the calculator, and then compare graphical derivative with that obtained with the symmetric difference quotient.

Graphing technology is a willing partner in finding local maximum and minimum points, inflection points, the general shape of a graph. The student must still verify analytically that the function is differentiable.

IV <u>Integration</u>:

There are two ways to find definite integrals:

1. Use the Second Fundamental Theorem of Integral Calculus: use the symbolic indefinite integral, $F(x)$, and substitute the limits a and b into that solution, obtaining $A = F(b) - F(a)$.

2. Use numerical techniques for integration.

Much graphing technology can return a closed form solution to finding an indefinite integral. However, since a vast number of functions have no closed form antiderivative, it makes good sense to learn numerical techniques. As teachers, we are also freed from the constraint that all problem sets contain artificially "easy" integrands. On the other hand, we are perfectly justified in teaching "techniques of integration" for those indefinite integrals which do have a closed form solution.

Exercises using graphing technology that emphasize understanding of the First Fundamental Theorem of Calculus can be extremely useful for a student. The Advanced Placement exams in mathematics particularly use questions relating to it.

The student may be asked to find $A(c)$ or $A'(c)$ for some c, $a \le c \le b$ if

$$A(x) = \int_a^x f(t)\,dt$$

and only a graph of f is given for $a \le x \le b$.

Another exercise might be:

Given

$$A(1.3) = \int_1^{1.3} \cos x^2\,dx$$

Verify that

$$A'(x) = \cos x^2 .$$

We could find

$$K(b) = \frac{A(b+h) - A(b-h)}{2h}$$

(the symmetric difference quotient for numerically finding a derivative) using a small value of h, say 0.1, and do this for each of the thirteen values of b (.1, 0.2,0.3....1.3) if the increment used is 0.1. These calculations would be found using the calculator's or software's numerical integrator. The calculations could be stored in a list for later graphing.

To do this:

1. **Dimension a list, L1, say,**
   13 [DimL] L1 for 13 elements
2. **Calculate**
   fnint(cos x^2, x, .1, (L1(1)-L1(0)/.2)

   .
   .
   .

   fnint(cos x^2, x, .1, (L1(13)-L1(12))/1.3)
   (12 of these)
3. **Plot the points (b_i $A(b_i)$) as follows:**
   Pt(On) (.1, L1(1)

   .
   .
   .

   Pt(On) (1.3, L1(1.3))

The graph of the dots, K should resemble that of
f(x) - cos x^2 .
We should then be able to graph the points (b, K(b)). The graph should resemble f(x) = cos x^2.

APPENDIX

Programs used are for the TI-85:

PROGRAM: GRQUAD
```
:ClDrw
:Prompt A,B,H
:FC=A*x^2+B*x+H
:y1=FC
:DispG
:Return
```

PROGRAM: QUAD2
```
:GRQUAD
:D=B^2-4A*H
:If D<0
:Then
:Disp "IMAG ROOTS"
:Else
:E=((-B+√D)/2A)
:F=((-B-√D)/2A)
:End
:Disp E,F
```

PROGRAM: SYN
```
:Prompt M,R
:M→dim A
:For(N,1,M,1)
:Prompt P
:P→A(N)
:Disp A(N)
:End
:For(N,2,M,1)
:(A(N)+A(N-1)*R)→A(N)
:Disp A(N)
:End
:Disp A
```
[To be able to scroll through A, invoke it on the Home Screen, then use ▸]

PROGRAM: NEWTON (for a function stored in y2 and its derivative stored in y3)
```
:Prompt A
:evalF(y2,x,A)→G
:evalF(y3,x,A)→H
:A-(G/H)→NP
:Disp NP
```

.STATISTICS LAB-123

USING ELECTRONIC SPREADSHEET

Tower Chen, University of Guam

Statistics is the science of collecting and interpreting data. There are four elements common to statistical problems: defining the population of interest, selecting a random sample from this population, making inference about the population based on this random sample, and measuring the reliability of the inference based on the sampling distribution of samples.

It is helpful to use computer to solve statistics problems by its calculating facility. If we don't understand a random sample selected, applied formula, and calculating procedure, it will be very hard for us to interpret the result. It is very important for statisticians to know how to interpret the result.

We can classify statistics problems into different types, there is the same procedure for problems of the same type. When data are plugged into the formula, each data is processed as the same way as the other. This type of way to process data should be handled by a computer program with "tabling facility" instead of "listing facility". After entering data, you will get results without showing steps of calculation. Even some new statistics programs use "tabling facility" only for entering data and does not showing steps of calculation from data. For education purpose, students does not gain too much by using this kind of statistics program. After entering data, students only can get result but not steps. They can not compare their manual steps with computer steps, if they make wrong caculation.

There is a new computer tool having "tabling facility" which is called a spreadsheet. The most popular spreadsheet is Lotus-123. The special facility and the advantange of Lotal-123 spread sheet are:

1. Spreadsheet use the location of a block to represent a mathematical variable and the relationship among blocks to represent a formula.

2. There are two layers in the program of spreadsheet because a hidden formula is on the bottom layer of a block and the result is on the top layer of a block. There is only one layer in a program of traditional computer languages (Basic, Fortran,C).

3. "Two layers" facility can help us easily arrange data and results in "tabling form" which is more readable layout than "listing form".

4. "Hidden" formulas of the bottom layer do not show on the spreadsheet program, but if there is need for us to check a formula we are able to call it out instantly.

5. It is easy to increase or decrease the value of data to see how a result changed immediately without re-compiling program. This kind of facility of spreadsheet is very helpful for educational purpose.

6. The result of any step can be displayed in the spreadsheet to compare with the result we have done manually. It is very helpful for learning purpose.

7. It is easy to change data and also to modify some parts of the program for similar problems. To be able to positively interact with the program is the key to keep interesting in this subject.

8. The graphing facility of spreadsheet can help us to describe data and results in graph easily. My programs are developed by taking the advantage of "tabling tacility" of this tool. From the spreadsheet students can see in "Statistics Lab-123". there are twelve labs: Lab 1: Using Graphs to Describe Data

Lab 2: Using Numbers to Describe Data

Lab 3: Calculating the Means & the Variance of the Sampling Distribution of Means

Lab 4: Making Inference about μ from a Random Sample Mean Which Sampling Distribution Is Normal

Lab 5: Making Inference about P from a Random Sample Proportic Which Sampling Distribution Is Binomial

Lab 6: Making Inference about (u_1-u_2) from Two Independent Large-Samples

Lab 7: Making Inference about (P_1-P_2) from Two Independent Large-Samples

Lab 8: Making Inference about u(d) from Pair of ependent Small-Samples

Lab 9: Find the Linear Regression & Its Properties

Lab 10: Making Inference about u_1, u_2 ··· u_k from Many Independent Large-Samples

Lab 11: Making Inference about P from a Random Sample Multinomial Proportions

Lab 12: Making Inference about the Distributions of Populations by Nonparametric Test

Each lab is related to one statistics concept to solve one type of problems. These twelve labs covers most basic and applied staistics problems. All labs is designed similar to physics labs starting with purpose following by introduction and theory. Each lab is given one application to show how this program works. From the spreadsheet student can see clearly how each block work and how formulas integrate with blocks. Students have chance interact with program, are able to see each step of procedures and practice theory and formulas by using "Statistics Lab-123" program.

The dis-advantage of "Statistics Lab-123" is that there is a need to search a constant from a table to continue the program. On the other hand, this kind of interacting with the program will test our understanding the concept.

"Statistics Lab-123" is the program of "what you see is what you get". It is not only are suitable for students to use but also very helpful for professional use.

AUTOMATED GEOMETRY THEOREM PROVING AND GEOMETRY EDUCATION †

Shang-Ching Chou, Xiao-Shan Gao[1], Jing-Zhong Zhang[2],

Department of Computer Science, The Wichita State University
Wichita KS 67260-0083, USA
e-mail: (chou,gao,zhang)@cs.twsu.edu

Abstract. We give a brief introduction to the area method which has been used to produce short and human-readable proofs for hundreds of geometry theorems. We also report a computer program, GE (Geometry Expert), which implements the area method and provides sophisticated graphical interface.

1 The Area Method

Although René Descartes' great work on the algebraization of geometry via the notion of a coordinate system was regarded as a universal method of dealing with geometry problems, the precise method of mechanical geometry theorem proving along the way paved by Descartes was discover much later. In the 1930s, A. Tarski gave a decision procedure for the elementary Euclidean geometry [10]. In 1978, Wen–Tsün Wu gave an algebraic method, which for the first time was used to prove hundreds of difficult geometry theorems [12, 13, 2]. In essence, all the above algebraic methods can be characterized as to "trade qualitative difficulty for quantitative complexity" (words of H. Wang [11]), i.e., the proofs are carried out by massive algebraic computation. Proofs thus produced generally lose the elegant manner of the traditional geometry proofs. On the other hand, the artificial intelligence approaches to automated geometry theorem proving [1, 6, 9], though have produced human-readable proofs for many relatively simple geometry theorems, are still unable to prove most of the moderately difficult geometry theorems. By *human-readable proofs*, we mean the proofs are short enough for people to repeat with pencil and paper easily and each step of the proofs has clear geometric meaning.

In this paper, we introduce the area method which is capable of producing human-readable proofs for difficult geometry theorems efficiently. The area method extends the idea of eliminating variables to the idea of eliminating points and is complete for a class of constructive geometry

†This work was supported in part by the NSF Grant CCR-9117870 and a Grant from Chinese NSF.

[1]On leave from Institute of Systems Sciences, Academia Sinica.

[2]On leave from Chengdu Institute of Computer Applications, Academia Sinica and Department of Math., Gungzhou Normal University, 510400 Gungzhou, China.

statements. Our computer program GE (Geometry Expert), which implements the area method, for the first time has produced elegant, short (sometimes even shorter than those given by geometry experts), and human-readable proofs of over 400 difficult theorems [3]. The proofs are in shapes that a student of mathematics could reproduce with pencil and paper easily. This feature makes the area method a promising new approach to the teaching of geometry.

To further illustrate the usefulness of the method (and our program), we have searched the Problem Sections and articles in *American Mathematical Monthly.* We have selected 135 geometry problems and used our computer program to produce solutions for them [5]. Most of the solutions are very short.

We will give a brief introduction to the key parts of the area method. The complete description of the area method and its extensions to the *volume method*, the *full-angle method*, and the method based on *vectors and complex numbers* can be found in [3].

1.1 Signed Area. The *signed area* S_{ABC} of triangle ABC is the usual area with a sign depending on the order of the three vertices of the triangle. In our area method, the signed area is used as a basic (undefined) geometry quantity which satisfies the following properties.

Proposition 1.1 (1) $S_{ABC} = S_{CAB} = S_{BCA} = -S_{BAC} = -S_{CBA} = -S_{ACB}$.

(2) *Points $A, B,$ and C are collinear iff $S_{ABC} = 0$.*

(3) *For any four points A, B, C, and D, we have $S_{ABC} = S_{ABD} + S_{ADC} + S_{DBC}$.*

(4) (**The Co-side Theorem**) *Let M be the intersection of two non-parallel lines AB and PQ and $Q \neq M$. Then $\frac{\overline{PM}}{\overline{QM}} = \frac{S_{PAB}}{S_{QAB}}, \frac{\overline{PM}}{\overline{PQ}} = \frac{S_{PAB}}{S_{PAQB}}, \frac{\overline{QM}}{\overline{PQ}} = \frac{S_{QAB}}{S_{PAQB}}$, where $S_{PAQB} = S_{PAB} - S_{QAB}$ is the signed area of a quadrilateral $PAQB$.*

Proposition 1.2 *Let R be a point on line PQ. Then for any two points A and B, $S_{RAB} = \frac{\overline{PR}}{\overline{PQ}}S_{QAB} + \frac{\overline{RQ}}{\overline{PQ}}S_{PAB}$.*

Proof. By the co-side theorem $\frac{S_{RAP}}{S_{QAP}} = \frac{\overline{PR}}{\overline{PQ}}, \frac{S_{RPB}}{S_{QPB}} = \frac{\overline{PR}}{\overline{PQ}}$. By (3) of Proposition 1.1, $S_{RAB} = S_{PAB} + S_{RAP} + S_{RPB} = S_{PAB} + \frac{\overline{PR}}{\overline{PQ}}(S_{APQ} - S_{BPQ}) = S_{PAB} + \frac{\overline{PR}}{\overline{PQ}}(S_{APB} + S_{ABQ}) = (1 - \frac{\overline{PR}}{\overline{PQ}})S_{PAB} + \frac{\overline{PR}}{\overline{PQ}}S_{QAB} = \frac{\overline{PR}}{\overline{PQ}}S_{QAB} + \frac{\overline{RQ}}{\overline{PQ}}S_{PAB}$. ∎

We use the notation $AB \parallel CD$ to denote the fact that $A, B, C,$ and D satisfy one of the following conditions: (1) lines AB and CD are parallel; (2) $A = B$ or $C = D$; or (3) A, B, C and D are on the same line.

Proposition 1.3 *$PQ \parallel AB$ iff $S_{PAB} = S_{QAB}$, i.e., iff $S_{PAQB} = 0$.*

Proof. If $S_{PAB} \neq S_{QAB}$, it is possible to take a point O on line PQ such that $\frac{\overline{PO}}{\overline{PQ}} = \frac{S_{PAB}}{S_{PAQB}}$. Thus $\frac{\overline{OQ}}{\overline{PQ}} = -\frac{S_{QAB}}{S_{PAQB}}$. By Proposition 1.2, $S_{OAB} = \frac{\overline{PO}}{\overline{PQ}}S_{QAB} + \frac{\overline{OQ}}{\overline{PQ}}S_{PAB} = 0$. By (2) of Proposition

1.1, point O is also on line AB, i.e., AB is not parallel to line PQ. Conversely, if $PQ \not\parallel AB$ then AB and PQ intersect at a unique point O. By the co-side theorem, $\frac{\overline{OP}}{\overline{OQ}} = \frac{S_{PAB}}{S_{QAB}} = 1$. Thus $P = Q$ which is a contradiction. ∎

Proposition 1.4 *Let $ABCD$ be a parallelogram. Then for two points P and Q, we have $S_{APQ} + S_{CPQ} = S_{BPQ} + S_{DPQ}$ or $S_{APBQ} = S_{DPCQ}$;*

Proof. Let O be the intersection of AC and BD. Since O is the midpoint of AC, by Proposition 1.2, $S_{APQ} + S_{CPQ} = 2S_{OPQ}$. For the same reason, $S_{BPQ} + S_{DPQ} = 2S_{OPQ}$. ∎

We use the following simple example to show how the area method works.

Example 1.5 *Let O be the intersection of the two diagonals AC and BD of a parallelogram $ABCD$. Show that $\overline{AO} = \overline{OC}$, or $\frac{\overline{AO}}{\overline{OC}} = 1$.*

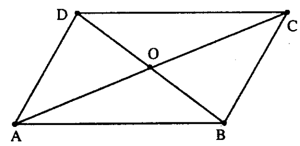


Figure 1

Proof. The proof is to eliminate points from the conclusion. By the co-side theorem, we can eliminate O: $\frac{\overline{AO}}{\overline{OC}} = \frac{S_{ABD}}{S_{BCD}}$. By Proposition 1.3, we can eliminate D: $S_{ABD} = S_{ABC}$, $S_{BCD} = S_{BCA}$. Thus we have the following proof: $\frac{\overline{AO}}{\overline{OC}} = \frac{S_{ABD}}{S_{BCD}} = \frac{S_{ABC}}{S_{BCA}} = \frac{S_{ABC}}{S_{ABC}} = 1$. ∎

1.2. Pythagorean Difference. For three points A, B, and C, the *Pythagorean difference*, P_{ABC}, is defined to be $P_{ABC} = \overline{AB}^2 + \overline{CB}^2 - \overline{AC}^2$. It is easy to check that (1) $P_{AAB} = 0$; $P_{ABC} = P_{CBA}$; (2) $P_{ABA} = 2\overline{AB}^2$; (3) If A, B, C are collinear, $P_{ABC} = 2\overline{BA} \cdot \overline{BC}$. For a quadrilateral $ABCD$, let $P_{ABCD} = P_{ABD} - P_{CBD} = \overline{AB}^2 + \overline{CD}^2 - \overline{BC}^2 - \overline{DA}^2$. The Pythagorean theorem is taken as a basic (unproven) property of the Pythagorean difference.

Proposition 1.6 (Pythagorean Theorem) *$AB \perp BC$ iff $P_{ABC} = \overline{AB}^2 + \overline{BC}^2 - \overline{AC}^2 = 0$.*

For four points A, B, C, and D, the notation $AB \perp CD$ implies that one of the following conditions is true: $A = B$, or $C = D$, or the line AB is perpendicular to line CD.

Proposition 1.7 *$AC \perp BD$ iff $P_{ABD} = P_{CBD}$ or $P_{ABCD} = 0$.*

Proof. Let M and N be the feet of the perpendiculars from points A and C to line BD. Then by the Pythagorean theorem, $P_{ABD} = \overline{AB}^2 + \overline{BD}^2 - \overline{AD}^2 = \overline{AM}^2 + \overline{BM}^2 + \overline{BD}^2 - \overline{AM}^2 - \overline{MD}^2 = \overline{BM}^2 + \overline{BD}^2 - (\overline{BD} - \overline{BM})^2 = 2\overline{BM} \cdot \overline{BD}$. Similarly $P_{CBD} = 2\overline{BN} \cdot \overline{BD}$. We have $P_{ABD} = P_{CBD}$ if and only if $\overline{MB} = \overline{NB}$, i.e., $M = N$ which is equivalent to $AC \perp BD$. ∎

Proposition 1.8 *Let R be a point on line PQ with position ratios $r_1 = \frac{\overline{PR}}{\overline{PQ}}, r_2 = \frac{\overline{RQ}}{\overline{PQ}}$. Then for points A, B, we have $P_{RAB} = r_1 P_{QAB} + r_2 P_{PAB}$ and $P_{ARB} = r_1 P_{AQB} + r_2 P_{APB} - r_1 r_2 P_{PQP}$.*

Proof. We first assume

$$\overline{RA}^2 = r_1\overline{QA}^2 + r_2\overline{PA}^2 - r_1r_2\overline{PQ}^2 \quad (1)$$
$$\overline{RB}^2 = r_1\overline{QB}^2 + r_2\overline{PB}^2 - r_1r_2\overline{PQ}^2. \quad (2)$$

Then $P_{RAB} = \overline{RA}^2 + \overline{AB}^2 - \overline{RB}^2 = r_1(\overline{QA}^2 + \overline{AB}^2 - \overline{QB}^2) + r_2(\overline{PA}^2 + \overline{AB}^2 - \overline{PB}^2) = r_1P_{QAB} + r_2P_{PAB}$. The second result can be proved similarly. To prove (1), let the foot of the perpendicular from A to PQ be A_1. By Proposition 1.7, $\frac{P_{APR}}{P_{APQ}} = \frac{P_{A_1PR}}{P_{A_1PQ}} = \frac{\overline{A_1P}\cdot\overline{PR}}{\overline{A_1P}\cdot\overline{PQ}} = \frac{\overline{PR}}{\overline{PQ}} = r_1$. Then $r_1\overline{QA}^2 + r_2\overline{PA}^2 - r_1r_2\overline{PQ}^2 = r_1\overline{QA}^2 + (1-r_1)\overline{PA}^2 - r_1(1-r_1)\overline{PQ}^2 = \overline{PA}^2 + r_1(\overline{QA}^2 - \overline{PA}^2 - \overline{PQ}) + r_1^2\overline{PQ}^2 = \overline{PA}^2 + \overline{PR}^2 - r_1P_{APQ} = \overline{PA}^2 + \overline{PR}^2 - P_{APR} = \overline{AR}^2$. ∎

Proposition 1.9 *Let AB and PQ be two non-perpendicular lines and Y be the intersection of line PQ and the line passing through A and perpendicular to AB. Then* $\frac{\overline{PY}}{\overline{QY}} = \frac{P_{PAB}}{P_{QAB}}, \frac{\overline{PY}}{\overline{PQ}} = \frac{P_{PAB}}{P_{PAQB}}, \frac{\overline{QY}}{\overline{PQ}} = \frac{P_{QAB}}{P_{PAQB}}$.

Proof. We only need to show the first equation. Let P_1 and Q_1 be the orthogonal projections from P and Q to line AB respectively. By Proposition 1.7, $\frac{P_{PAB}}{P_{QAB}} = \frac{P_{P_1AB}}{P_{Q_1AB}} = \frac{\overline{AP_1}\cdot\overline{AB}}{\overline{AQ_1}\cdot\overline{AB}} = \frac{\overline{AP_1}}{\overline{AQ_1}} = \frac{\overline{PY}}{\overline{QY}}$. ∎

Proposition 1.10 *Let $ABCD$ be a parallelogram. Then for any points P and Q, we have $P_{APQ} + P_{CPQ} = P_{BPQ} + P_{DPQ}$ or $P_{APBQ} = P_{DPCQ}$.*

Proof. This proposition can be proved similarly to Proposition 1.4. ∎

Example 1.11 (The Orthocenter Theorem) *Let the two altitudes AF and BE of triangle ABC meet in H. Show that $CH \perp AB$.*

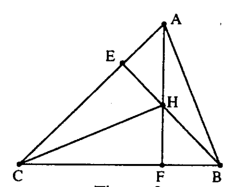


Figure 2

Proof. By Proposition 1.7, we need only to show $P_{ACH} = P_{BCH}$. Since $BH \perp AC$ and $AH \perp BC$, by Proposition 1.7 we can eliminate point H: $P_{ACH} = P_{ACB}$; $P_{BCH} = P_{BCA}$. Thus we have the following proof. $\frac{P_{ACH}}{P_{BCH}} = \frac{P_{ACB}}{P_{BCA}} = \frac{P_{ACB}}{P_{ACB}} = 1$. ∎

1.3 Constructive Geometry Statements. The area method is complete for the class of *constructive geometry theorems* which is actually certain statements about geometric configurations that can be drawn with a ruler and a compass.

By *geometry quantities*, we mean ratios, signed areas, or Pythagorean differences. A *construction* is one of the following ways of introducing new points from old ones.

C1 (POINTS $Y_1, \cdots, Y_m$) Take arbitrary points $Y_1, \cdots, Y_m$ in the plane. Y_i are free points.

C2 (ON Y *ob*). Take a point Y on the geometric object *ob*. A geometric object could be a line or a circle.

We have four kind of lines: (LINE U V) is the line passing through U and V; (PLINE W U V) is the line passing through W and parallel to (LINE U V); (TLINE W U V)

is the line passing through point W and perpendicular to (LINE U V); (BLINE U V) is the perpendicular-bisector of UV. For each of the above lines, we introduce the *non-degenerate (ndg) condition*: $U \neq V$. A *circle* with point O as its center and passing through a point U is denoted by (CIR O U).

C3 (INTER Y $ob1$ $ob2$). Point Y is the intersection of object $ob1$ and object $ob2$. The ndg condition is that $ob1$ and $ob2$ should have a normal intersection. More details may be found in [3].

C4 (PRATIO Y R P Q λ) Take a point Y on the line passing through R and parallel to PQ such that $\overline{RY} = \lambda\overline{PQ}$ where λ can be a rational number, a variable, or an expression in geometry quantities. The nondegenerate condition is $P \neq Q$.

C5 (TRATIO Y U V r). Take a point Y on line (TLINE U U V) such that $r = \frac{4S_{UVY}}{P_{UVU}} (= \frac{UY}{UV})$, where r can be a rational number, a rational expression in geometric quantities, or a variable. The ndg condition is $U \neq V$.

Since there are four kinds of lines, constructions C2 and C3 have 4 and 15 possible forms respectively. Thus, we have 22 different forms of constructions. The 22 constructions are not independent to each other. We now introduce a minimal set of constructions which are equivalent to all the 22 constructions but much few in number.

A *minimal set of constructions* consists of C4, C5, and the following construction.

C6 (INTER Y (LINE P Q) (LINE U V)) Take the intersection Y of line PQ and line UV. The nondegenerate condition is $PQ \not\parallel UV$.

It is not difficult to prove that this minimal set of constructions is equivalent to all the 22 constructions.

A *constructive geometry statement* can be represented by a list $S = (C_1, C_2, \ldots, C_k, G)$ where each construction C_i introduces a new point from the points introduced by the previous C_j, $j = 1, \cdots i-1$, and G, the conclusion of the statement, is either a geometry predicate like collinear, parallel, perpendicular, or an equation $E_1 = E_2$ of polynomials in geometric quantities about the points introduced by the constructions C_i.

Example 1.12 *The following is a constructive description for Example 1.5: ((POINTS A B C) (INTER D (PLINE A B C) (PLINE C A B)) (INTER O (LINE A C) (LINE B D)) (MIDPOINT O A C)). The ndg condition is that A, B, and C are not collinear.*

The following is a constructive description for Example 1.11: ((POINTS A B C) (INTER E (LINE A C) (TLINE B A C)) (INTER F (LINE B C) (TLINE A B C)) (INTER H (LINE B E) (LINE A F)) (PERPENDICULAR C H A B)). The ndg condition is that A, B, and C are not collinear.

1.4. A Method of Eliminating Points. As shown in Examples 1.5 and 1.11, the area method is to *eliminate the constructed points* from geometry quantities. It is clear that we need only to give elimination methods for the minimal set of constructions.

Let $G(Y)$ be one of the quantities: $S_{ABY}, S_{ABCY}, P_{ABY}$, or P_{ABCY} for distinct points A, B, C, and Y. For three collinear points Y, U, and V, by Propositions 1.2 and 1.8 we have

$$(I) \qquad G(Y) = \frac{\overline{UY}}{\overline{UV}} G(V) + \frac{\overline{YV}}{\overline{UV}} G(U).$$

We call $G(Y)$ a *linear geometry quantity* for variable Y.

Lemma 1.13 *Let $G(Y)$ be a linear geometry quantity. Then $G(Y)$ equals to*

$$\begin{cases} G(W) + r(G(V) - G(U)) & \text{if } Y \text{ is introduced by (PRATIO } Y\ W\ U\ V\ r). \\ \frac{S_{UPQ}G(V) - S_{VPQ}G(U)}{S_{UPVQ}} & \text{if } Y \text{ is introduced by (INTER } Y \text{ (LINE } U\ V) \text{ (LINE } P\ Q)). \end{cases}$$

Proof. For the first case, take a point S such that $\frac{\overline{WS}}{\overline{UV}} = 1$. By (I)

$$G(Y) = \frac{\overline{WY}}{\overline{WS}} G(S) + \frac{\overline{YS}}{\overline{WS}} G(W) = rG(S) + (1 - r)G(W).$$

By Propositions 1.4 and 1.10, $G(S) = G(W) + G(V) - G(U)$. Substituting this into the above equation, we obtain the result. The second case comes from (I) and the co-side theorem. ∎

Let $G(Y) = P_{AYB}$. Since we have obtained the position ratios $\frac{\overline{UY}}{\overline{UV}}, \frac{\overline{YV}}{\overline{UV}}$ for Y when it is introduced by $C4$ and $C6$ in the above lemma, we can substitute them into the second formula in Proposition 1.8 to eliminate point Y from $G(Y)$.

Lemma 1.14 *Let Y be introduced by (TRATIO Y P Q r). Then we have $S_{ABY} = S_{ABP} - \frac{r}{4} P_{PAQB}$.*

Proof. Let A_1 be the orthogonal projection from A to PQ. Then by Proposition 1.3 $\frac{S_{PAY}}{S_{PQY}} = \frac{S_{PA_1Y}}{S_{PQY}} = \frac{\overline{PA_1}}{\overline{PQ}} = \frac{P_{A_1PQ}}{P_{QPQ}} = \frac{P_{APQ}}{P_{QPQ}}$. Thus $S_{PAY} = \frac{P_{APQ}}{P_{QPQ}} S_{PQY} = \frac{r}{4} P_{APQ}$. Similarly, $S_{PBY} = \frac{P_{BPQ}}{P_{QPQ}} S_{PQY} = \frac{r}{4} P_{BPQ}$. Now $S_{ABY} = S_{ABP} + S_{PBY} - S_{PAY} = S_{ABP} - \frac{r}{4} P_{PAQB}$. ∎

Lemma 1.15 *Let Y be introduced by (TRATIO Y P Q r). Then we have $P_{ABY} = P_{ABP} - 4rS_{PAQB}$.*

Proof. Let the orthogonal projections from A and B to PY be A_1 and B_1. Then $\frac{P_{BPAY}}{P_{YPY}} = \frac{P_{B_1PA_1Y}}{P_{YPY}} = \frac{\overline{A_1B_1}}{\overline{PY}} = \frac{S_{PA_1QB_1}}{S_{PQY}} = \frac{S_{PAQB}}{S_{PQY}}$. Since $PY \perp PQ$, $S_{PQY}^2 = \frac{1}{4} \overline{PQ}^2 \cdot \overline{PY}^2$. Then $P_{YPY} = 2\overline{PY}^2 = 4rS_{PQY}$. We have $P_{ABY} = P_{ABP} - P_{BPAY} = P_{ABP} - 4rS_{PAQB}$. ∎

Now we consider how to eliminate points from the ratio of lengths.

Lemma 1.16 *Let point Y be introduced by construction (PRATIO Y R P Q). Then we have $\frac{\overline{AY}}{\overline{CD}} = \frac{\frac{\overline{AR}}{\overline{PQ}} + r}{\frac{\overline{CD}}{\overline{PQ}}}$ if $A \in RY$; $\frac{\overline{AY}}{\overline{CD}} = \frac{S_{APRQ}}{S_{CPDQ}}$ if $A \notin RY$.*

Proof. The first case is obvious. For the second case, take points T and S such that $\frac{\overline{RT}}{\overline{PQ}} = 1$ and $\frac{\overline{AS}}{\overline{CD}} = 1$. By the co-side theorem and Proposition 1.4, $\frac{\overline{AY}}{\overline{CD}} = \frac{\overline{AY}}{\overline{AS}} = \frac{S_{ART}}{S_{ARST}} = \frac{S_{APRQ}}{S_{CPDQ}}$. ∎

Lemma 1.17 *Let Y be introduced by (TRATIO Y P Q r). Then* $\frac{\overline{AY}}{\overline{CD}} = \frac{P_{APQ}}{P_{CPDQ}}$ *if* $A \notin PY$; $\frac{\overline{AY}}{\overline{CD}} = \frac{S_{APQ}-\frac{r}{4}P_{PQP}}{S_{CPDQ}}$ *if* $A \in PY$.

Proof. The first case is a direct consequence of Proposition 1.9. If $A \in PY$, then $\frac{\overline{AY}}{\overline{CD}} = \frac{\overline{AP}}{\overline{CD}} - \frac{\overline{YP}}{\overline{CD}}$. By the co-side theorem, $\frac{\overline{AP}}{\overline{CD}} = \frac{S_{APQ}}{S_{CPDQ}}$; $\frac{\overline{YP}}{\overline{CD}} = \frac{S_{YPQ}}{S_{CPDQ}} = \frac{rP_{PQP}}{4S_{CPDQ}}$. Now the second result follows immediately. ∎

Lemma 1.18 *Let point Y be introduced by (INTER Y (LINE P Q) (LINE U V)). Then we have* $\frac{\overline{AY}}{\overline{CD}} = \frac{S_{AUV}}{S_{CUDV}}$ *if A is not on UV; otherwise* $\frac{\overline{AY}}{\overline{CD}} = \frac{S_{APQ}}{S_{CPDQ}}$.

Proof. This lemma can be proved as Lemma 1.16 similarly. ∎

Theorem 1.19 *We have a complete method of proving or disproving constructive geometry theorems.*

Proof. Let $E = F$ be the conclusion for a geometry statement. Using Lemmas 1.13–1.18, we can eliminate all the non-free points in the statement from E and F to obtain two new expressions E' and F'. If $E' = F'$ is an identity then the statement is true. Otherwise let E'' and F'' be the expressions obtained by replacing the areas and Pythagorean differences of free points in E' and F' by *area coordinates* (see [3]). Since area coordinates of free points are independent variables, the statement is true iff $E'' = F''$ is an indentity ∎

1.5 Geometry Problems from A.M.M. We have collect 135 geometry theorems mainly from the Problem Section of the Monthly. [5] is a collection of these theorems including the complete machine proofs for 92 geometry theorems. The following table contains timing and proof length statistics for the 135 geometry problems solved by our computer program. Maxterm means the number of terms of the maximal polynomial occurring in a proof. Lemmano is the number of elimination lemmas used to eliminate points from geometry quantities. In other words, lemmano is the number of deduction steps in the proof.

Proving Time		Proof Length		Deduction Step	
Time (secs)	% of Thms	Maxterm	% of Thms	Lemmano	% of Thms
$t \leq 0.1$	18%	$m = 1$	11%	$l \leq 5$	11%
$t \leq 0.5$	61%	$m \leq 2$	36%	$l \leq 10$	48%
$t \leq 1$	74%	$m \leq 5$	55%	$l \leq 20$	72%
$t \leq 5$	92%	$m \leq 10$	78%	$l \leq 30$	85%
$t \leq 30$	100%	$m \leq 52$	100%	$l \leq 68$	100%

2 Area Method and Geometry Education

There are two approaches to use the area method to geometry education. First, We may teach geometry courses in high schools with area as the central concept. Traditional geometry concepts, such as similar triangles and ratios, can be treated elegantly in the new framework. Textbooks based on this idea have been written [15] and experimental classes have been taught in Chengdu, China. Second, we have developed a computer program, GEOMETRY EXPERT[3], which has a three-fold purpose: (1) It is a powerful geometry theorem prover. (2) It is a geometric diagram editor. (3) It is a geometry tutor.

2.1. GE as a Geometry Theorem Prover. GE is a master level geometry theorem prover. Within its domain, it invites comparison with the best of human geometry provers. The following geometry theorems (all of them are built-in examples of GE) should provide some illustration of the prover's capabilities:

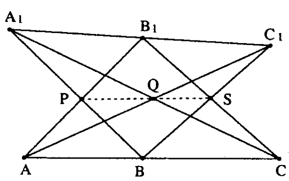


Figure 3

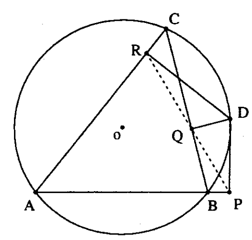


Figure 4

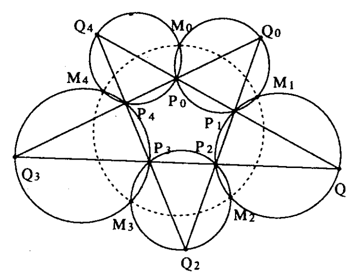


Figure 5

Example 2.1 (Pappus' Theorem) *Let points A, B and C be on one line, and A_1, B_1 and C_1 be on another line. Let $P = AB_1 \cap A_1B$, $Q = AC_1 \cap A_1C$, and $S = BC_1 \cap B_1C$. Show that P, Q, and S are collinear.*

Example 2.2 (Simson's Theorem) *Let D be a point on the circumscribed circle of triangle ABC. From D three perpendiculars are drawn to the three sides BC, AC, and AB of triangle ABC. Let Q, R, and P be the three feet respectively. Show that P, Q and R are collinear.*

Example 2.3 (Miquel circle for five lines) *Given five lines, we can determine five Miquel points by considering each four of the five lines. Show that the five Miquel points are on the same circle.*

Besides the area method, GE also implements the following proving methods.

1. *Wu's method.* Wu's method is a coordinate-based method. It first transforms geometry conditions into polynomial equations in the coordinates of the involving points, then deals with the polynomial equations with the characteristic set method. This method has been used to prove more than 600 geometry theorems. See [13, 2] for detailed description.

[3]Preliminary versions of the system are now available via anonymous ftp at emcity.cs.twsu.edu: pub/geometry/software: ge_sun.tar.Z is a SUN workstation version and euc_pc.zip is a PC version.

2. *Vector method.* This method is a variant of the area method and is based on the calculation of vectors and complex numbers. For many geometry problems, this method can produce very elegant proofs. See [3] for detailed description.

3. *Full-angle method.* This method is based on the calculation of full-angles (directed angles between two lines).

4. *The Grobner basis method.* The Gröbner basis method is also a coordinate-based method which uses the Gröbner basis to deal with the polynomial equations. See [2, 8, 7] for detailed description.

Why do we use so many methods in the prover? First, with these methods, for the same theorem, the prover can produce a variety of proofs with different styles. Second, for a certain class of geometry theorems, a particular method may produce much shorter proofs than other methods. Third it is important for the method to be used in geometry education, since different methods allow students to explore different and better proofs.

2.2. GE as a Geometry Diagram Editor. For a geometry statement described constructively, we can use GE to draw its diagram automatically. The program will ask the user to select the positions for the free and semi-free points in the statement using the mouse pointer, and compute the positions for the fixed points automatically.

GE also provides a tool, the *graphical input*, for the user to input a geometric statement directly using the mouse via a set of constructions. The process is to *draw* the diagram and the prover will transform the drawing process to corresponding geometric conditions. GE also provides the following distinctive edit features:

1. *Drag*. The user can drag certain points in a diagram to change the shape of the diagram. As a result, the user may see intuitively whether certain properties are valid in the diagram.

2. *Animation.* The user can apply animation to the diagram by causing a particular point in the diagram to move in a prescribed orbit. During the process, the diagram changes continuously.

3. *Printing*. The user may save the figure in PostScript form for printing.

2.3. GE as a Geometry Tutor. The geometry tutor, not yet completed, will guide geometry students as they learn to implement the five methods for proving geometry theorems. The tutor will employ the following methods.

1. *Hierarchical Proofs.* As a geometry tutor, the system will combine the area method with a sophisticated graphics display. A proof produced by the system is displayed on the screen at various hierarchical levels, the highest level being similar to one presented by a geometry teacher. For a moderately difficult theorem, only 4–8 steps are required to complete the proof. The visual representation of the proof is accompanied by a diagram in which each step is illustrated by the highlighting of corresponding geometric objects. A student may click the mouse button at any step in the proof to open an auxiliary text

window at level 2, (or level 3, etc.) to have a more detailed explanation of that step. In this way, the student can understand a proof in full detail, while keeping the highest level as the main focus.

2. *Step by step proof.* In proving a geometry theorem, the tutor will ask the student to guess the next step of the proof. If the student is right, the tutor will accept the response and continue with the proof. Otherwise, the computer may give hints or may show all the possible choices for the next step of the proof and let the student choose one.

References

[1] J. R. Anderson, C. F. Boyle, A. Corbett, & M. Lewis, The Geometry Tutor, in *Proc. of the IJCAI*, Los Angeles, USA, 1985, p. 1–7.

[2] S.C. Chou, *Mechanical Geometry Theorem Proving*, D.Reidel Publishing Company, Dordrecht, Netherlands, 1988.

[3] S.C. Chou, X.S. Gao, & J.Z. Zhang, *Machine Proofs in Geometry*, World Scientific, Singapore, 1994.

[4] S.C. Chou, X.S. Gao, & J.Z. Zhang, Automated Production of Traditional Proofs for Constructive Geometry Theorems, *Proc. of Eighth IEEE Symposium on Logic in Computer Science*, p.48–56, IEEE Computer Society Press, 1993.

[5] S. C. Chou, X. S. Gao, & J. Z. Zhang, Automated Solution of 135 Geometry Problems from American Mathematical Monthly, TR-94-8, CS Dept., WSU, August, 1994.

[6] H. Gelentner, J.R. Hanson, and D.W. Loveland, Empirical Explorations of the Geometry-theorem Proving Machine, *Proc. West. Joint Computer Conf.*, 143-147, 1960.

[7] D. Kapur, Geometry Theorem Proving Using Hilbert's Nullstellensatz, *Proc. of SYMSAC'86*, Waterloo, 1986, 202–208.

[8] B. Kutzler & S. Stifter, Automated Geometry Theorem Proving Using Buchberger's Algorithm, *Proc. of SYMSAC'86*, Waterloo, 1986, 209–214.

[9] A. J. Nevins, Plane Geometry Theorem Proving Using Forward Chaining, *Artificial Intelligence*, **6**, 1-23.

[10] A. Tarski, *A Decision Method for Elementary Algebra and Geometry* , Univ. of California Press, Berkeley, Calif., 1951.

[11] H. Wang, A Variant to Turing's Theory of Computing Machines, *J. of ACM*, vol. 4, 1957, p.63–92.

[12] Wu Wen-tsün, On the Decision Problem and the Mechanization of Theorem in Elementary Geometry, *Scientia Sinica* 21(1978), 159–172; Also in *Automated Theorem Proving: After 25 years*, A.M.S., Contemporary Mathematics, 29(1984), 213–234.

[13] Wu Wen-tsün, *Basic Principles of Mechanical Theorem Proving in Geometries*, Volume I: Part of Elementary Geometries, Science Press, Beijing (in Chinese), 1984.

[14] L. Yang, J.Z. Zhang, & X.R. Hou, A Criterion of Dependency Between Algebraic Equations and Its Applications, *Proc. of the 1992 international Workshop on Mechanization of Mathematics*, p.110-134, Inter. Academic Publishers.

[15] J. Z. Zhang, *A New Approach to Plane Geometry,* Sichuan Educational Publishing Inc., (in Chinese) 1992.

Geometric Constructions with the Computer

Jen-chung Chuan
Department of Mathematics
National Tsing Hua University
Hsinchu, Taiwan 300
jcchuan@math.nthu.edu.tw

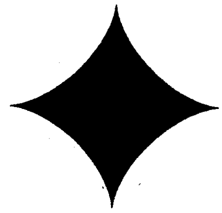
1. Introduction

Computer can be used as a tool to represent and communicate geometric knowledge. With the appropriate software, the geometric diagram can be manipulated through series of animation that offer more than one particular snapshot as shown in a traditional mathematical text. Geometric constructions with the computer enables the learner to see and understand a diagram in different ways. Engaging in constructing the animation encourages the learner to go through the abstract process of formulation of conjectures, generalization, condition-simplification and classification. In what follows we shall offer the visual experience of geometric animation by guiding the audience through demonstrations of geometric construction from topics in enveloping curves, linkage, polynomial interpolation, inversion, hypocycloid and epicycloid.

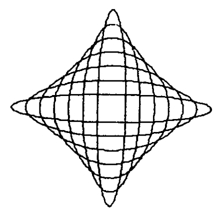
2. Envelop

When a system of lines or curves expressed in the form of an algebraic equation, it is not clear immediately if there is another curve touching every member of the system. A precise drawing allows the learner to perceive this notion of enveloping curve visually. Interesting enveloping curves occur when the system obeys certain common geometric conditions. It is thus desirable to generate the curve by means of the direct geometric notion. Consequently a dynamic geometric environment is particularly suitable for such exploration.

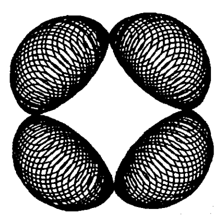
An enveloping curve may appear when only straight line segments are drawn:



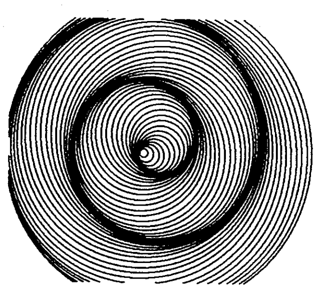
In this case all segments have the same length and have endpoints lying on two mutually perpendicular lines. The same envelop may appear as a curve touching a family of ellipses:



All ellipses have the same major axis, the same minor axis, and the same sum of distances from the foci. This astroid also appears as a part of enveloping curve associated with a family of circles:



The centers of these circles lie on one single (hidden) circle and each circle is tangent to the astroid. An enveloping curve needs not lie on the boundary:



Notice that a slower computer may even serve as a better pedagogical tool, since the attention of the learner has been drawn by the delayed image display.

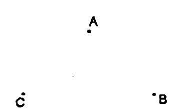
3. Linkage

At MIT library one can retrieve a massive book that weighs over 3 kilograms and was published in 1951 in fulfillment of an MS thesis. The title of this dinosaur is called "Analysis of the Four Bar Linkage".

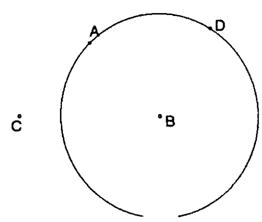
The four-bar linkage is one of the simplest mechanism and may be regarded as a basic mechanism. The equations which describe the displacement, velocity and acceleration of points on the connecting rod are long and difficult to handle, making mathematical synthesis a very tedious process which few designers will undertake. The publication claims to includes over 7000 displacement paths with the velocity given at 72 equal intervals of drive crank angle along each path and therefore represents about 500,000 solutions of interest to engineers.

Geometric construction can be adopted to simulate the movement of a linkage. We now show how to construct the Fermat linkage, a particular four-bar linkage used to draw a lemniscate of Bernoulli.

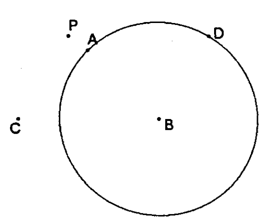
1) Take an equilateral right triangle ABC :



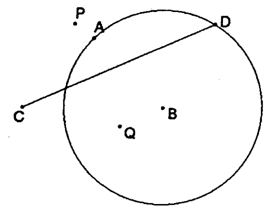
2) With B as center, draw a circle passing through A and then take a point D confined on the circle:



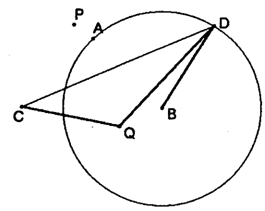
3) Find the point P so that $CBDP$ is a parallelogram.



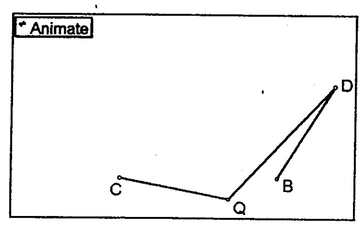
4) Construct the reflection Q of P with respect to the line CD :



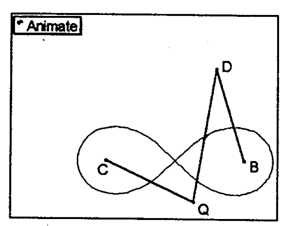
5) Connect the line segments CQ, QD, DB. These are the bars in the linkage:



6) Make the point D travel along the circle and then hide the circle, the line segment CD as well as the points A and P :



7) This is the locus of the mid-point of the segment DQ, when the Animate button is pressed:



4. Polynomial Interpolation

Suppose that we are given points $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ in the coordinate plane with the x_i distinct. The problem of polynomial interpolation asks for a polynomial function whose graph passes through these points.

Of the many methods of solution to this problem, the ones suggested by Aitken and Neville can be most readily implemented under the dynamical geometric environment.

The principle behind constructing the graph of the interpolating polynomial is the following lemma of Aitken:

Lemma. Suppose that the functions f and g satisfy $f(x) = g(x)$ for $x \in T$. If $a \neq b$ then the function

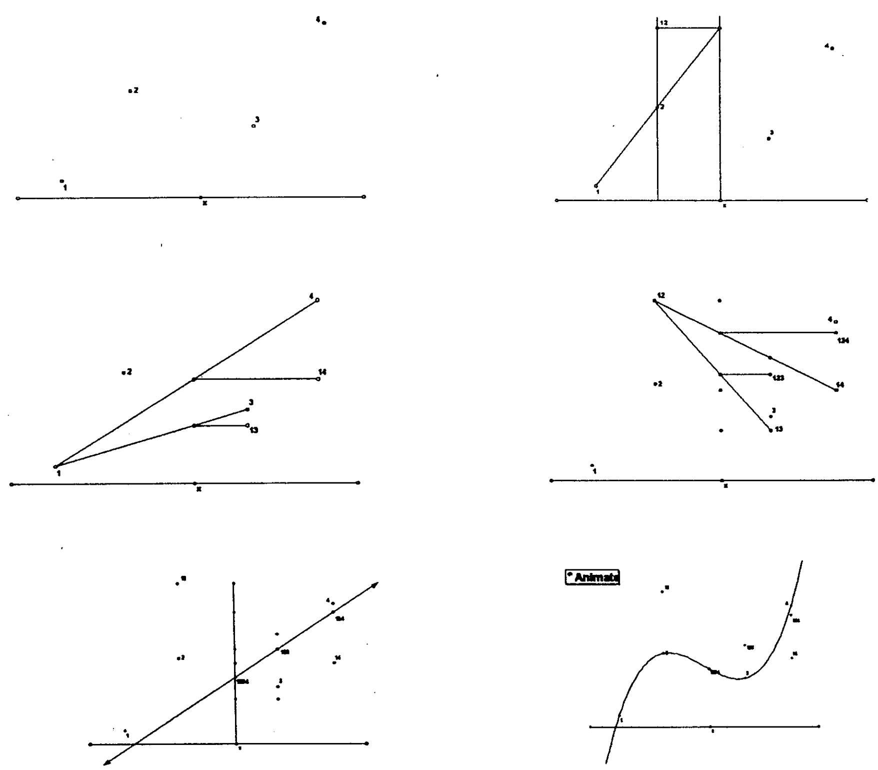
$$h(x) = \frac{(b-x)f(x) + (x-a)g(x)}{b-a}$$

satisfies $f(x) = g(x)$ for $x \in T$, $h(a) = f(a)$, $h(b) = g(b)$.

Let $x_1, x_2, \cdots, x_n$ be distinct points. Let $y_1, y_2, \cdots, y_n$ be given points. For each nonempty subset T of $\{1, 2, ..., n\}$, let p_T be the (unique) polynomial of degree $\leq |T| - 1$ satisfying $p_T(x_j) = y_j$ for $j \in T$. If $|S \backslash T| = 1 = |T \backslash S|$, Aitken's lemma shows us how to produce $p_{S \cup T}$ from p_S and p_T. The required interpolating polynomial $p_{\{1,2,\cdots,n\}} \equiv p_{12\cdots n}$ can be computed from this table:

p_1	p_{12}	p_{123}	p_{1234}	p_{12345}
p_2	p_{13}	p_{124}	p_{1235}	
p_3	p_{14}	p_{125}		
p_4	p_{15}			
p_5				

with the obvious initial data $p_j \equiv y_j$ for $j = 1, 2, \cdots, n$.

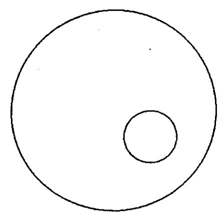


The preceding procedure may be altered by finding the interpolating polynomial $p_{12\cdots n}$ building through this table:

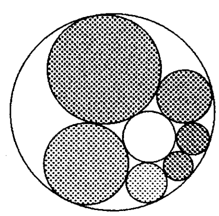
p_1	p_{12}	p_{123}	p_{1234}	p_{12345}
p_2	p_{23}	p_{234}	p_{2345}	
p_3	p_{34}	p_{345}		
p_4	p_{45}			
p_5				

5. Inversion

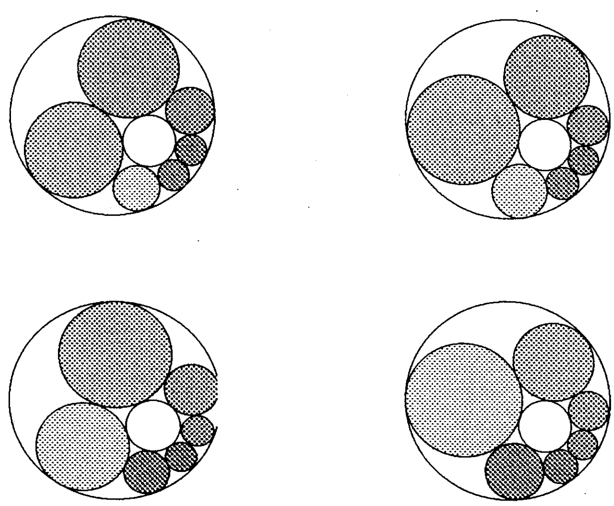
Take two circles, one inside another:



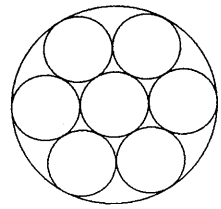
It is not always the case that one can find a series of circles kissing each other so precisely as thus:



Steiner's porism states that if such situation arises once, it will occur infinitely often:



What is the secret behind this animation of Steiner porism? Answer: the inversion. We begin with the particular situation



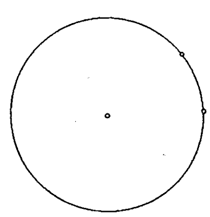
and then invert the configuration with respect to another circle which is to be hidden. As this symmetric figure rotates, the resulting inverted figure, obeying the law that tangent circles be inverted to tangent circles, achieves the desired result. The method of this demonstration actually guides us to the proof of Steiner's porism! Just invert the two disjoint circles into concentric ones. The configuration of kissing circles close up if and only the inverted figure is symmetric. QED.

6. Epicycloid and Hypocycloid

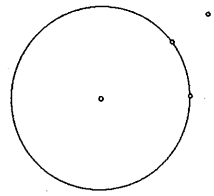
An epicycloid is the curve traced by a point in the circumference of a circle rolling outside a fixed circle. A hypocycloid is the curve traced by a point in the circumference of a circle rolling inside a fixed circle. These curves possess many remarkable properties that fascinate both mathematicians and nonmathematicians alike. It is possible to explore the magic and beauty of these curves through purely geometric constructions. Computer makes it possible for the learner to be exposed to the geometric approach before hooked up to analytical and algebraic methods.

We now show how to generate the epicycloid pictorially:

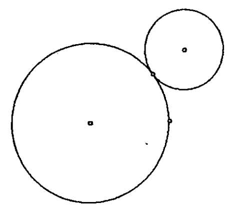
Starting with a circle with "the point of contact" attached.



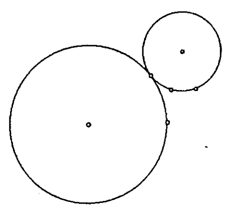
Next, locate the center of the rolling circle.



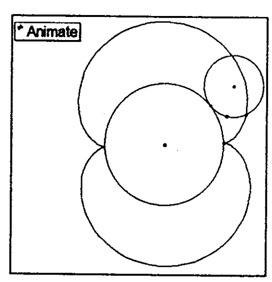
Draw the rolling circle.



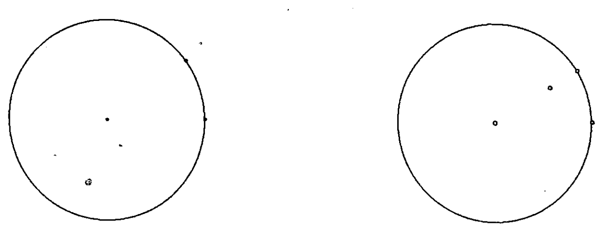
Construct the "moving point" by suitably rotating the point of contact w.r.t. the center of the moving circle.

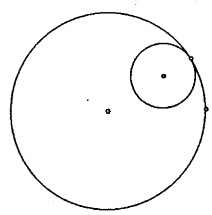


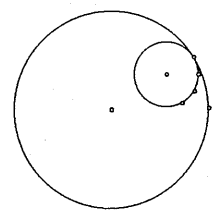
This is the result of pressing the "Animate" button:

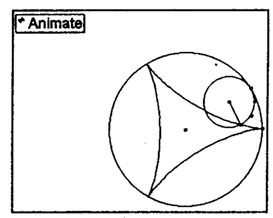


The animation of a hypocycloid is created through a similar process:

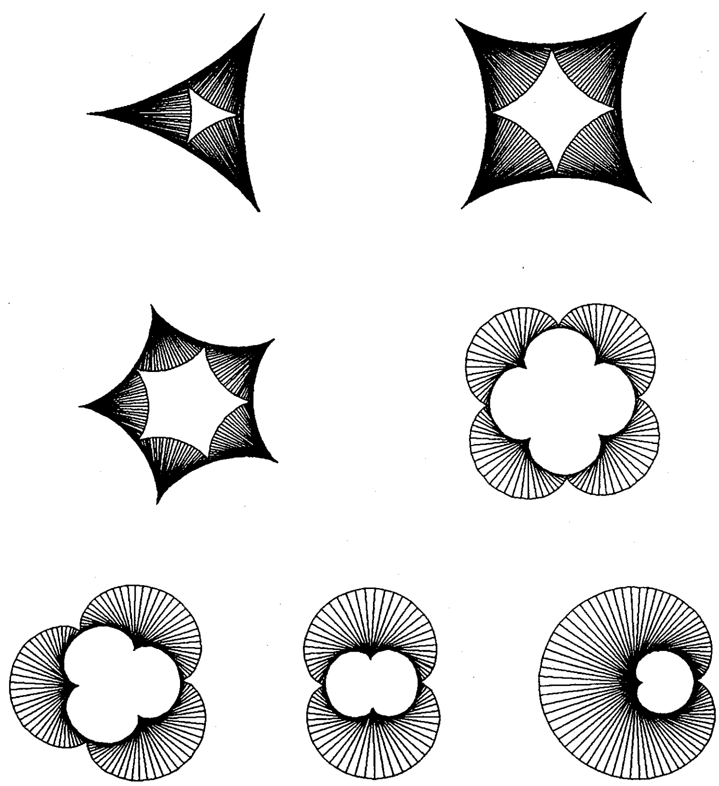






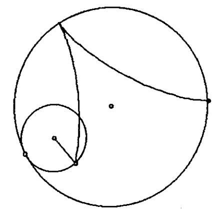


There is a whole host of computer activities related to epicycloids, hypocycloids by drawing the series of line segments joining various points of the curve to the corresponding center of curvature.

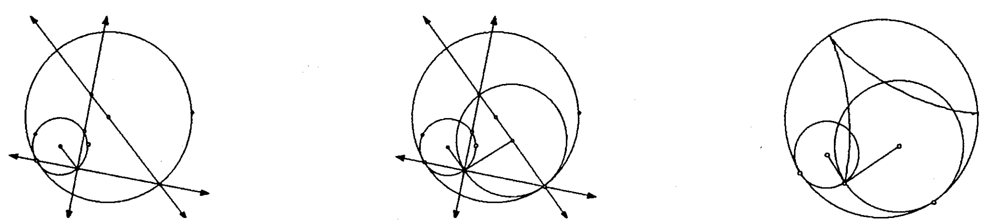


7. Double Generation

Every hypocycloid and every epicycloid can be generated in two ways. Consider the case the hypocycloid traced by a point on the circumference of a rolling circle:



This rolling circle determines another circle rolling in the opposite direction containing the same fixed point on its circumference that traces the hypocycloid. It is difficult to perform a dynamic demonstrate the double generation by means of mechanical devices. With the assistance of computer, however, the animation of such phenomenon can be vividly shown.



Not convinced just by looking at these pictures? Then you have got to see the demonstration.

References

1. H.S.M. Coxeter, Introduction to Geometry, Wiley, New York, 1961.
2. J. A. Hrones and G. L. Nelson, Analysis of the Four Bar Linkage, Wiley, 1951.
3. E.H. Lockwood, A Book of Curves, Cambridge University Press, 1961.
4. K. Rektorys (Editor), Survey of Applicable Mathematics, Illiffe, London, 1969.
5. J. Todd, Survey of Numerical Analysis, McGraw-Hill, New York, 1962.

USING ISETL TO MASTER THE CONCEPT OF FUNCTION

Dr. Julia M. Clark & Dr. Robert L. Pour
Emory & Henry College
Department of Mathematics
Emory & Henry College
Emory, Virginia 24327-0947
U.S.A.
mathphys@delphi.com

The mathematics community has continually reviewed and revised both the methodology of teaching and the content of the curriculum. In the past decade, however, we have seen a phenomenal growth in the mathematics reform movement. The increased importance of mathematical literacy in an increasingly technological world (see [Glimm] p 9), and the dismal performance of U.S. students compared with students of other nations (see [McKnight, et al. p 12]) has focused attention, and renewed interest in the teaching of mathematics. At the secondary level the National Council of Teachers of Mathematics has adopted Curriculum and Evaluation Standards for School Mathematics [NCTM], while the Mathematical Association of America has encouraged reform through the publication of monographs such as Reshaping College Mathematics [Steen, 1989], Calculus for a New Century [Steen, 1987], and Computers and Mathematics [Smith, et al]. We wish to focus in this paper on one aspect of this reform movement, namely the use of symbolic computation as an aid in teaching and learning mathematics.

In the past fifteen years computers have become increasingly available to both students and faculty. Their use in schools, elementary, secondary and post-secondary, has been promoted, albeit often without a clear purpose. The use of computers in mathematics classes does not automatically insure an increase in student performance, indeed it can have the opposite effect (see [Dubinsky, p 73]). Without an underlying epistomological theory, it is difficult to develop or evaluate pedagogical practices.

Seymore Papert of MIT, a former student of Piaget, is well known for his development of the computer language LOGO. The development and application of this software as a learning tool was in no small part influenced by the learning theories of Piaget.

Papert says "The understanding of learning must be genetic. It must refer to the genesis of knowledge." (see [Papert, p vii]). He begins with Piaget's idea of "assimilation". When confronted with a new concept, a child relates the new concept to those concepts previously learned. They attempt incorporate it into their existing conceptual framework in a way meaningful, and consistent to them. It may be possible to incorporate an idea into the existing model, and hence achieve an understanding of it, or it may contradict existing knowledge, in which case the old knowledge may have to be abandoned. In either case the process must be an active one. Unfortunately the model followed by many teachers of mathematics is what Papert calls a "dissociated" model (See [Papert, pp 47-48]). Rote learning emphasizes the acceptance of information which is essentially meaningless, hence need not be assimilated into an existing framework. Piaget emphasizes that any real learning occurs in a context. The constructivist theory of learning follows in the tradition of Piaget and Papert. Its fundamental premise is that the learner must construct his/her individual realization of a mathematical concept.

The concept of a function is a fundamental building block in any mathematics course. In particular the Calculus is essentially a tool to study the behavior of functions. There is, however, widespread misunderstanding of the function concept. The constructivist theory holds that a mathematical notion must be viewed as an **action**, a **process**, an **object**, and combinations of these three (see [Cuoco,pp 121-140]). The most rudimentary understanding of a function is as an ***action***. At this level the student views a function as a sequence of discrete, isolated computations. A function such as f: x--> 3x-2 would only have meaning for specific values of x. Each evaluation of a function is an experience bearing little relationship to previous evaluations; it is also unlikely that similarities would be observed between closely related functions (e.g. f: x-->3x-2, and g: x-->2x-1).

At the next level of understanding of the function concept the student is able to view the function as a ***process***. Here the form of a function is realized. A function such as h: x-->3x-2 is interpreted as the repeatable process of tripling a number and then subtracting two. This level of abstraction enables the student to see essential similarities between functions such as f: x-->3x-2 and g: x-->2x-1.

After years of studying arithmetic the student tends to operate with numbers as the quintessential mathematical object. Numbers can be combined by addition, subtraction, multiplication, division. They can be transformed by exponentiation, etc. Many areas of mathematics, however, require this higher level of understanding of the concept of function. In calculus and many other areas, functions to must be viewed as ***objects*** which can be combined and transformed. This level of abstraction is vital for the understanding of the idea of composition, differentiation, integration, solutions of differential equations and many other applications.

The appropriate use of thoughtfully designed software can be a powerful tool for the student to construct his/her concept of function . ISETL is an excellent example of such software (another is Papert's LOGO). ISETL is an inpreted version of the compiled programming language SETL, which was designed at the Courant Institute by Schwartz et al in 1980. Gary Levin, at Clarkson University, developed ISETL as a tool for teaching and learning traditional topics in discrete mathematics. It is written in C, can be run on most personal computers and main-frames and is interactive, so one does not have to wait for compilation. The language supports functions as first-class objects, infinite precision arithmetic, some graphics features, set formers, and existential and universal quantifiers. Jennie Dautermann, [Dautermann]has written a fairly good reference for learning ISETL.

ISETL (Interactive SET Language) is a powerful, interactive programming language, which can easily be used to help students construct the concept of functions as processes, and then objects, because ISETL's syntax is almost identical to standard mathematical notation. (Writing ISETL code fragments at the very least helps students to write precise mathematical statements!) Since ISETL itself treats mathematical entities as both processes and objects, students can use ISETL to explore functions as procedures or algorithms, tables, sets of ordered pairs (tuples), and graphs.

For example, consider the function $f(x) = x^2$. There are a variety of ways to implement this function using ISETL:

```
1)   f:=func(x);
       return x**2;
     end;

2)   g:= |x -> (x**2)|;

3)   h:={[x,x**2]: x in [1..100]}

4)   r:=[i**2: i in [1..20]];
```

The code for f is an example of a "computer function" which treats the function as a process. An argument(x)is input into the function, the input is manipulated, and the results of the manipulation are output. Students can evaluate the function at various input values -- for example typing f(4); results in the ISETL output "16". The representation of g is similar, but is written as a mapping or expression for squaring: g maps an input to its square. Students can even check that although f and g agree on any finite set of domain values, ISETL does not view f and g as being equal objects. For example,

```
f = g;
```

results in "false" but

```
forall x in {1,1.2.. 10}| (f(x) = g(x)};
```

results in "true".

Function h is of course a representation of squaring as a set of ordered pairs, while r represents the function as a sequence. The queries

```
r(4);  h(4);  f(4);  g(4);
```

will all result in the output 16, but no two of the functions are viewed by ISETL as functionally equivalent. It is a simple matter to write an ISETL function, TABLE, which converts any function defined as f or g above into a table or set of order pairs for a given finite domain interval [a,b] and number of pairs:

```
TABLE:=func(f,a,b,n);
     return {[x,f(x)]: x in [a,a+N..b]};
end;
```

The code

```
SEQUENCE:=[f(x): x in [1..20]];
```

makes a similar conversion from function expression into a squence. Sets of ordered pairs can also be input directly as a function, then plotted, which enables one to deal with raw data, and with implicitly defined functions.

In order to utilize ISETL effectively, one must provide students with suitable exercises and explorations. Dubinsky, et al have written a Calculus text [Dubinsky, & Schwingendorf]which contains computer activities (using ISETL or some other mathematical programming language) which are particularly effective in developing student constructions of the concept of functions as processes and eventually objects. For example, students are asked to represent the relationship between federal income tax and income as an expression, a computer functions, a table, and a graph. This enables the students to readily switch back and forth between the various representations of a function and to explore various properties of functions. These initial activites are designed to internalize students' understanding of functions as processes.

Later activities ask students to write functions whose outputs are not simply numbers, but are boolean. The action of these functions might be to check whether an ordered pair (x,y) satisfies a given mathematical expression. This encourages students to view even implicitly defined functions as objects, and to internalize the implied process. Eventually the students are manipulating the functions themselves, composing functions, and even writing functions which have other functions as both inputs and outputs:

```
Derivative:=func(F,h);
     return func(x);
          return(F(x+h) - F(x))/h;
     end;
end;
```

Similar texts are available for PreCalculus [Dubinsky & Kiaie], Discrete Mathematics [Baxter, 1989] and Abstract Algebra [Dubinsky & Leron]. These texts are being used at several test sites throughout the Unites States and the results have been incouraging. "Working with ISETL is consistent with a theory of learning according to which the acquisition of mathematical knowledge consists of constructing dynamic mental processes which can be reversed in thought, as well as encapsulated to form new mathematical objects." [Dubinsky] The idea is that what one constructs on the computer with full awareness, is necessarily constructed in his or her mind. Teachers who have used ISETL in this manner report that students "come to understand the connection between a function process and a set of ordered pairs..., work well with composition of functions,... [and] enjoy learning mathematics this way and work more intensely." [Dubinsky] (See [Baxter,1933], and [Monteferrante].)

REFERENCES

[Baxter, 1993] Baxter, Nancy Hood. Understanding How Students Acquire Concepts Underlying Sets. Dubinsky, Ed & James J. Kaput. (Ed.) *Research Issues in Undergraduate Mathematics Learning: Preliminary Analyses and Results*. 99-106. Mathematical Association of America. 1993.

[Baxter, 1989] Baxter, Nancy. *Learning Discrete Mathematics with ISETL*. Springer Verlag. 1989.

[Breidenbach, Hawks & Nichols] Breidenbach, Daniel, Julie Hawks, and Devilyna Nichols. *Development of the process conception of Function.* Educational Studies in Mathematics; v23. 247-285 June 1992.

[Cornu & Dubinsky] Cornu, Bernard and Ed Dubinsky, Ed. *Using a Cognitive Theory to Design Educational Software*. Education and Computing; v5. 73-80. 1989.

[Cuoco] Cuoco, Albert A. Multiple Representations for Functions. Dubinsky, Ed & James J. Kaput. (Ed.) *Research Issues in Undergraduate Mathematics Learning: Preliminary Analyses and Results.* 121-140. Mathematical Association of America. 1993.

[Dautermann] Dautermann, Jennie. *Using ISETL -- A Language for Learning Mathematics*. West Publishing. 1992.

[Dubinsky] Dubinsky, Ed, Using the Isetl Programming Language in teaching Mathematics, in Johnson, David C. and Lovis Frank (Editors). *Informatics And The Teaching of Mathematics.* 141-142. North Holland. 1987.

[Dubinsky & Kiaie] Dubinsky, Ed, Catherine Kiaie, Jerry Przybylski, Barbara Reynolds and Keith Schwingendorf. *Precalculus and Problem Solving: A Computer Aided Approach*. In preparation. 1994.

[Dubinsky, & Schwingendorf] Dubinsky, Ed , Keith E. Schwingendorf, & David M. Mathews. *Calculus, Concepts & Computers*. McGraw Hill, Inc. 1995.

[Dubinsky & Leron] Dubinsky, Ed & Uri Leron. *Learning Abstract Algebra with ISETL*. Springer-Verlag. 1994.

[Earnest] Earnest, Paul. *Mathematics, Education & Philosophy.* Falmer Press. 1994.

[Furth] Furth, Hans G. *Piaget and Knowledge.* Prentice Hall. 1969.

[Glimm] Glimm, James G. (Editor). *Mathematical Sciences. Technology. and Economic Competitiveness.* National Academy Press. 1991.

[Harel& Dubinsky] Harel, Guershon and Ed Dubinsky, Ed. (Editors). *The Concept of Function.* Mathematical Association of America. 1992.

[Jaworski] Jaworski, Barbara. *Investigating Mathematics Teaching.* Falmer Press. 1994.

[Papert] Papert, Seymour. *Mindstorms.* Basic Books 1980.

[McKnight, et al] McKnight, CC, F.J. Crosswhite, J.A. Dossey, E. Kifer, J.O. Swafford, K.J. Travers, & T.J. Cooney. *The Underachieving Curriculum: Assessing U.S. School Mathematics from an International Perspective.* Stipes Publishing Co. 1987.

[Monteferrante] Monteferrante, Sandra. Implementation of Calculus, Concepts and Computers at Dowling College. *Collegiate Microcomputer.* XI; 94-98. May 1993.

[NCTM] Nation Council of Teacher of Mathematics. *Professional Standards for Teaching Mathematics.* NCTM. 1991.

[Sierpinska] Sierpinska, Anna. *Understanding Mathematics.* Falmer Press. 1994.

[Simon] Simon, Martin A. Reconstructing Mathematics Pedagogy From a Constructivist Prespective. *Journal for Research in Mathematics Education.* 26; 114-145. 1995.

[Smith, et al] Smith, D.A., G.J. Porter, L.C. Leinbach, & R.H. Wenger, Ed. *Computers and Mathematics: The Use of Computers in Undergraduate Instruction.* Mathematical Association of America. 1989.

[Steen, 1987] Steen, L.A. Ed. *Calculus for a New Century: A Pump, not a Filter*. Mathematical Association of America. 1987.

[Steen, 1989] Steen, L. *Reshaping College Mathematics: A Project of the Committee on the Undergraduate Program in Mathematics*. The Mathematical Association of America. 1989.

[Tucker, Leitzel] Tucker, Alan C. and Leitzel, James R.C. (Editors). *Assessing Calculus Reform Efforts.* The Mathematical Association of America 1995.

[Von Glaserfeld] Von Glaserfeld, Ernst. *Radical Constructivism.* Falmer Press 1995.

TEACHING MATHEMATICAL CONCEPTS BY ANIMATION.

Ivan Cnop,
Vrije Universiteit Brussel,
Pleinlaan 2, B 1050 Brussels, Belgium
icnop @ vnet3.vub.ac.be

•

Introduction.

Why do most students have more problems understanding calculus and elementary analysis than other areas of mathematics? In many areas, the subject can be clarified by a picture. Most topics in geometry or algebra and discrete mathematics are finite and a single picture will do. Nobody will ask questions when he sees the drawing of a couple of intersecting lines, a square with its right angles indicated, or some arrows describing a permutation of the elements in a finite set. In analysis however, many objects are defined by a limiting process of an infinite collection and it is not possible to illustrate this by a single, or even a finite number of pictures.

Calculus describes change. Some quantities considered in the process of modelling are formally labeled functions of some variables and their behaviour is analysed. Calculus is mainly interested in the rate of change, the extreme values, the cumulative effect of change, comparing growth and decay. The changes considered may be both quantitative (e.g. increase or decrease) and qualitative (e.g. smooth behaviour versus discontinuities or chaos). Changes are difficult to put down in writing since printed texts are, by their nature, frozen. Change can be written down in the form of formulas but these may hide the essential behaviour. Even if a calculus course develops good computing skills, the results still have to be interpreted to exhibit the consequences upon the behavior of the quantities under consideration. Drawings in textbooks have been used for a long time, but they in turn may introduce unwanted artifacts or misinterpretations.

Historically, it has been difficult to develop the underlying theory of real (and complex) analysis. Working with functions was clumsy for a long time and the notations remained obscure until the notion of variable was introduced. Defining change was even more complicated. G. Leibniz introduced some infinitely small quantities dx over which increments were measured. Great mathematicians after him were very careful, even shy, expressing these concepts in writing. Even A. Cauchy himself wrote long sentences circumscribing them, about variables which tended to some value but never quite attained it. Eventually K. Weierstrass formalised the limit concept in the $\varepsilon - \delta$ way. This was further generalised by N. Bourbaki [Bo] and E. Cech [Ce] into topological concepts and their terminology was finally adopted by educational reforms in many countries a couple of decades ago. With hindsight many people now feel these reforms were overdone.

The essential idea of change is no longer apparent in the $\varepsilon - \delta$ terminology. It will only express qualitative change (existence of a limit or a discontinuity) if used correctly. Students find it difficult to grasp and to apply it correctly. They ask: "What ε should one consider and how many does one need? What is the relationship between those ε and δ and how does one find it ?" . The correct answer is: all ε, that is infinitely many of them, and for each of them, a corresponding δ. It is difficult to write this correspondence down unless one knows some kind of rigid behaviour, a Lipschitz property for instance. If ε / δ remains bounded, the function is Lipschitz at the point. But how does one verify this in general?

Further analysis is necessary if one wants to measure quantitative changes (smoothness, Lipschitz behaviour). Many instructors, led by Lee P.Y. [L] are therefore trying to avoid the use of the $\varepsilon - \delta$ terminology. Following ideas of A. Robinson the Paris School is experimenting with using the abstract setting of non-standard analysis as a basis for some analysis courses [R]. It is still too early to decide if their approach will be succesful, although it seems to be catching on the scientific community, mainly because intuitively speaking infinitesimals look familiar. To make correct statements, the difficulty of $\varepsilon - \delta$ terminology has to be replaced by some non-trivial system of axioms in set theory.

The role of animation.

It is impossible to describe change by a static medium. The conceptual transfer would be much easier if we describe change by something which does indeed change. As a matter of fact, our senses are trained to become aware of things that change: moving objects, panoramic views, music patterns, live performances, traffic, ... Our attention is immediately drawn by movement and memories are more intense. We all have become addicted to moving images since the film and television industries caught hold of our cultures. It will be an unfair fight if we try to offer our students mathematics concepts by printed text, classroom teaching and drill exercices after they have been overwhelmed by the possibilities of sattelite television, video games and computer networks. This battle will certainly be lost. It may even explain to a great extent the dwindling interest in mathematics studying which is now apparent in many developed countries.

The technology for making animations exists. It is built-in in computer algebra packages such as Mathematica™ [W], running on all recent computers, either personal or over networks. Other authors are also using Mathematica™ animations to make calculus courses more attractive [PW].

The problem is to represent mathematical concepts in a faithful way without introducing artifacts due to the finiteness of the software package or the machine on which it is running.

For some time, there has been a debate among mathematicians wether a machine can prove a theorem. The first theorem which relied on a computer for its proof was the solution to the four color problem [AH], and computers are now frequently used for difficult combinatorial or group theoretical proofs. For problems requiring infinite reasoning, this remains debatable. But without giving complete proofs, the computer may still help to convince by presenting on request a large number of examples.

Although a single counterexample disproves an assertion, one cannot prove a theorem by an example or a drawing. Only the knowledge that in all instances (special cases) the assertion verifies will be sufficient for a proof. Therefore, a picture or an example should be as general as possible. If a reasoning is performed in a generic situation (one which can be shown to cover all possibilities), then we do have a proof.

The same ideas are valid while working with animations. As an example we start by the representation of numbers on an axis. While scientists give their measurement up to a certain precision depending on their instruments, a real number in mathematics has an (possibly) infinite decimal expansion. To locate it, one needs an infinite sequence of approximations for its value. This corresponds to an infinite number of dots on some axis. In general, a finite number of dots will only take us within some small distance of its value even if the difference cannot be discerned on a computer screen (or any other kind of drawing). But if we are allowed to choose subsequences as many times as we want, by zooming in with arbitrary magnification or equivalently, defining a better resolution on the axis, then the definition of our real number is as correct as if done by the standard limiting argument.

A limit of a numerical sequence can be drawn in the same way. If the entries converge to some value, they will eventually get ε - close to this value if the index is bigger than the appropriate $1/\delta$. If not, for some subsequence the entries will stay away from that value. The "for all ε" part in the $\varepsilon - \delta$ condition is replaced by the possibility of choosing any finite subsequence, probing farther for greater precision.

What have we gained so far? Showing a limit value as a succession of values gives an intuitive idea of approximation. The viewer who is not yet convinced that the ultimate value is indeed what is asserted, may zoom in more times until he gets convinced. Beyond this final precision which he decides to be small enough, he no longer cares. We are quite close to the approach of the nonstandard analysis theory, where beyond some (finite) set of standard numbers, everything becomes infinitesimal, i.e. infinitesimally small. A precise definition is beyond the scope of this paper.

Something more important is obtained. The sequence of images will give a good idea about the speed with which the convergence happens. It is shown in the following examples that it is not easy to develop insight in the growth or decay of functions. A student who sees many images will most certainly profit from it.

Examples of growth.

1. Trying to graph the function $x \to 2^{2^x}$ on a graphics calculator, one runs immediately into problems. On the interval [-5,3], the graph looks nice but already for $x = 5$ the vertical size of a drawing with unit 1 cm should be more than the length of the earth equator! Indeed:

$$2^{32} \text{ cm} > 4 \times 1000^3 \text{ cm} = 40000 \text{ km}$$

2. Every mathematician should know that the harmonic series is divergent and that its partial sums tend to infinity as the logarithm. In fact, a much more precise estimate is possible: the difference between a partial sum and the corresponding logarithm tends to the Euler γ constant, but the sequence of differences tends to γ in a tantalisingly slow way!
3. A sequence which behaves well in many cases is the Newton approximation for the root of a polynomial or some other well-behaved function, starting from an initial guess which is close enough to this root. Each step in the approximation will add a couple of new correct decimals to the value of the root if the function is sufficiently smooth.
4. Definite integrals can be estimated efficiently by the Simpson method which replaces the function by a piecewise quadratic interpolation. The accuracy is of order four if the function has a bounded fourth derivative.
5. Interesting graphs arise if one mixes the (slow) growth of logarithmic functions with those of some power. The graphs of $\sqrt{x}\log x$, $x \log x$, $\log x / x$, $(\log x)^3 / x$ should be magnified many times to find out their exact behaviour near zero:

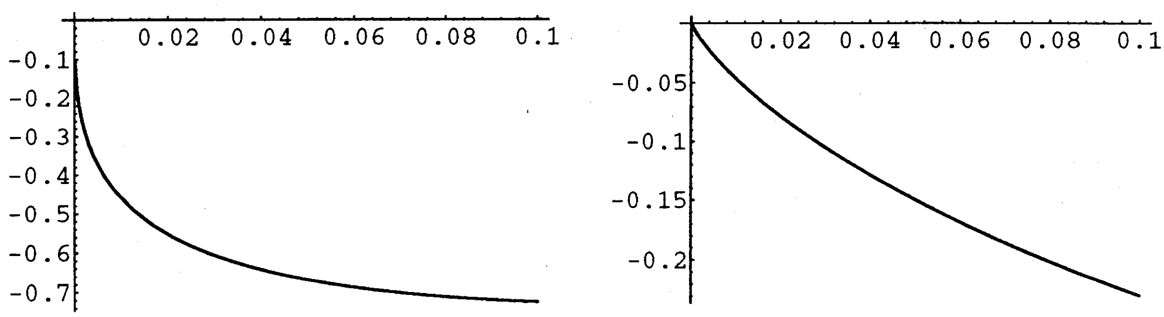


Fig 1: behaviour of $\sqrt{x}\log x$, $x \log x$ near zero

While the first magnification gives useful information near 0, the second graph is still not showing its behaviour which should be tangent to the y-axis. In some cases, graphs of funtions should be condensed far away in order to guess the behaviour near infinity:

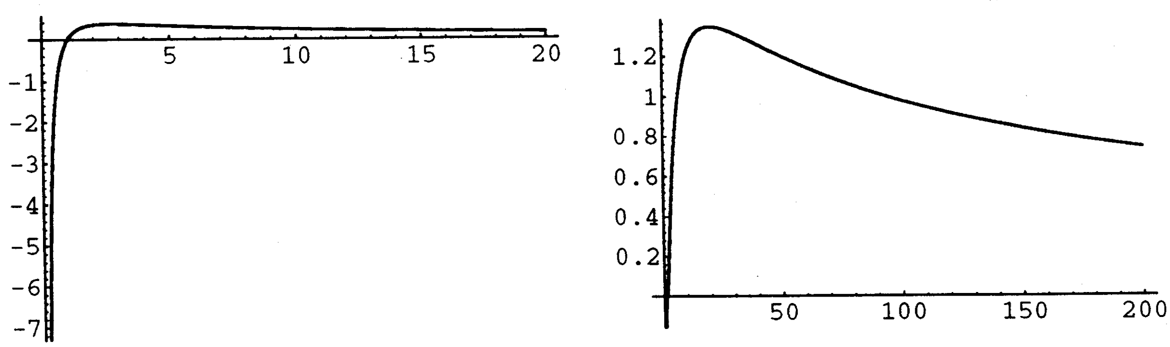


Fig 2: behaviour of $\frac{\log x}{x}$, $\frac{(\log x)^3}{x}$ at infinity

The behaviour is still worse if one mixes logarithms and exponentials as in $x^x \log(x)$. The following graph gives good information for the behaviour at infinity (left hand graph), but miss the fact that the function is tending to $-\infty$ near 0 since x^x tends to 1 near zero (right hand graph):

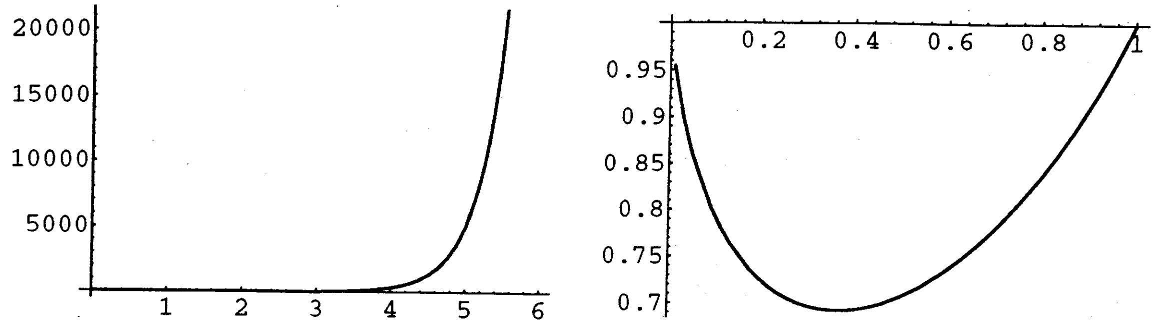


Fig 3: behaviour of $x^x \log x$ at infinity and of x^x near zero

Quantisation of limits

The notion of limit, such as in

$$\lim_{x \to +\infty} f(x) = +\infty$$

$$\lim_{x \to a;\, x > a} f(x) = +0$$

or

the limit of f(x,y) at the origin (0,0) does not exist,

does hide a lot of useful information. In many cases the way the function tends to its limit at a certain point is as interesting as the value itself. This is hidden by the definition in terms of ε and δ, or any topological definition. Proving that this definition is satisfied involves in many cases obtaining a lot of metric information in the form of inequalities which yield the right δ for a given ε .

The inverse of this correspondence, ε as a function of δ is called the modulus of continuity. In the case the ratio ε/δ is bounded, the function is Lipschitz of order 1 (at the given point x_0):

$$|f(x) - f(x_0)| < K\,|x - x_0|$$

and if moreover the constant K does not depend on the given point, the function is uniformly Lipschitz.

If this cannot be achieved, in many cases there is a Lipschitz-α $(0 < \alpha < 1)$ correspondence which can be obtained and which also implies the continuity of the function:

$$|f(x) - f(x_0)| < K\,|x - x_0|^{\alpha}$$

at the given point.

In cases where it is difficult to obtain such direct correspondence between e and d, the continuity or the existence and value of a limit is still established by comparing the function to other functions by setting up inequalities. Well-known examples of this technique are the computation of the limit values of $\sin x / x$, $x \log x$ and x^x at the point 0 , or $\log|x|/x$ at infinity.

Having at our disposal a collection of examples of functions exhibiting certain growth behavior is a reliable technique for computing such limits. Many teachers and students refer to l'Hospitals's rule for doing tricky limits. This is done automatically, without bothering if it makes sense to apply this rule or to check if the conditions for applying the rule are all valid. Moreover, this rule does not give an answer if the derivatives are not continuous or not monotonic, such as in

$$\sin x \log x / x$$

Moreover, the rule does not lead to generalisations.

We see that the role of inequalities is essential. It is difficult to learn students how to handle inequalities. One has to replace the exact value by some (smaller or larger) estimate and this involves the risk of losing information if the estimate is not chosen in the right way: if it is too coarse it may lead nowhere, if it is too precise the computations may become forbiddingly complicated. Either way, one will not be able to arrive at the goals set.

Learning inequalities is important. They arise in all subjects in mathematics. If one has gained experience in comparing growth of functions, one is able to solve a lot of problems. Let me give some examples from integration theory, discrete mathematics, and geometry.

If a positive function in one variable goes to infinity at some point more slowly than the function $\log |x|$ at 0, then the first function will be integrable near that point. In the case of several variables this growth estimate can be relaxed significantly. Such simple remarks are important in physics, where the integration of force fields is done.

For computer scientists analysing algorithms, it makes a tremendous difference wether the complexity (either worst-case or on the average) of a certain algorithm is proportional to $\log n$, n , n^2 , n^3 , 2^n , or even $n!$. In many cases the best he can hope for is an algorithm with a complexity not worse than $n \log n$ instead of n^2 and $n^{3/2}$ or $n^2 \log n$ instead of n^3 . Problems having worse complexities are untractable in practice. Fine analysis is carried out, including the estimates of the constants arising in the inequalities to gain insight in the best performance regarding execution time or memory usage.

In differential geometry, the notion of contact is important. The graph of a function makes contact of order ≥ 2 with the x-axis at 0 if the function remains below some constant times x^2 . This gives more information than the mere fact that its tangent is the x-axis at 0 . A graph going through the origin which is not Lipschitz-1/2 in the origin makes contact of order ≥ 2 with the y-axis.

By now, it should be clear for the reader that having ideas on how to compare growth and decay of functions is of utmost importance. Formal definitions can be given.

O( ),o( ) and Θ( ) notations.

The O and o notations were introduced by Landau to make precise statements about growth of functions. These notations are commonplace and universally adopted nowadays. Given two functions f and g , one has

$f \in O(g)$ (" f is dominated by g ")

if and only if f/g is bounded by some constant near the point under consideration, which may be finite or one of the points at infinity. Likewise,

$f \in o(g)$ (" f is small compared to g ")

if and only if f/g tends to 0 near the point under consideration.

These concepts are useful if the functions f and (or) g have zero or infinite limits at the point under consideration. They compare in a straightforward way the steepness (growth or decline) of f with that of g , and a lot of results are readily proved. The relations Œ O and Œ o are reflexive and transitive but not symmetric or antisymmetric:

$x \in o(x^2)$ at infinity but not conversely;

$x + 1 \in O(x - 1)$ and conversely, but the functions are unequal.

A more symmetric behavior is given by

$f \in \Theta(g)$ (" f is exactly of the order of g ")

which is true if and only if $f \in O(g)$ and $g \in O(f)$. In all these relations some constants arise and it is sometimes worthwile keeping track of them.

Another closely related notation was introduced by Du Bois-Reymond. His work was expanded by G. Hardy in a thin monograh [Ha] dating from 1910 and reprinted in 1971 because of the timeless value of its contents. It contains some 60 pages of rather strong results relative to the growth of logarithmic and exponential functions.

Motion and speed.

Most human activities (work, games, sports, traffic, television, ...) involve movement. Almost everything surrounding us is moving. We are therefore keen on seeing motion, and well-trained in finding moving targets and assessing their speeds and trajectories. We all catch a ball without making the appropriate computations of Newtonian mechanics.

A limit will be easier to understand if we can visualise it as a dynamic process. Limits are dynamic processes by their definition:

choose a continuous range of ε tending to 0, and watch the corresponding δ as they shrink to o

(if they do) or

select a sequence of x_n tending to x_o and watch the corresponding values of f

This visualisation can be done easily by modern computing software. It should be remarked that in this paper we only pay attention to the visualisation of limits. The reader interested in an overview of the effect of modern technology on other aspects of the teaching of mathematics should consult [Ho] or [CR].

In the case of a right hand limit at x_o we choose a sequence of intervals $[a_n,b_n]$ which are overlapping in the sense that $b_{n+1} \in [a_n,b_n]$ and have endpoints decreasing to x_o , set a range of the picture containing all these intervals and the ranges of the function on them, and plot the function on each interval. When these pictures are viewed as a rapid sequence we see the graph of the function going to its limit. If the limit exists it actually will attain the limit if the last picture in this sequence is constructed on an interval with x_o as its left endpoint. What I suggest is choosing a decreasing sequence b_n starting with a large number of values not too close and ending with values very close to x_0 , and setting $a_n= b_{n+2}$.

It may be the case that a first choice of sequences of endpoints gives a sequence of pictures which gets too rapidly or too slowly to the given limit. This makes the movie uninteresting, but this can be corrected by choosing a different frame speed. Of course one has to set up these choices in such a way that choosing different speeds can be done easily.

Since changing the viewing speed up (or down) the sequence of frames is always done by some lineair rule (each frame is shown the same fraction of a second), dramatic increases (or decreases) in speeds can only be achieved if this much faster (or slower) behavior is already built-in in the definition of the sequence b_n . This is the reason why the sequence should be non-linear. An example of such a sequence, which can mimic a large number of speeds is obtained by setting, for N sufficiently large (e. g. N = 100)

$$b_n = x_0 + \frac{1}{\tan(\frac{n}{2}\frac{\pi}{N})} \quad \text{if } n < N \text{ and } b_N = x_0$$

but many other choices are possible.

In fact getting the feeling for choosing a good sequence and setting the right speed gives us an estimate on how the limit is achieved.

Implementation.

A graphics pocket calculator offers the possibility of zooming in (a limited number of times on its screen. This feature can only be used for clarifying somewhat the behaviour near a particular point on the screen. It takes time both for setting the zoom and each screen buildup. What is offered on a graphics calculator with today's technology can hardly be called an animation.

An animation (movie) on a computer consists of a finite number of screen drawings which is shown at a set speed. The problem consists in reaching infinity in this way. Let us analyse this problem in detail.

1. The finite character of the whole setup can never be avoided. We cannot generate an infinite number of frames, it cannot be stored, we cannot show them and nobody is interested in seeing an infinite number of frames passing by.
2. More important than the finiteness of the sequence of frames is the choice of the values of the parameter for which a frame is made. If we choose in a standard fashion $\delta = 1/n$ for the sequences discussed above, this may be all right for illustrating certain phenomena for certain functions, but may give poor results for other phenomena or other functions. A simple example is the way the functions "square" and "square root" tend to zero on this sequence of δ .
3. These values should not be linear in the sense of n going to infinity, or not even as some exponential. While it is true that 2^n tends to infinity as n grows bigger, theoretically we have to wait for an infinitely long time to see this happen.
4. The viewer will be disappointed if he sees all frames running by at a fixed speed. Not only may the nature of a phenomenon pass by too rapid or too slow for him to see, but his ability to capture essential behaviour may depend on his experience.

5. The technical limitations make that we cannot show frames in very high speed succession. Moreover, the eye will not catch frames that are shown for less than five hundreds of a second.

The idea could be to produce, for each example, a large number of frames depending on a parameter which goes asymptotically to infinity on a finite interval, and show a subset of it at the user's request. The choice of a subset would be dictated by his setting of the speed. The limit situation is the one for which the image stabilises. If it is not reached on a choosen subset, a faster speed should be selected. It is clear that the movie has to be set up in such a way that if the full sequence is shown, the final image becomes very stable.

Showing frames at preset speeds may be dangerous for another reason now: if some phenomenon happens if a full sequence is shown, it may remain hidden on a given subsequence and will only be visible if the subsequence is changed, i.e. if the speed is set in a different way. We all know a lot of sequences having subsequences behaving in a peculiar way. This can be avoided by:

- hiding the parameter choice from the user;
- choosing a parameter which uses a transcendental function as the tangent function above (so that it is more difficult for the viewer to guess);
- encouraging the viewing under several choices of viewing speeds;
- suggesting several choices of viewing speeds (for which the phenomena are most explicit).

The user may even learn something from his setting of different speeds. For example: if he sees a certain behaviour for the function square root at a certain speed, he will see the behaviour of the function square if the speed setting relates to the original as $n \to n^4$. Playing around with different speeds will give him some insight into orders of magnitude. He will learn, among other things the tremendous difference between the polynomial orders of growth, logaritms and exponentials. Depending on the example the viewer is offered certain preset speeds, for which the interesting behaviour is apparent. Suppose now he wants to set another speed. A speed is any increasing function from a bounded interval (say $[0,1]$) to the interval $[1,+\infty]$. If the choice of the frame parameter reduces the image to this interval, we can define a speed as any increasing function from $[0,1]$ to $[0,1]$. The viewer has a tremendous freedom of choice now: he can trace any increasing function in a square and the sequence of frames is chosen according to it.

To illustrate the importance of right choices of speeds, the appendix contains the code for generating frames how $\sqrt{x}\sin(1/x)$ and $\sqrt{x^3}\sin(1/x)$ go to 0 on the positive real axis at 0 for the speed using 1/n. The resulting movie is slow for one function and rather rapid for the other.

The above methodology can be used not only to show limits and growth of functions, but also to show to beginning students a whole range of subjects in real analysis. Most subjects can be illustrated by sequences of pictures (but not by a single picture). Let me list the following:

- least upper (and greatest lower) bounds;
- summation of series;
- uniform continuity;
- approximations of the definite integral;
- curve length, surfaces and volumes by giving suited approximations;
- numerical approximations of the integral;
- derivatives by using cords or linear approximation of functions;
- derivative and integral as inverse operations;
- approximation of functions by polynomials of higher order;
- pointwise and uniform convergence;
- series of functions;
- approximations by Fourier theory;
- convergence in the C^1 topology, i.e. together with the derivative;
- two dimensional limits, limits along a curve in the plane;

and many more phenomena in analysis, which nowadays are forbiddingly difficult for our beginning students to grasp. My favourite is the approximation of distributions by smooth functions. Generalised functions or distributions in the sense of L. Schwartz [S] are, in fact, sequences of smooth functions and therefore best visualised by the animation of this sequence.

References.

[AH] Appel, K. and Haken, W., Every Planar Map is Four Colorable, Bulletin of the American Mathematical Society 82, 1976.

[B] Bourbaki, N., Elements de Mathématique, Hermann, 1960-1980.

[Ce] Cech, E. , Topological Spaces, J. Wiley-Interscience, 1966.

[CR] Cornu, B. and Ralston, A., The influence of computers and informatics on Mathematics and its teaching, UNESCO studies Science and technology education, 44, 1992.

[Ha] Hardy, G.H., Orders of infinity. Hafner Publshing, New York, 2nd ed,1971.

[Ho] Howson, G. et al, Mathematics as a service subject, ICMI Study Series, 1987.

[L] Lee, Peng Yee, Calculus, National University of Singapore, 1993

[M] Mandelbrot, B.B., Fractals, Form Chance and Dimension, Freeman 1977.

[PW] Packel, E. and Wagon, S., Animating Calculus (Mathematica Notebooks for the laboratory), W. H. Freeman & Co, 1994.

[R] Robert, A., Nonstandard Analysis, J. Wiley, 1988.

[S] Schwartz, L., Théorie des Distributions , Hermann, 2nd ed 1966.

[W] Wolfram, S., Mathematica, A System for Doing Mathematics by Computer, Addison-Wesley, 2nd ed,1991.

Appendix: Mathematica™ code.

```
m=20
a= Table [1/j,{j,1,m}]

f[x_] = x^3/2 Sin[1/x]
h[j_,x_] = If[a[[j-1]]<x<a[[j+1]],f[x],]
H = Table[Plot[h[j,x],{x,0,1},
          PlotRange->{{0,1},{-1,1}}]
     ,{j,1,m-1}]
Show[Release[H]]
```

```
m=20
a= Table [1/j,{j,1,m}]

f[x_] = x^1/2 Sin[1/x]
h[j_,x_] = If[a[[j-1]]<x<a[[j+1]],f[x],]
H = Table[Plot[h[j,x],{x,0,1},
          PlotRange->{{0,1},{-1,1}}]
     ,{j,1,m-1}]
Show[Release[H]]
```

■ Sample dialog

Double-click small rectangles to the right to open each dialog area

□ Choose Point x0

□ Choose left=0 or 1

□ Choose right = 0 or 1

□ Choose horizontal range

□ Choose vertical range

□ Choose number of frames

□ Choose speed

□ Choose function of x

Finally, click in the rightmost margin to see the following generation of frames:

```
alfa[k_]:=alfa[k]=N[(Tan[Pi/4 (number-k)/number])^speed];
Table[alfa[k], {k,1,number-1}]

gL=Table[
     If[left>0,
          Plot[f[x],{x,x0-alfa[k-1],x0-alfa[k+1]}
               ,DisplayFunction->Identity],
          Plot[{},{x,x0-alfa[k-1],x0-alfa[k+1]}
               ,DisplayFunction->Identity]
          ]
     ,{k,2,number-1}];
gR=Table[
     If[right>0,
          Plot[f[x],{x,x0+alfa[k-1],x0+alfa[k+1]}
               ,DisplayFunction->Identity],
          Plot[{},{x,x0+alfa[k-1],x0+alfa[k+1]}
               ,DisplayFunction->Identity]
          ]
     ,{k,2,number-1}];

Do[Show[{gL[[j]],gR[[j]]}
     ,PlotRange->{{x0-hor,x0+hor},
               {N[f[x0]]-vert,N[f[x0]]+vert}}
     ,DisplayFunction->$DisplayFunction]
     ,{j,2,number-1}]
```

Paper presented for the First ATCM meeting in Singapore

Abstract

Many mathematical concepts in analysis (and in most other subfields of mathematics, except perhaps finite mathematics) involve "change" and some kind of limiting process. This makes them difficult to understand for many students both in secondary and tertiary education. Nowadays such limiting processes can be simulated using animations available in computer algebra packages running on all state of the art personal computers.
In the implementation of such simulations it is necessary to avoid creating artifacts due to the finiteness of both soft- and hardware. The user should be able to choose the speed and other settings involved in the simulation, and should be able to create his own examples according to the same template. If this is achieved, the animations can greatly improve understanding of the concepts and ideas involved. Examples of this methodology are given using Mathematica™ .

SOME FOUNDATIONAL ISSUES, ILLUSTRATED BY CONSIDERATION OF TECHNOLOGIES FOR THE TEACHING OF: (1) CALCULUS USING INFINITESIMALS AND (2) MATHEMATICAL LOGIC USING FUZZY LOGICS

Don Faust
Department of Mathematics and Computer Science, Northern Michigan University
Marquette, Michigan 49855, USA
current mailing address (1995 and 1996): IKATAN (TNB)
Jalan Serdang, 43009 Kajang, Selangor, MALAYSIA

There is an exciting synergy surfacing between mathematics pedagogy and the various technologies being brought to bear in the teaching of mathematics. For example, the computer-based technologies for teaching mathematics have played an important role in helping us: (A) to see more clearly the pedagogical advantages of having students look carefully at examples and applications of the concepts they are trying to understand; and, (B) to more carefully analyze *the various processes involved in the learning of mathematical concepts and applications*, and thereby to build improved pathways over which this learning can take place. The latter has been particularly facilitated through efforts to build Intelligent Computer Assisted Instruction (ICAI) systems, since the construction of such systems focuses our attention on the finer detail of what learning involves, and the insights gained have been studied and successfully used by workers in mathematics pedagogy to improve our understanding of the processes involved in the learning of mathematics.

At the same time, it is well to remember that at the base, of these efforts to improve the teaching of mathematics with the appropriate use of new technologies, is indeed *mathematics itself*. Keeping this in mind, part of our effort should be directed toward

> (*) looking at recent new ideas in mathematics itself and, with an eye toward the future implementation of these new ideas using computer-based technologies, considering both (a) the possible pedagogical value of these ideas and (b) how computer-based technologies may profitably be used in implementing these ideas.

In this paper we will briefly exemplify effort (*) by looking at two relatively new and fundamentally important ideas in mathematics: (1) infinitesimals, as rigorized by Abraham Robinson, and (2) nonstandard logics, for the representation and processing of uncertain information. Following these examples, in Section 3 we will consider some of the broader issues surrounding effort (*) made apparent by the examples.

1. Teaching Calculus Using Infinitesimals

[Abraham Robinson was] the one mathematical logician who accomplished incomparably more than anybody else in making this science fruitful for mathematics. I am sure his name will be remembered by mathematicians for centuries.

Kurt Goedel

The new foundation for the Calculus, based on the seminal work of Abraham Robinson (see [Robinson 1961 and 1966]) but now being widely investigated and used (see [Henle & Kleinberg and Keisler]), is truly remarkable. The informal use of infinitesimals in learning and applying the Calculus has indeed been the technology, in a broad sense of the term, which has been used intuitively by all learners and users of the Calculus for over 300 years, beginning with Newton and Leibniz. However, for 200 years, roughly 1650-1850, the Calculus existed and was used without its having any foundation at all. The first rigorization of the Calculus, it should be recalled in order to have sufficient perspective about this matter, came around the mid-nineteenth century (the so-called ε–δ approach) and it took about seventy years for it to become standard in the elementary Calculus. With this perspective, and a familiarity with *the very clear way in which the infinitesimal approach provides a formalism which, unlike the ε–δ approach, nicely matches the intuitive way we think about the Calculus*, let us consider the following possibility: it may indeed take a rather long period of time (seventy years, again?) for the infinitesimal approach to supplant the ε–δ approach, but this is ultimately what will happen if our education systems, however slowly, move to incorporate this superior learning technology. It may in fact take longer, since the ε–δ approach was providing a foundation where none previously existed whereas the infinitesimal approach is providing a pedagogically superior potential replacement foundation for a foundation (the ε–δ approach) which is already in place. In this regard, the reader is encouraged to read the concluding Chapter X of [Robinson 1966], which provides a truly excellent historical perspective on the Calculus.

Let us turn briefly to giving some indication of this infinitesimal approach to the Calculus; the reader is encouraged to consult [Robinson 1966, Keisler 1976, and Faust 1981 and 1985] for further details. Using techniques of mathematical logic and model theory, Robinson was able to construct the hyperreal numbers H which contains the real numbers R, contains nonzero infinitesimals (numbers infinitely close to 0, i.e. numbers whose absolute value is both positive and yet less than every positive real number), contains numbers infinitely close to *each* real number x which are yet different from x (numbers which differ from x by a nonzero infinitesimal; note that in our presentation here zero is considered to be an infinitesimal), and contains both negative and positive infinite numbers (numbers whose reciprocals are some nonzero infinitesimal). Consider the following four propositions, easily provable within any of the usual formal frameworks for the infinitesimal Calculus, where $a \sim b$ (for a and b in H) means that a and b are infinitely close (i.e. a-b is an infinitesimal).

Proposition 1. ~ is an equivalence relation.
Proposition 2. If u and v are infinitesimal, then so is u + v.
Proposition 3. If u is an infinitesimal and x is a finite hyperreal, then ux is infinitesimal.
Proposition 4. If a, b, c, and d are finite hyperreals with a~b and c~d, then
a+c ~ b+d and ac ~ bd.

Already, from even these simple propositions, one can begin to get a flavor of infinitesimals and how the Calculus proceeds if using them. For example, consider

limit f(x) = L,
x --> c

which in the infinitesimal approach formalism means: for all x, if x is infinitely close to c (but x is unequal to c), then f(x) is infinitely close to L. In particular, consider

limit (3x+4) = 10.
x --> 2

In the ε–δ approach, we of course *talk* about this in the intuitive way it has been *thought* about for 300 years: as x gets closer and closer to 2 (with in general, however, x not equal to 2), 3x+4 gets closer and closer to 10. Then we prove it, and other more complicated limits as well, depending on the instructor's love for the ε–δ formalism and the pain threshhold of the students, for in droves they *do* find the ε–δ formalism painfully assistive at best (or even downright unhelpful): let ε > 0; let us seek a δ > 0 such that .... Enough said! In the infinitesimal approach, in sharp contrast to the ε–δ approach, the formalism matches nicely with the intuitions the student is trying to construct and build her/his understanding upon. The proof , in the infinitesimal approach formalism, is simply as follows:

x ~ 2 (and x not equal to 2) ===> 3x ~ 6 by Proposition 4
===> 3x+4 ~ 10 by Proposition 4.

Regarding continuity, in the infinitesimal approach f(x) is continuous at x=c if (for all x) x ~ c ==> f(x) ~ f(c). In integration, to quickly mention something further on in the Calculus, in the infinitesimal approach a Riemann sum is actually an infinite sum of rectangles each of which is of nonzero infinitesimal width.

A number of efforts in the direction of computer-based implementation of nonstandard analysis have taken place, for example the work of Jensen [see Jensen 1972]. In this brief overview I would like simply to make some general remarks about such implementations, hoping to stress pedagogical aspects of these efforts and to provide a bit more insight into the nature of the hyperreals.

Extensions of the currently widely used Calculus packages, such as Maple or Mathematica, might be investigated, which incorporate the infinitesimal approach and allow the student to interact with the computer in manipulating the basic concepts of the Calculus utilizing this infinitesimal spproach which so well matches our intuitions. For example, the algebra of the congruence relation "~" described briefly above might be axiomatized (programmed) into the package. Then, for instance, the student could interact with the package in the computation of the derivative of $f(x) = x^2$, possibly as follows:

$$h \sim 0, \text{ h not equal to } 0 \implies \frac{(x+h)^2 - x^2}{h} = \frac{x^2 + 2xh + h^2 - x^2}{h} = 2x + h \sim 2x,$$

and hence $f'(x) = 2x$. Such packages, including symbolic computation with infinitesimals, provide a formalism which indeed matches the intuitions the students are building upon.

Beyond the level of symbolic computation just mentioned lies the implementation of the hyperreal numbers themselves. To gain some further insight into the nature of the hyperreals, let us briefly consider this as well. Of course, to keep our expectations reasonable, we should begin by recalling that naturally only a finite subset of the reals is actually implemented in any implementation the reals, so that we should expect the same serious limitation to be present in any implementation of the hyperreals as well. Now, the standard realization of the hyperreals views the hyperreals as infinite sequences of reals with equality meaning agreement on a 'big' set (an element of a fixed ultrafilter on the set of positive integers). This definition of equality among hyperreals makes the implementation of this realization quite hopeless since every element of the ultrafilter is an infinite set. Another common realization of the hyperreals views the hyperreals as infinite decimals of order type $\omega + \omega^* + \omega$, where as usual ω is the order type of $\langle 0,1,2,...\rangle$ while ω^* is the order type of $\langle ...,2,1,0\rangle$. In this realization the first ω-sequence codes the real part of the hyperreal while the subsequent $\omega^* + \omega$ - sequence codes the infinitesimal part of the hyperreal. This realization is of course problematic as well for any finitistic implementation, but it offers some interesting possibilities the consideration of which, in this short overview, we must leave for the reader.

To understand a bit more about the nature of the hyperreals, note first that one can think of the finite hyperreal line as having been gotten from the real line by adding, as we have briefly sketched above, new numbers (nonreal of course!) infinitely close to each real number. In fact, as the Propositions above, particularly Proposition 3, imply, for each real x lots of new numbers are added which are infinitely close to x. Intuitively one can think of these new numbers infinitely close to x as forming a 'cloud' of hyperreals with x at its center, what is called in nonstandard analysis the monad of x. In fact, any finite hyperreal x is in the monad of some real x, and hence $x = x + u$ for some infinitesimal u. Again, to increase our intuitional understanding of the hyperreals, let us consider some aspects of an implementation based on this representation of the hyperreals. The most direct thing to do, namely implementing a hyperreal as a pair $\langle x,u\rangle$, will not do at all (the reader can work this out for her/himself or peek at the example discussed below). So let's consider the implementation of the hyperreals as (n+1)-tuples of reals (that is, of course, of the finite subset of implemented reals) where the first coordinate is the real part of the hyperreal and the n subsequent coordinates code the infinitesimal part of the hyperreal. On this set of tuples, let us stipulate that the ordering is defined to be lexicographic and addition is defined coordinatewise. Defining a product is more interesting and, to elucidate better what is involved, let us make first a rather poor choice for a product: for represented hyperreals $x = \langle x,u_1,\ldots,u_n \rangle$ and $y = \langle y,v_1,\ldots,v_n \rangle$, let the product be the *compacted polynomial product* $z = \langle z,w_1,\ldots,w_n \rangle$ where $z = xy$ and either n=1 and $w_1 = u_1y + v_1x + u_1v_1$ or $n > 1$ and $w_1 = u_1y + v_1x$ while for all k = 2,...,n-1

$$w_k = u_{k-1}v_{k-1} + u_k y + u_k \sum_{i=1}^{k-1} v_i + v_k x + v_k \sum_{i=1}^{k-1} u_i$$

$$\text{and } w_n = u_{n-1}v_{n-1} + u_n y + u_n \sum_{i=1}^{n-1} v_i + v_n x + v_n \sum_{i=1}^{n-1} u_i + u_n v_n .$$

With this product we are considering a representation for the infinitesimal part which preserves at least some of the arithmetic of the hyperreals and yet requires no more space for the product (just an (n+1)-tuple) than is used for representing the hyperreals themselves. Some insight into the problems here can be gotten by considering the following example, illustrating why ordered pairs (i.e. use of a single infinitesimal coordinate) provides a poor implementation and why increasing the size of the tuples improves the implementation in a way very similar to the way in which increasing the number of machine decimal places improves the implementation of the reals. Since the sentence "for all x such that $0<x<1$, $x^2<x$" holds in the reals, and the hyperreals are an elementary extension of the reals, this sentence holds also in the hyperreals. But in an implementation with n=1, $\langle 0,0 \rangle < \langle 0,3 \rangle < \langle 1,0 \rangle$ while $\langle 0,3 \rangle^2$ is, with the compacted polynomial product, $\langle 0,9 \rangle$ which is not less than $\langle 0,3 \rangle$. In contrast, however, in an implementation with n=2 (i.e. two infinitesimal coordinates), $\langle 0,3 \rangle$ is $\langle 0,3,0 \rangle$ and $\langle 0,3,0 \rangle^2$ is $\langle 0,0,9 \rangle$ under the compacted polynomial product. But then, with n=2, other problems now arise (viz. $\langle 0,0,3 \rangle^2$) .... These considerations lead to defining the product in a less miserly fashion, allowing a (2n+1)-tuple (instead of an (n+1)-tuple) for the product of two (n+1)-tuples, as the familiar *formal polynomial product*, where the hyperreals $x = \langle x_0,x_1,\ldots,x_n \rangle$ and $y = \langle y_0,y_1,\ldots,y_n \rangle$ (with real parts x_0 and y_0 respectively) have product $z = \langle z_0,z_1,\ldots,z_{2n} \rangle$ where for all k = 0,1,...,2n

$$z_k = \sum_{\substack{i+j=k \\ 0 \le i,j \le n}} x_i y_j .$$

Note that this latter product, although it provides a representation better suited to preserving the arithmetic of hyperreals, requires a data structure essentially twice as large -- a characteristic this multiplication shares with the usual machine implementation of real number multiplication. In fact, of course, real multiplication is essentially formal polynomial multiplication with reals represented as polynomials of positive and negative powers of ten. The analogy is in fact closer: for in both the usual computer realization of the reals and the above computer realization of the hyperreals longer representations give rise to better representations and less inaccurate arithmetic.

Just as with the implementation of the reals where space / time constraints determine the number of decimal places implemented (and hence the size of the actual finite subset of the reals which gets implemented), here also with the hyperreals space / time constraints will determine n. Note also that attempting to mitigate problems like those exemplified above by using the reals in the open interval from -1 to 1, instead of all the reals, in the infinitesimal coordinates, would make problematic the closure we naturally want under addition and multiplication. Thus we have gained some initial understanding of the infinitesimals by getting a taste of the hyperreals and how they are used in the Calculus together with a bit of the interesting nature of the problem of representing reasonably, while of course necessarily incompletely, the essentially infinitistic hyperreals in a finitistic manner.

2. Teaching Mathematical Logic Using Fuzzy Logics

Vagueness and precision alike are characteristics which can only belong to a representation .... They have to do with a representation and that which it represents. Apart from representation, whether cognitive or mechanical, there can be no such thing as vagueness or precision; things are what they are, and there is an end of it. Nothing is more or less what it is, or to a certain extent possessed of the properties which it possesses.

Bertrand Russell, p. 84

During the nineteenth century logic progressed greatly; in this regard Boole and Frege are often and rightly mentioned, the former in regard to the algebra of logic and the latter in regard to the predicate calculus. Further development of logic has continued throughout our century, two of the often cited leaders being Goedel and Tarski. The semantic side of logic has now blossomed, giving rise to the rich and fruitful area of model theory. One example of this fruitfulness is so close at hand as Section 1 above, the infinitesimal approach to the Calculus as rigorized by Robinson with the model H for the hyperreal numbers. Other examples are the nonstandard models of number theory and the topological spaces (Stone Spaces) of models of formal languages used so fruitfully in studying the algebraic and recursive structure of formal languages (for the latter, see e.g. [Faust 1982]).

In parallel with these developments, but overshadowed until recently by a bias in favor of two-valued logic strongly advocated by Frege and his followers, has been the steady development of logics which seek to go beyond two-valued logic, which I shall here rather roughly refer to as many-valued logics or fuzzy logics. The Polish logicians made seminal contributions to many-valued logics in the early part of our century. It is coming to be realized that modeling our knowledge with a two-valued logic is hardly adequate: our knowledge is often less than certain, often evidential and a matter of degree, yet we must be able to represent and process this knowledge; more important still, if we are to move forward with the construction of more intelligent computer-based systems, we must develop (and understand well) logical formalisms which are rich enough to handle such uncertainty and which are amenable to machine implementation (see [Faust 1990 and 1992]).

As a typical example of such a logic, upon which we can base our discussions here, we will now briefly describe Evidence Logic (EL) (see [Faust 1994]). In EL predications can be made with an attached confidence or evidence level. These levels arise naturally in practice, for example as measures of the degree of evidence confirmatory of, or refutatory of, the circumstance described by the predication. For evidence levels, let us use here real numbers in the closed interval from 0 to 1, where 0 denotes a total lack of evidence and 1 denotes complete evidence. So for example, consider a unary predicate (property) P. P_ca:.7 asserts that there is evidence at the .7 level confirming that the object a has property P, while P_ra:.4 asserts that there is evidence at the .4 level refuting that the object a has property P. It is important to notice the distinction between, for example, *the absence of evidence at the .4 level confirming that a has property P* (which is asserted by $\neg$ P_ca:.4) and *the presence of evidence at the .4 level refuting that a has property P* (which is asserted by P_ra:.4); in fact, the ability to represent and process this distinction is what makes EL an enrichment of classical logic, an enrichment which provides for a broader scope of knowledge representation and logical analysis.

In application, EL functions much like any logical system, except that its ability to represent evidence levels considerably increases its utility and its potential in the direction of construction of kernel inference engines flexible enough to handle uncertainty in the complex ways inherent in intelligent behavior. Here is an example of such an environment: data enters the system, for example from analogue sensors of various types if the system is monitoring some physical process, or from the human learner if the system is some sort of ICAI tool; preprocessors analyse the data, determining what evidential knowledge is supplied by what is present (and what is absent) in the data (e.g., the data provides .7-confirmation that the process has such and such a characteristic, or the data provides .8-confirmation that the learner does not yet understand the concept of a function); these evidential statements, the representation of which EL is particularly suited for, then become the "time t" axioms for the EL inference engine; deductions are produced by the inference engine, using the current axioms together with further underlying axioms which reflect the rules and goals of the system; these deductions give rise to actions (e.g., further testing or modification of the physical process, or further appropriately chosen activities for the human learner); then the cycle is repeated, beginning again with data acquisition.

On the implementation side, EL has been prototyped, and experimented with, in Prolog by simulating the enriched predications of EL with Prolog predicates containing extra arguments which code whether the predication is confirmatory or refutatory and its evidence level. Formal analysis of EL, namely in regard to the Boolean algebras of sentences which arise under the various stipulations of predicates for the language and an arbitrary finite linearly ordered space of n evidence levels (n greater than or equal to 2), has been accomplished and is reported in [Faust 1994]. In implementation, the reals are of course an inappropriate choice as a data type to represent evidence levels, since tests for equality of evidence levels will commonly occur as an integral part of the resolution-based logical processing of EL; hence, in the prototype developed we used the integer subrange 0..MAXINT for evidence levels.

But beyond this much remains to be done. Certainly logic is becoming a more important tool in science, as evidenced in the initial paragraphs of this section. It may well be that some understanding of mathematical logic will become a standard part of many undergraduate majors. For, science majors generally, as well as business majors and many liberal arts majors such as pre-law and pre-medical/dental, will find their fields making heavy use of *model-building*, the construction and use of sophisticated knowledge representation and modeling schemes to help them make reasonable use of the burgeoning data in their field. Business and law are certainly good examples. Increasingly, expert system based *advisors* will be used routinely on a daily basis. A rudimentary exposure to logic and model theory will provide a conceptual base upon which users of these systems can erect for themselves a basic understanding of what these systems are (and are not!) doing, and hence they will be better able to make use of them in appropriate ways. Tools for teaching logic, such as Tarski's World, can form a base over which to erect teaching tools for such logics as EL, logics which go needfully beyond classical logic and reach out to encompass knowledge as it so commonly exists in the real world, as for example EL provides machinery for representing and processing the uncertainties which indeed pervade real world knowledge.

Finally, let us note that this progress in logic and in tools for teaching logic is in fact a *double-edged sword.* For, beyond the discipline of logic itself and its growing utility in individual disciplines lies *the relevance of logic to the construction of computer-based intelligent systems and, in regard to teaching and learning, ICAI systems for <u>any and all disciplines</u>*: as is now widely appreciated, computer-based systems of logic will increasingly form an integral part of the kernel of computer-based intelligent systems. Especially important will be logical systems, like EL, which incorporate technologies for the representation and processing of uncertainties, since so much of our intelligent behavior is dependent on our ability to handle them.

3. Some Foundational Considerations

The expansion of number systems has occured often in the history of mathematics and has usually marked a major turning point .... At each step the growth was the result of a need .... At each step expansion of the number system met with opposition, and at each step the new numbers were formally accepted long before they were given the status of numbers.

Henle & Kleinberg, pp. 7,9

In Sections 1 and 2 above we have surveyed two relatively recent mathematical ideas. In this section we will briefly explore what these examples might have to say about the implementation of technology in the teaching of mathematics.

In our first example we saw that with the infinitesimals we extend the reals, and with this larger set of numbers we find a new formal framework which matches well with our intuitions and aids us in increasing our understanding of the limit processes which are at the basis of the Calculus. In the second example we saw that with a variety of number systems which extend the classical set of truth values {0,1}, we find new formal frameworks for logic which extend classical logic and which incorporate mechanisms for representing and processing the labyrinth of uncertain knowledge usually found in the real world. Both examples involve, then, an expansion of number systems and, as we noted in each case, and exemplifying the quotation above, there is a clear historical record of considerable resistance to their introduction. Yet, their constructions are each clearly based on real needs: in the first case a need for a simple yet rigorous foundation for the Calculus which matches well the intuitive way the Calculus has always been worked with; in the second case a need for logical systems which, unlike classical logic, can represent and grapple with the evidential nature of our knowledge of the real world.

This problem, involving a reticence to the implementation of new and clearly superior mathematical ideas, may indeed be somewhat alleviated by our simply being aware of it. When we contemplate this problem in terms of the present melieu of computer-based implementation of new teaching and learning technologies, it is readily seen that it dovetails with

a) the increased specialization of most producers of new mathematical ideas, and

b) the increased specialization of most implementors of computer-based mathematics teaching and learning technologies.

There is certainly a great challenge for us here, for in spite of these problems, the pedagogical value of substantial new mathematical ideas such as we have exemplified in Sections 1 and 2 deserve investigation, especially in terms of how computer-based learning technologies might profitably utilize these ideas.

To meet this challenge, we need to encourage inter-disciplinary research and communication. We need to carefully nurture flexible inter-disciplinary programs which will produce a wide variety of "generalist mathematicians" who can catalyze communication and collaborative efforts between workers in mathematics pedagogy and workers in computer-based tools for teaching and learning mathematics, and who can help to increase the awareness of the producers of new mathematical ideas as to the needs of workers in pedagogy and implementation so that some of their effort can be better focused on meeting these needs.

Especially in this time of rapid expansion of computer-based technologies, we can benefit from a broad perspective which recognizes the long history of development of technologies for mathematics, reaching back to the earliest number systems, including even the lunar notation systems of the Upper Paleolithic (see [Faust 1989, Section 1.1]). Thereby, we can see the importance of working constantly to bring together the producers of new mathematical ideas, the workers in mathematics pedagogy, and the implementors of computer-based tools, whose combined efforts will produce the superior learning and teaching technologies of the future.

REFERENCES

Faust, Don, (*abstract*) "The concept of evidence", Journal of Symbolic Logic 59 (March 1994), 347-8.

_________,"Observations on the Construction of a Foundation for the Handling of Uncertainty", Proceedings of the Asian Mathematics Conference 1990, World Scientific Press 1992, 140-143.

_________, "The Representation and Processing of Uncertain Information: Status Report and Prospectus for Information Technology Applications", Proceedings of the International Conference on Information Technology (ICIT-90), E 275 - E 295.

_________, "What is Mathematics? Another Example: Computer Science", International Journal of Mathematical Education in Science and Technology 20 (1989), 307-315.

_________, "Infinitely Small Numbers and Infinitely Large Numbers: the Hyperreal Number System", Southeast Asian Bulletin of Mathematics 9 (1985), 40-49.

_________, "The Boolean Algebra of Formulas of First-Order Logic", Annals of Mathematical Logic 23 (1982), 27-53.

_________, Calculus Using Infinitesimals, 1981 (available from the author).

Henle, J., & E. Kleinberg, Infinitesimal Calculus, MIT Press 1979.

Jensen, Anton, "A Computer Oriented Version of 'Nonstandard Analysis'", Contributions to Nonstandard Analysis (ed. by W. A. J. Luxemburg and A. Robinson), North-Holland Publ. Co. 1972, 281-289.

Keisler, H. Jerome, Elementary Calculus, PWS Publ. Co. 1976.

_______________, "Reshuffling Theorems for Finite Markov Processes", AMS Notices (January 1980).

Robinson, Abraham, "Nonstandard Analysis", Proc. Royal Acad. of Science A64 (1961).

_________________, Nonstandard Analysis, North-Holland Publ. Co. 1966.

Russell, Bertrand, "Vagueness", Australian Journal of Phil. and Psych. (1923), 84-92.

THE MAPLE COMPUTER ALGEBRA SYSTEM

Keith Geddes and George Labahn
Department of Computer Science
University of Waterloo, Waterloo, Ontario, Canada
e–mail: {kogeddes, glabahn}@daisy.uwaterloo.ca

1 Introduction

Maple is a computer algebra system that provides a unified environment for symbolic, numeric, and graphical computation. It has its own programming language, an interactive environment, a library of mathematical software, and both published and on-line documentation. It comes with a large library of specialized mathematical packages. In this paper we will give a brief introduction to some of the capabilities of the Maple computer algebra system.

Maple developed out of a research project started in 1980 by Gaston Gonnet and Keith Geddes at the University of Waterloo. The central focus of the project has been, and continues to be, the design and implementation of the Maple computer algebra system [1]. The original motivation of the project was to create a computer algebra system that was able to run on the smaller computers that were starting to gain prominence at that time. Maple has a small compiled kernel written in C with the remaining mathematical knowledge coded using its own user-level programming language. The small kernel facilitates portability over a wide range of computer platforms (Unix, Windows, MacIntosh) and allows Maple to run on computers with only modest memory capacity. Its large mathematical library, on the other hand, allows Maple to be powerful enough to meet the mathematical needs of a wide range of users. Maple has a large world-wide base of users in both research laboratories and classrooms.

2 Basic Facilities

2.1 Numerical Computation

The arithmetic supported by systems such as Maple is exact arithmetic rather than the floating-point arithmetic used by typical calculators. In addition, the arithmetic is not limited to the few digits that one encounters with calculators.

```
> 42!;
        1405006117752879898543142606244511569936384000000000
```

```
> ifactor( " );
```

$$(2)^{39}(3)^{19}(5)^{9}(7)^{6}(11)^{3}(13)^{3}(17)^{2}(19)^{2}(23)(29)(31)(37)(41)$$

```
> ( 1117 / 23125 + 441/7)^3 - 1/5;
```

$$\frac{30993104206529198 63}{12366455078125}$$

Real numbers are not required to be floating-point numbers in Maple; rather they can be symbolic quantities. Of course, such symbols have properties known to the Maple system. For example,

```
> sqrt( 2 );
```

$$\sqrt{2}$$

```
> " ^ 2 ;
```

$$2$$

and

```
> sin( 3*Pi );
```

$$0$$

In the above examples " denotes the previous expression. One can also obtain floating-point approximations to real quantities using the command **evalf** (for evaluation over the floating-point numbers):

```
> Pi^2;
```

$$\pi^2$$

```
> evalf( " );
```

$$9.869604404$$

Maple also supports arbitrary-precision arithmetic, that is, the number of digits to be carried can be set by the user. (The default is 10 digits.) For example, a more accurate decimal expansion of π^2 is given by

```
> evalf( "" , 50 );
```

$$9.8696044010893586188344909998761511353136994072408$$

Additional number systems handled by Maple include complex numbers

```
> ( 2 + 3 * I ) / ( 4 - 5 * I );
```

$$-\frac{7}{41}+\frac{22}{41}I$$

where I denotes the square root of -1, and integers modulo m for a specified integer m.

When computing in exact symbolic mode, one may wish to compute exact roots of polynomials and to manipulate such quantities. It is not sufficient to represent such values by numerical approximations. In mathematics there is a formal structure, namely algebraic extension fields, for the representation and manipulation of roots of polynomials. In Maple, algebraic number arithmetic is invoked via the `evala` evaluator. For example, we can let α denote a root of the irreducible polynomial $x^5 + 2x + 2$ by

```
>  alias( alpha = RootOf( x^5 + 2*x + 2, x ) ):
```

and then compute the following expression in $\mathbf{Q}(\alpha)$:

```
>  (alpha^4 - alpha + 3)^2;
```

$$\left(\alpha^4 - \alpha + 3\right)^2$$

```
>   evala( " );
```

$$4\,\alpha^4 - 2\,\alpha^3 + \alpha^2 - 2\,\alpha + 13$$

Computer algebra algorithms that require the manipulation of roots of polynomials (e.g. integration of rational functions, laplace transforms, etc.) often make use of the `RootOf` construct.

2.2 Symbolic Computation

One of the main differences between computer algebra systems and traditional calculators is the ability to compute with symbols. In Maple, one can do algebraic manipulations such as expanding and factoring polynomials.

```
>  (x^2*y + y^2 - 1)^3;
```

$$(x^2\,y + y^2 - 1)^3$$

```
>  expand( " );
```

$$x^6\,y^3 + 3\,x^4\,y^4 - 3\,x^4\,y^2 + 3\,x^2\,y^5 - 6\,x^2\,y^3 + 3\,x^2\,y + y^6 - 3\,y^4 + 3\,y^2 - 1$$

```
>  factor( " );
```

$$(x^2\,y + y^2 - 1)^3$$

and normalizing rational functions

```
>  ( x^5 - y^5)/ (x^3 - y^3);
```

$$\frac{x^5 - y^5}{x^3 - y^3}$$

```
> normal( " );
```

$$\frac{x^4 + y\,x^3 + x^2\,y^2 + x\,y^3 + y^4}{x^2 + x\,y + y^2}$$

Similar operations are also possible for symbolic quantities that do not represent polynomial data. For example,

```
> sin( x + 2*y );
```

$$\sin(\,x + 2\,y\,)$$

```
> expand( " );
```

$$2\sin(\,x\,)\cos(\,y\,)^2 - \sin(\,x\,) + 2\cos(\,x\,)\sin(\,y\,)\cos(\,y\,)$$

```
> combine( " , trig );
```

$$\sin(\,x + 2\,y\,)$$

Operations such as factorization and taking the greatest common divisor can be performed in domains such as **Q**[x]. In addition, one can compute in domains such as $\mathbf{Z}_p[x]$

```
> x^24 - 2;
```

$$x^{24} - 2$$

```
> Factor( " ) mod 7;
```

$$(\,x^3 + 4\,)\,(\,x^6 + 2\,x^3 + 2\,)\,(\,x^6 + 2\,)\,(\,x^3 + 3\,)\,(\,x^6 + 5\,x^3 + 2\,)$$

```
> Gcd( "" , x^7 - x^6 + 5*x - 5) mod 7;
```

$$x^6 + 5$$

Maple also has a wide variety of solvers. For example, exact solutions for equations are computed as in

```
> solve( x^3+x^2*a+4*x*a+2*x^2+4*a = 0, x);
```

$$-2, -\frac{1}{2}\,a + \frac{1}{2}\,\sqrt{a^2 - 8\,a}, -\frac{1}{2}\,a - \frac{1}{2}\,\sqrt{a^2 - 8\,a}$$

while floating-point solutions are computed via the `fsolve` procedure.

```
> fsolve( x^5 + 2*x + 2, x, complex );
```

$$-.8174710190, -.6111632249 - .9892348119\,I, -.6111632249 + .9892348119\,I,$$
$$1.019898734 - .8770734498\,I, 1.019898734 + .8770734498\,I$$

Other solvers include `rsolve` for solving recurrence equations

```
> rsolve({f(n) = -3*f(n-1) - 2*f(n-2), f(1) = 1, f(2)=1}, {f});
```

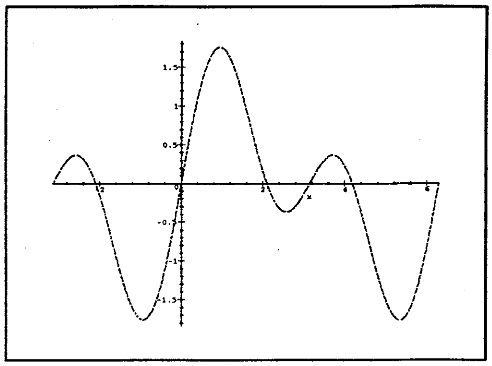
$$\{\mathrm{f}(\,n\,) = -3\,(\,-1\,)^n + (\,-2\,)^n\}$$

`dsolve` for symbolic or numeric solution of differential equations, `pdesolve` for partial differential equations, `linsolve` for linear equations, `isolve` for equations over the integers, and `msolve` for equations over a finite field to name just a few.

2.3 Graphics

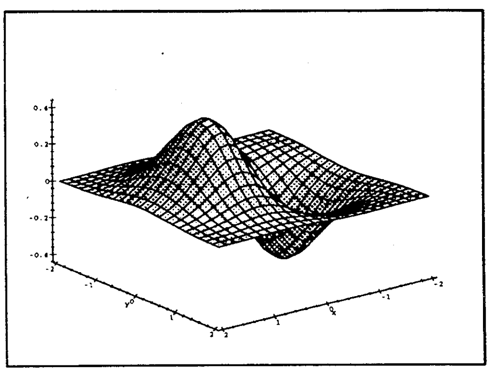
Maple supports a wide variety of graphic functions for visualizing mathematical expressions. One can plot in two-dimensions

```
> plot( sin(x) + sin(2*x), x=-Pi..2*Pi );
```



or three-dimensions

```
> plot3d( x*exp(-x^2 - y^2), x=-2..2, y=-2..2 );
```

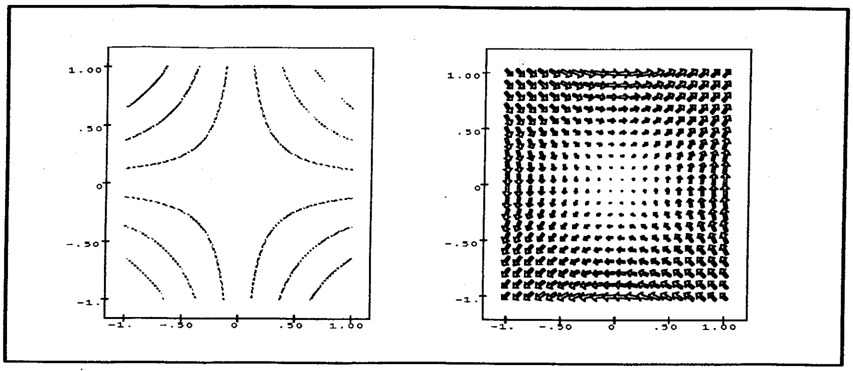


Alternate coordinate systems, drawing styles, and other optional arguments such as axis style, coloring, etc., are available for these commands. There are also a number of specialized plotting routines available in the `plots` and `plottools` packages. The `plots` package includes commands to generate contours, points, gradient, and other vector fields, implicit plots in both two and three dimensions, along with many other types of plots. For example, one can generate the contour plot and the gradient plot of the surface of $z = sin(xy)$

```
> with(plots):
> p:=contourplot(sin(x*y), x= -1..1, y=-1..1, axes=boxed,contours=7):
> q:=gradplot(sin(x*y), x= -1..1,y=-1..1,axes=boxed,arrows=THICK):
```

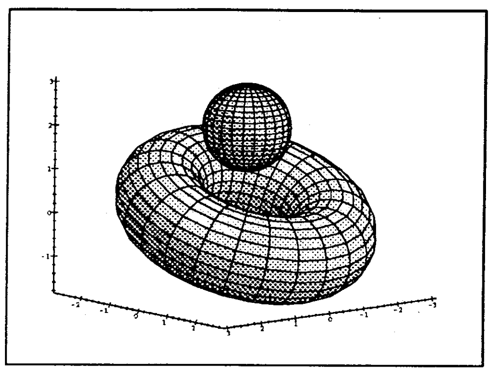
and display the two plots side by side in an array of plots via

```
> display( array( [p, q] ) );
```



The plottools package allows one to manipulate with basic graphical objects, for example

```
> with(plottools):
> display( sphere([ 0, 0, 2]) ,  rotate( torus([0,0,0]), Pi/8,0,0) );
```



Additional objects include geometric shapes such as the tetrahedron, octahedron, and ellipses, to name just a few. One can rotate, scale, and translate such objects or transform them using any function.

2.4 Programming

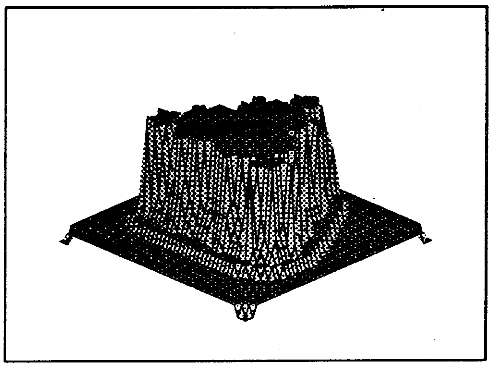
Maple comes with its own programming language. This language is Pascal-like with all the usual constructions such as if-then-else and for and while loops. For example, the procedure

```
> mandelbrotSet := proc(x, y)
>     local z, m;
>         z := evalf( x + y*I );  m := 0;
>         to 10 while abs(z) < 2 do
>             z := z^4 + ( x + y*I );
```

```
>                m := m + 1
>           od;
>           m
>   end:
```

counts the number of iterations (to a maximum of ten) needed to exit the disk of radius two from various starting points. In this case one can visualize such a function using plot3d:

```
> plot3d( mandelbrotSet , -1.5..1.5, -1.5..1.5, grid = [50, 50] );
```



In addition to the basic facilities found in most computer languages, Maple supports a large set of programming tools and constructions that work with mathematical expressions.

2.5 Language Translation

One very useful feature of Maple is its ability to translate mathematical expressions into alternate computer languages. For example, the mathematical expression

```
> s := ln(x)^2 + 2*ln(x)^3;
```

$$s := \ln(x)^2 + 2\ln(x)^3$$

can be differentiated

```
> diff(s,x);
```

$$2\frac{\ln(x)}{x} + 6\frac{\ln(x)^2}{x}$$

and then converted to a Fortran expression

```
> fortran(");

      t0 = 2*alog(x)/x+6*alog(x)**2/x
```

One can ask that the Fortran code be optimized in the sense of pulling out common subexpressions to achieve more efficient evaluation.

```
> fortran(",optimized);

      t1 = alog(x)
      t2 = 1/x
      t4 = t1**2
      t6 = 2*t1*t2+6*t4*t2
```

There are similar commands for translating Maple expressions into the C programming language. In addition, one can convert Maple expressions into text processing languages such as `latex` or eqn:

```
> linalg[hilbert](4);
```

$$\left[\begin{array}{cccc} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{array}\right]$$

```
> latex( " );

\left [\begin {array}{cccc} 1&1/2&1/3&1/4
\\\noalign{\medskip}1/2&1/3&1/4&1/5\\\noalign{\medskip}1
/3&1/4&1/5&1/6\\\noalign{\medskip}1/4&1/5&1/6&1/7
\end {array}\right ]
```

This allows users to incorporate the results of Maple computations into their documents in an error-free manner.

3 Calculus

Since Maple has the ability to do algebraic manipulations, it does not come as a surprise that one can preform many computations from calculus.

```
> Limit( sin(Pi*x) / (x-1), x=1 ) = limit( sin(Pi*x) / (x-1), x=1 );
```

$$\lim_{x \to 1} \frac{\sin(\pi x)}{x-1} = -\pi$$

```
> f := x^(x^x);
```

$$f := x^{(x^x)}$$

```
> diff(f, x);
```

$$x^{(x^x)}\left(x^x(\ln(x)+1)\ln(x)+\frac{x^x}{x}\right)$$

```
> series(cos(x)*exp( a*x^2), x);
```

$$1+\left(a-\frac{1}{2}\right)x^2+\left(\frac{1}{2}a^2-\frac{1}{2}a+\frac{1}{24}\right)x^4+O(x^6)$$

Even indefinite integration

```
> g := sin(a*x)*exp(-b*x);
```

$$g := \sin(x\,a)\,e^{(-b\,x)}$$

```
> int(g, x);
```

$$-\frac{a\,e^{(-b\,x)}\cos(x\,a)}{b^2+a^2}-\frac{b\,e^{(-b\,x)}\sin(x\,a)}{b^2+a^2}$$

is possible with Maple. One can verify the answer by differentiating

```
> diff(" , x);
```

$$\frac{a^2\,e^{(-b\,x)}\sin(x\,a)}{b^2+a^2}+\frac{b^2\,e^{(-b\,x)}\sin(x\,a)}{b^2+a^2}$$

and then simplifying to get the original function

```
> simplify( " );
```

$$\sin(x\,a)\,e^{(-b\,x)}$$

The topic of indefinite integration is an interesting one because Maple does not rely only on heuristics (i.e. intelligent guess work) to solve integrals. Rather, it uses the Risch integration algorithm [2], a decision procedure for determining if closed form solutions exist. Thus, for the integral

```
> int( exp(-x^3)*x^4*log(x), x);
```

$$\int e^{(-x^3)}\,x^4\ln(x)\,dx$$

it is not simply the case that Maple cannot find an answer. Rather, Maple has used the Risch algorithm to prove that no closed form solution exists, at least no closed form solution in terms of elementary functions such as exponentials, logarithms, and algebraic functions.

One can perform manipulations with integrals even when no closed form solution can be determined. For example, one can differentiate or take series expansions

```
> series( " , x, 14);
```

$$\left(\frac{1}{5}\ln(x)-\frac{1}{25}\right)x^5+\left(-\frac{1}{8}\ln(x)+\frac{1}{64}\right)x^8+\left(\frac{1}{22}\ln(x)-\frac{1}{242}\right)x^{11}+O(x^{14})$$

In addition, not finding a closed form expression for an indefinite integral does not necessarily imply that there is no closed form expression for a corresponding definite integral. Thus we have

```
> int( exp(-x^3)*x^4*log(x), x=0..infinity );
```

$$\frac{1}{9}\Gamma\left(\frac{2}{3}\right)-\frac{2}{27}\Gamma\left(\frac{2}{3}\right)\gamma+\frac{1}{81}\Gamma\left(\frac{2}{3}\right)\pi\sqrt{3}-\frac{1}{9}\Gamma\left(\frac{2}{3}\right)\ln(3)$$

One can also calculate formulas for special integrals, for example elliptic integrals. Such integrals, which appear in numerous applications (e.g. computing the arc length of a trigonometric function) reduce to a normal form in terms of Legendre elliptic functions. We can show one of these reduction formulas by taking advantage of Maple's `assume` facility. For example,

```
> assume( k>0, k<1 );
> Int(x^2/sqrt( (1-x^2)*(1-k^2*x^2)), x= 0..1/k);
```

$$\int_0^{\frac{1}{k^{\sim}}} \frac{x^2}{\sqrt{(1-x^2)(1-k^{\sim 2}x^2)}}\,dx$$

```
> value( " );
```

$$\frac{\mathrm{EllipticK}(k^{\sim})}{k^{\sim 2}}-\frac{\mathrm{EllipticE}(k^{\sim})}{k^{\sim 2}}-I\,\mathrm{EllipticPi}\left(1-k^{\sim 2},\sqrt{1-k^{\sim 2}}\right)$$

gives a formula in the case where the integrand crosses one of the branch cuts of the square root function.

4 Conclusions

We have presented a brief introduction to the functionality which is included in the Maple computer algebra system. We have not described such topics as special functions, their manipulation and evaluation to arbitrary precision over the complex plane, advanced graphics, linear algebra, algebraic functions, solvers such as those for exact solutions of partial differential equations, and many other topics.

References

[1] B.W. Char, K.O. Geddes, W.M. Gentleman, and G.H. Gonnet. The design of Maple: A compact, portable, and powerful computer algebra system. Appears in *Computer Algebra (Proceedings of EUROCAL '83)*, J.A. van Hulzen (ed.), Lecture Notes in Computer Science, No. 162, Springer-Verlag, Berlin, 1983, pp. 101-115.

[2] K.O. Geddes, S.R. Czapor and G. Labahn. Algorithms for Computer Algebra. Kluwer Academic Publishers, Boston, 1992, 585 pages.

SYMBOLIC AND NUMERIC INTEGRATION IN MAPLE

Keith Geddes and George Labahn
Department of Computer Science
University of Waterloo, Waterloo, Ontario, Canada
e–mail: {kogeddes, glabahn}@daisy.uwaterloo.ca

1 Introduction

The Maple system is designed to support both symbolic and numeric modes of mathematical computation. In this paper, we give a survey of the algorithms used in Maple to deal with the integration problem of calculus, in both symbolic and numeric contexts.

2 Indefinite Integration

Given a function $f(x)$, the indefinite integration problem asks whether there exists a function $g(x)$ such that $g'(x) = f(x)$. We assume that $f(x)$ and $g(x)$ are both represented as Maple expressions; specifically, as rational expressions in x and various mathematical functions. We will write

$$\int \mathrm{f}(x)\, dx = \mathrm{g}(x)$$

where we do not explicitly write the arbitrary constant of integration.

The algorithmic approach to the indefinite integration problem is based on the Risch integration algorithm. In this presentation, it is only possible to give a very brief outline of a part of the Risch algorithm. For a more detailed exposition of the algorithm, see [1],[4].

2.1 Liouville's Principle

Recall that a field $K(x)$ of rational functions in the variable x over a constant field K consists of all functions which can be represented as a quotient of polynomials, where the numerator and denominator are each polynomials in x with coefficients from the constant field K. For our purposes, the constant field K will be a subfield of the field of complex numbers. For example, K could be the field Q of rational numbers or, more generally, an extension field

$$K = \mathrm{Q}\,(\alpha_1, ..., \alpha_k)$$

where each α_i is an algebraic extension (i.e. the solution of a polynomial equation over a lower field in the hierarchy). Examples of constant algebraic extensions are the complex number i satisfying $i^2 + 1 = 0$, $\sqrt{2}$, or any n-th root of a constant.

To define a new class of functions, we extend the field $K(x)$ to include additional functions. The notation

$$F_n = \mathrm{K}\,(x, \theta_1, ..., \theta_n)$$

denotes the set of all rational expressions in the symbols $x, \theta_1, ..., \theta_n$ over the constant field K. Each extension θ_i will be a new function of x. The extension θ_i is called an *elementary extension* of the field $F_{i-1} = K(x, \theta_1, ..., \theta_{i-1})$ if it can be expressed in one of the following forms: (i) $\theta_i = log(u)$ for some $u \in F_{i-1}$; (ii) $\theta_i = exp(u)$ for some $u \in F_{i-1}$; or (iii) θ_i is *algebraic* over F_{i-1} (i.e. there exists a polynomial $p \in F_{i-1}[z]$ such that $p(\theta_i) = 0$).

The more restrictive class of *transcendental elementary extensions* are those elementary extensions which are not algebraic. A function field F_n is called a *field of elementary functions* if it can be expressed in the above form where each θ_i is an elementary extension, and it is called a *field of transcendental elementary functions* if each θ_i is a transcendental elementary extension.

Note that the field of elementary functions includes all of the mathematical functions commonly encountered in a beginning calculus course, including the exponential, logarithm, trigonometric, inverse trigonometric, hyperbolic, and inverse hyperbolic functions, as well as the square root, cube root, and all other algebraic functions. This relies on the fact that functions such as `sin(x)`, `arctan(x)`, `cosh(x)`, `arcsech(x)` (and the others) can be expressed in terms of exponentials and logarithms if we allow complex numbers in the constant field. The advantage of expressing all of the elementary functions in terms of exponentials, logarithms, and algebraic functions is that this imposes some structure on the possible form of the integral.

For the case of rational function integration, if $f(x) \in K(x)$ then there exists an elementary function $g(x)$ which is an indefinite integral of $f(x)$. We cannot say that $g(x)$ is necessarily a rational function. However, we can state that $g(x)$ lies in an extension field $K(x, \theta_1, ..., \theta_n)$ where each θ_i is a logarithmic extension (with possibly some algebraic number extensions of the constant field K). In other words, the only new functions required are logarithms. Surprisingly, a similar result holds for the field of elementary functions.

Liouville's Principle: Let F be a field of elementary functions with constant field K. For $f \in F$ suppose that the integral

$$\mathrm{g}(\,x\,) = \int \mathrm{f}(x)\,dx$$

is an elementary function. Specifically, suppose that $g \in G$ where G is an elementary extension of the field F having the same constant field K. Then there exist functions $v_0, v_1, ..., v_m \in F$ and constants $c_1, ..., c_m \in K$ such that

$$g(x) = v_0 + \left(\sum_{i=1}^{m} c_i \ln(v_i)\right)$$

2.2 The Risch algorithm for a transcendental log extension

In the space available here, the general form of the Risch algorithm for the case of transcendental elementary functions will be outlined by concentrating on the case of a logarithmic extension. The case of an exponential extension is quite similar. Computation with algebraic functions involves a rather different set of mathematical concepts and will not be pursued in this presentation.

Let $f \in K(x, \theta_1, ..., \theta_n)$ and suppose that the last extension $\theta = \theta_n$ is a transcendental log extension, say $\theta = log(u)$ for $u \in F_{n-1} = K(x, \theta_1, ..., \theta_{n-1})$. We view the integrand as a rational function in the variable θ; specifically, $f(\theta) = p(\theta)/q(\theta) \in F_{n-1}(\theta)$. Without loss of generality, we may assume that $f(\theta)$ is normalized such that $p(\theta), q(\theta) \in F_{n-1}[\theta]$ satisfy $GCD(p(\theta), q(\theta)) = 1$ and that $q(\theta)$ is monic.

The steps in the Risch algorithm are as follows.

1. Apply Euclidean division to the polynomials $p(\theta), q(\theta) \in F_{n-1}[\theta]$ yielding polynomials $s(\theta), r(\theta) \in F_{n-1}[\theta]$ such that $p(\theta) = q(\theta)s(\theta) + r(\theta)$ with $r(\theta) = 0$ or $deg(r(\theta)) < deg(q(\theta))$. Then

$$\int f(\theta)\, dx = \int s(\theta)\, dx + \int \frac{r(\theta)}{q(\theta)}\, dx$$

   The first term on the right hand side is called the integral of the *polynomial part* and the second term is called the integral of the *rational part*.

Integration of the Rational Part

2. Compute the square-free factorization of the denominator $q(\theta) \in F_{n-1}[\theta]$:

$$q(\theta) = \prod_{i=1}^{k} q_i(\theta)^i$$

   where each $q_i(\theta)$ $(1 <= i <= k)$ is monic and square-free.

3. Compute the partial fraction expansion

$$\frac{r(\theta)}{q(\theta)} = \sum_{i=1}^{k} \sum_{j=1}^{i} \frac{r_{i,j}(\theta)}{q_i(\theta)^j}$$

   where each $r_{ij}(\theta) \in F_{n-1}[\theta]$ and $deg(r_{ij}(\theta)) < deg(q_i(\theta))$. Then

$$\int \frac{r(\theta)}{q(\theta)}\, dx = \sum_{i=1}^{k} \sum_{j=1}^{i} \int \frac{r_{i,j}(\theta)}{q_i(\theta)^j}\, dx$$

4. Apply Hermite reduction to reduce any power $j > 1$ in the denominator. For a particular term $r_{ij}(\theta)/q_i(\theta)^j$ with $j > 1$, by applying the Extended Euclidean Algorithm it is possible to compute polynomials $s(\theta), t(\theta) \in F_{n-1}[\theta]$ such that

$$\mathrm{s}(\theta)\, q_i(\theta) + \mathrm{t}(\theta)\,\frac{\partial q_i(\theta)}{\partial x} = r_{i,j}(\theta)$$

In terms of $s(\theta), t(\theta)$, the Hermite reduction formula takes the following form.

$$\int \frac{r_{i,j}(\theta)}{q_i(\theta)^j}\,dx = -\frac{\mathrm{t}(\theta)}{(j-1)\,q_i(\theta)^{(j-1)}} + \int \frac{\mathrm{s}(\theta) + \dfrac{1}{j-1}\dfrac{\partial t(\theta)}{\partial x}}{q_i(\theta)^{(j-1)}}\,dx$$

By repeated application of this reduction process, the integral of the rational part takes the following form:

$$\int \frac{\mathrm{r}(\theta)}{\mathrm{q}(\theta)}\,dx = \frac{\mathrm{c}(\theta)}{\mathrm{d}(\theta)} + \int \frac{\mathrm{a}(\theta)}{\mathrm{b}(\theta)}\,dx$$

where $a(\theta), b(\theta), c(\theta), d(\theta) \in F_{n-1}[\theta]$, $deg(a(\theta)) < deg(b(\theta))$, and $b(\theta)$ is monic and square-free.

5. *Rothstein/Trager method.* For the integration of the rational part, it remains to specify how to compute the integral of a proper rational function $a(\theta)/b(\theta)$. By a method attributed independently to Rothstein [9] and Trager [12], we compute the following *polynomial resultant* $R(z) \in F_{n-1}[z]$:

$$\mathrm{R}(z) = res_\theta(\mathrm{a}(\theta) - z\frac{\partial \mathrm{b}(\theta)}{\partial x}, \mathrm{b}(\theta))$$

Then the integral of $a(\theta)/b(\theta)$ is an elementary function if and only if all the roots of the polynomial $R(z)$ are constants (i.e. free of the integration variable x). If all roots of $R(z)$ are constants, let c_i $(1 <= i <= m)$ be the distinct roots of $R(z)$ and define $v_i(\theta)$ $(1 <= i <= m)$ by

$$v_i(\theta) = \mathrm{GCD}(\mathrm{a}(\theta) - c_i\frac{\partial \mathrm{b}(\theta)}{\partial x}, \mathrm{b}(\theta))$$

Then

$$\int \frac{\mathrm{a}(\theta)}{\mathrm{b}(\theta)}\,dx = \sum_{i=1}^{m} c_i \ln(v_i(\theta))$$

Example 2.1

If we apply the Risch algorithm to the integral

$$\int \frac{1}{x\ln(x)}\,dx$$

the resultant computation for this problem is

```
> a := 1:  b := x*theta:
> diff_b := diff(x*ln(x), x):
> diff_b := subs(ln(x) = theta, diff_b);
```

$$diff\_b := \theta + 1$$

```
> R := resultant(a - z*diff_b, b, theta);
```

$$R := x\,(-1+z)$$

The single root of $R(z)$ is $z = 1$ and the integral is elementary. Specifically, $c_1 = 1$, $v_1 = \theta$ and the integral is $c_1\, ln(v_1) = ln(\theta)$. In other words,

$$\int \frac{1}{x \ln(x)}\, dx = \ln(\ln(x))$$

Integration of the Polynomial Part

6. The problem is to compute (if it exists as an elementary function) the integral $q(\theta)$ defined by

$$\int \mathrm{p}(\theta)\, dx = \mathrm{q}(\theta)$$

where the given integrand is $p(\theta) \in F_{n-1}[\theta]$ and $\theta = ln(u)$ for some $u \in F_{n-1}$. A careful argument based on Liouville's Principle determines that if

$$\mathrm{p}(\theta) = p_l\, \theta^l + p_{l-1}\, \theta^{(l-1)} + \ldots + p_0$$

then $q(\theta)$, if it exists as an elementary function, must be of the form

$$\mathrm{q}(\theta) = q_{l+1}\, \theta^{(l+1)} + q_l\, \theta^l + \ldots + q_1\, \theta + q_0$$

Applying differentiation, the relationship between the given integrand $p(\theta)$ and the unknown result $q(\theta)$ yields the following system of equations to solve.

$$0 = \frac{\partial q_{l+1}}{\partial x}$$

$$p_l = (l+1)\, q_{l+1} \frac{\partial \theta}{\partial x} + \frac{\partial q_l}{\partial x}$$

$$p_{l-1} = l\, q_l \frac{\partial \theta}{\partial x} + \frac{\partial q_{l-1}}{\partial x}$$

$$\ldots$$

$$p_1 = 2\, q_2 \frac{\partial \theta}{\partial x} + \frac{\partial q_1}{\partial x}$$

$$p_0 = q_1 \frac{\partial \theta}{\partial x} + \frac{\partial q_0}{\partial x}$$

These equations can be solved systematically in the order stated, by recursively applying the integration algorithm in the field F_{n-1} which is free of the extension θ. Note that the given coefficients $p_i, 0 <= i <= l$ all lie in the field F_{n-1} and it is important when solving these equations that the solutions $q_i, 1 <= i <= l+1$ must all lie in the field F_{n-1} (possibly with some extensions to the constant field). However, the final coefficient q_0 is allowed to contain new log extensions.

The integration of the "polynomial part" involves recursive application of the Risch algorithm. If any of the sub-integrals which arise is not elementary then the original integral is not elementary. Furthermore, if any of the sub-integrals is elementary but contains a new log extension (other than the particular extension θ) then the original integral is not elementary.

2.3 Extensions of the Risch algorithm

When the Risch algorithm determines that a particular integral "cannot be expressed", this conclusion is relative to a particular class of "allowable functions". The Risch algorithm is usually presented in the context of the class of *elementary functions* described earlier. Ideally, one would like to broaden the class of functions to include the various Special Functions such as the error function, gamma function, exponential integrals, and many others. In fact, the theoretical framework of the Risch algorithm has been extended to only a few non-elementary extensions (see [2], [10].

A computer algebra system such as Maple has knowledge about various Special Functions and it expresses integrals in terms of these functions. However, failure to express a particular indefinite integral in closed form does not imply that a decision procedure has determined that there is no closed-form expression for the integral. For the case of definite integrals, various techniques are applied to attempt to express the result as a symbolic formula (see [6] and the references therein for details).

3 Definite Integration: Hybrid Symbolic-Numeric Techniques

We turn now to a different topic: *numerical integration.* For many definite integration problems, a computer algebra system may be unable to express the result in symbolic closed form. Even if there is a closed-form formula, perhaps all that is actually required is a numerical value for the definite integral (to some precision).

Our purpose here is to examine the potential advantages to be gained by exploiting a hybrid *symbolic-numeric* approach to the problem of obtaining a numerical value for a definite integral. For some additional details on this topic see [5].

Traditional numerical integration codes typically require the interval of integration to be finite and the integrand to be finite at every point of the interval, although some numerical routines have been written to handle specific classes of singularities. The Maple computer algebra system is a natural environment in which to treat singularities. By using generalized series expansions, variable transformations, and other techniques, it is possible to automatically handle the numerical integration problem in the presence of various types of singularities. Moreover, integrals can be computed to arbitrarily high precision.

Example 3.1

Maple provides a facility to "evaluate in floating-point mode" those definite integrals which are left unevaluated by the symbolic techniques of the `int` command, or which are deliberately left unevaluated by using the inert `Int` command.

```
> f := sin(x)*ln(x)*exp(-x^3):
> int(f, x=0..infinity);
```

$$\int_0^\infty \sin(x)\ln(x)\,e^{(-x^3)}\,dx$$

```
> evalf(");
```

$$-.1957885158$$

```
> evalf("", 25);
```

$$-.1957885158487997538390572$$

The computation of the above numerical results employs several of the techniques discussed in the following sections.

3.1 The case of an analytic integrand on a finite interval

Along with the techniques for symbolic analysis discussed here, we need an efficient routine for numerical integration in the simple case where the interval of integration is finite and the integrand has no singularities on the interval. Most traditional numerical integration methods could serve for this purpose. However, since the aim of the symbolic analysis phase is to present the numerical integration phase with an analytic integrand on a finite interval, we have found that the preferred default method is Clenshaw-Curtis quadrature [8], [3]. This method, named **ccquad**, is based on computing the Chebyshev series expansion of the integrand on the specified interval and then integrating term-by-term. Clenshaw-Curtis quadrature is at its best when the integrand is analytic in a sufficiently large region of the complex plane surrounding the interval of integration, in which case the Chebyshev series expansion converges quickly.

In Maple, ccquad is programmed to detect when convergence is too slow (i.e. there is a nearby singularity, perhaps on the real line, perhaps somewhere else in the complex plane) and to return control to the main routine when this occurs. The singularity-handling techniques are then invoked. If attempts to remove the singularity are unsuccessful, the code invokes another numerical integration routine named quadexp, an adaptive double-exponential quadrature method [11]. The strength of this method is that it is less sensitive than ccquad to the presence of singularities near to the interval of integration.

3.2 Generalized series expansions

The main tool for analyzing integrand singularities is a facility for computing generalized series expansions of expressions, called hierarchical series [7]. The standard Maple syntax for generating series expansions is

```
series(expr, var = value, ord);
```

although some of the most general series expansions which are needed are not returned by the `series` command in current versions of Maple. (There are special series expansion routines associated with the numerical integration routines.)

A generalized series expansion may not be a pure power series in the variable x but it may involve terms in $ln(x)$, $exp(-1/x)$, $exp(-1/x^2)$, etc. These series expansions will be understood to be one-sided expansions for $x > 0$. As long as the non-polynomial functions introduced into the series expansion can be bounded by powers of x, the mathematical validity of the series can be verified.

3.3 Transforming to an analytic integrand

If an integrand is found to be non-analytic at a point of the interval of integration, the first technique tried is to look for a transformation which will remove the singularity. The types of transformations which are attempted include: (i) subtracting off the singularity, and (ii) an algebraic transformation of variables. In each case, the fundamental tool for symbolic analysis is the generalized series expansion discussed in the preceding section.

Let f be the integrand, x be the variable of integration, $x = 0$ be the singular point, and s be the generalized series expansion of f about $x = 0$.

For case (i), the method used is to test each term in the expansion s to determine which terms are regular (i.e. have a Taylor series expansion at $x = 0$). If the number of non-regular terms is less than half the number of terms in s, then make the conjecture that the expression $f - q$ might be analytic at $x = 0$, where q denotes the sum of the non-regular terms. Test this conjecture, and if it is true then the numerical integration of $f - q$ can proceed normally. In some cases, the non-regular part q will be integrable by the symbolic integrator; otherwise, it will be passed on to the general techniques discussed in subsequent sections.

Example 3.2

```
> h := ln(1 - cos(2*x)):
```

The generalized series expansion of h at $x = 0$ takes the following form.

```
> series(h, x=0, 8);
```

$$\left(\ln(2) - 2\ln\left(\frac{1}{x}\right)\right) - \frac{1}{3}x^2 - \frac{1}{90}x^4 - \frac{2}{2835}x^6 + O(x^8)$$

The non-regular part is $q = -2\,ln(1/x)$. The new expression $h - q$ is analytic on the interval [0, 1] and therefore it can be integrated easily by ccquad.

```
> r1 := evalf(Int(h + 2*ln(1/x), x=0..1));
```

$$r1 := .5797067686$$

(computing to 10 digits of accuracy). Integrating q can be done symbolically.

```
> r2 := int(-2*ln(1/x), x=0..1);
```

$$r2 := -2$$

Finally, summing the two values yields the desired result.

```
> Int(h, x=0..1) = r1 + r2;
```

$$\int_0^1 \ln(1 - \cos(2x))\,dx = -1.420293231$$

Case (ii) is the case of an algebraic transformation of variables. This method comes into effect whenever there are fractional powers of x appearing in the series expansion s, whether or not there are other non-regular functions appearing in the expansion. The idea is to transform away algebraic singularities, and if other singularities remain they will be handled on a subsequent pass. The specific method used is to compute the least common multiple n of all denominators of the fractional powers, and then to apply the change of variables $t = x^{1/n}$.

3.4 Direct integration of a generalized series

Suppose that we have an integrand with a singularity at the left end-point $x = 0$. Suppose further that we have not succeeded in finding a transformation which removes the singularity. Then the method used is to treat the generalized series expansion as an approximation of the integrand near the singularity, and to directly integrate this generalized series over its interval of accuracy. The remainder of the interval can then be handled by the numerical integration method.

The success of this technique relies on the power of Maple's symbolic integrator to handle many of the singular functions which may arise. For series involving $ln(x)$, the indefinite integral of $ln(x)\,x^k$ is an elementary function. For series involving $exp(-1/x)$, the indefinite integral of $exp(-1/x)\,x^k$ can be expressed in terms of the exponential integral $Ei(t)$. Similarly, series involving the singular function $exp(-1/x^2)$ may lead to both the exponential integral and the error function $erf(t)$.

References

[1] M. Bronstein, Integration of elementary functions. J. Symbolic Comp, 9 (2), 1990, pp. 117-173.

[2] G.W. Cherry. Algorithms for Integrating Elementary Functions in Terms of Logarithmic Integrals and Error Functions. Ph.D Thesis, Univ. of Delaware, August 1983.

[3] K.O. Geddes. Remark on Algorithm 424: Clenshaw-Curtis quadrature. ACM Trans. Math. Software, 5 (2), 1979, p. 240.

[4] K.O. Geddes, S.R. Czapor, and G. Labahn. Algorithms for Computer Algebra. Kluwer Academic Publishers, Boston, 1992.

[5] K.O. Geddes and G.J. Fee. Hybrid symbolic-numeric integration in Maple. Proceedings of ISSAC '92, ACM Press, New York, 1992, pp. 36-41.

[6] K.O. Geddes, M.L. Glasser, R.A. Moore, and T.C. Scott. Evaluation of classes of definite integrals involving elementary functions via differentiation of special functions. Applicable Algebra in Engineering, Communication and Computing, 1 (2), November 1990, pp. 149-165.

[7] K.O. Geddes and G.H. Gonnet. A new algorithm for computing symbolic limits using hierarchical series. Symbolic and Algebraic Computation, P. Gianni (ed.), Lecture Notes in Computer Science, No. 358, Springer-Verlag, Berlin, 1989, pp. 490-495.

[8] W.M. Gentleman. Algorithm 424: Clenshaw-Curtis quadrature. Comm. ACM, 15 (5), 1972, pp. 353-355.

[9] M. Rothstein. Aspects of Symbolic Integration and Simplification of Exponential and Primitive Functions. Ph.D. Thesis, Univ. of Wisconsin, Madison, 1976.

[10] M. Singer, B. Saunders, and B.F. Caviness. An extension of Liouville's theorem on integration in finite terms. SIAM J. Computing, pp. 966-990, 1985.

[11] Hidetosi Takahasi and Masatake Mori. Double exponential formulas for numerical integration. Publ. RIMS, Kyoto Univ., 9, 1974, pp. 721-741.

[12] B. Trager. Integration of Algebraic Functions. Ph.D. Thesis, Dept. of EECS, M.I.T., 1984

Enhancing the Teaching of Mathematics to Engineers using Information Technology

Dr H W Gretton, Dr N V Challis and Dr P D Rangecroft
Mathematics Centre
Sheffield Hallam University
Sheffield S1 1WB
UK
E-mail H.Gretton@shu.ac.uk N.Challis@shu.ac.uk P.Rangecroft@shu.ac.uk

Introduction

At Sheffield Hallam University various course in Mathematics for Engineers at all levels are taught using a range of technologies. The TI-82 and TI-85 graphic calculators play a major role, along with the very recently introduced CBL (Calculator Based Laboratories); computer algebra packages such as Derive, and the spreadsheet Excel, are also utilised as part of an integrated technological approach to Mathematics (1). The students can use the various links available and incorporate all of these technologies to produce high quality case study reports using the Word package in the Microsoft Office. All students have internet and E-mail facilities enabling them to access the archives - such as the TI archive - for support.

The courses involved are in the Engineering Degree Programme which includes courses such as Computer Aided Engineering, Engineering with Business Studies, Mechanical Engineering, Electrical Engineering, Engineering Physics and so on, and the Construction Degree Programme containing degrees such as Civil Engineering, Construction Engineering, Quantity Surveying and Building Surveying. The number of students involved is over 400 and they come into the institution with a wide variety of mathematical backgrounds and numerical skills.

The current students in many cases have different skills from their predecessors; they possess fewer of the basic manipulative algebraic skills and less knowledge of the rules of mathematics governing these processes. This is a cause of some concern nationally in the UK at the present time (2). However, they do frequently have experience which can help to offset this disadvantage: they have in many cases been exposed to IT products such as computer games in their leisure pursuits, and machines such as the TI-82 and TI-85 can provide similar appeal and enable students to visualise abstract mathematical concepts and find fun in learning about mathematics.

Starting at the right place

Initially the students must spend time on consolidation of the basic algebraic skills because if they are not familiar with the rules, communicating with any machine is difficult. Therefore in the early sessions it is imperative to re-enforce the rules.

Many exercises are available for this purpose: one example involves students **themselves** using the SOLVER facility of the TI-85 to deal with **Errors and rearrangement of equations**:

The students use the SOLVER key in the following problem:

> If a metal rod is heated, the length increases according to the equation:
>
> $$L = L_0(1 + \alpha T)$$
>
> where L is the final length, L_0 is the initial length, T is the increase in temperature and α is the linear coefficient of expansion of the rod. If a steel rod of length 1 metre expands to a length of 1.004m when heated through 230°C, find α.
>
> If the measurements of the rod are only accurate to the nearest 0.5 mm, what are the limits for the value of α?

The equation can be set up using SOLVER, giving:

```
L=L0(1+α*T)
 L=1.004
 L0=1
▪α=1.7391304347826E-5
 T=230
 bound={-1E99,1E99}
▪left-rt=0
GRAPH RANGE ZOOM TRACE SOLVE
```

Hence α is found.

To address the matter of errors, the range of values for L can be entered, and the SOLVER used to explicitly give the tolerance on α:

```
L=L0(1+α*T)
 L=1.0045
 L0=1
▪α=1.9565217391441E-5
 T=230
 bound={-1E99,1E99}
▪left-rt=0
GRAPH RANGE ZOOM TRACE SOLVE
```

Thus a considerable amount of effort, as well as the need for calculus for error analysis, is avoided.

A more advanced topic

All engineering students must meet **differential equations**, but they often find solutions lengthy and difficult, and this can inhibit their understanding of the concepts therein. Here the TI-85 is useful both in solving differential equations directly and as a support tool for the classical methods.

Here is an illustration of the contribution the technology can make here:

Consider the solution of second order differential equations. The most useful one to study is the damped spring model, extended here to include non-linear damping and restoring force terms, that is, to models of the form

$$\frac{d^2y}{dt^2}+2b\frac{dy}{dt}+k^2y=f(t)$$

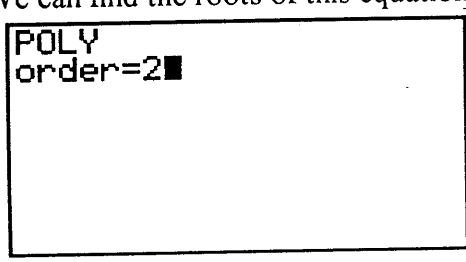
Here $b>0$ is the resistance coefficient, f(t) is the forcing term and k^2 is the internal resistance of the system. Many physical problems modelled by this differential equation contain non-linear elements that are approximated by terms linear in y or in dy/dt. The TI-85 will deal with non-linear extensions of the model.

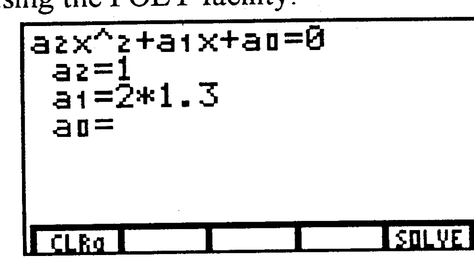
Let f(t) = 0 and b = 1.3 and k = 3.5.
Suppose the boundary conditions are: when t = 1 then y = 0.4 and dy/dt =3.6.

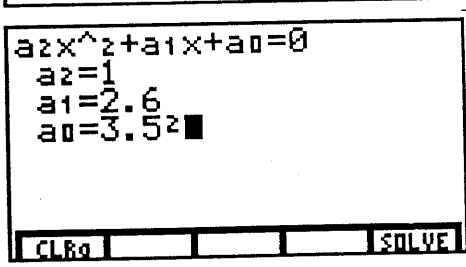
First study the contribution the technology can make in taking the drudgery out of the classical solution method:

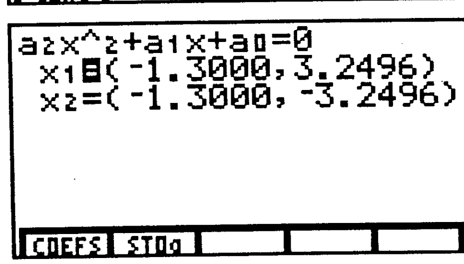
The auxiliary equation is $n^2+2\times1.3+3.5^2=$

We can find the roots of this equation using the POLY facility:









So in this case we have complex roots and the general solution (sometimes called the complete primitive) is:

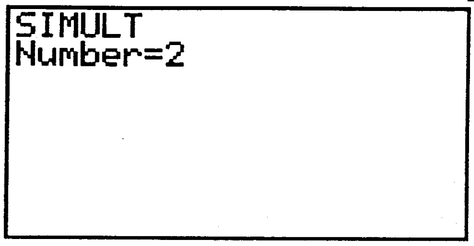
$$y=e^{-1.3t}(A\cos 3.249t+B\sin 3.249t)$$

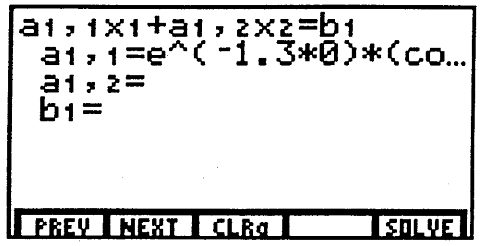
To find the arbitrary constants A and B, we use the boundary conditions and this leads to solving two simultaneous equations. Here the SIMUL facility is very powerful as it enables even derivatives to be put in as constants. So applying the boundary conditions to the solution yields (' has its usual meaning of differentiation)

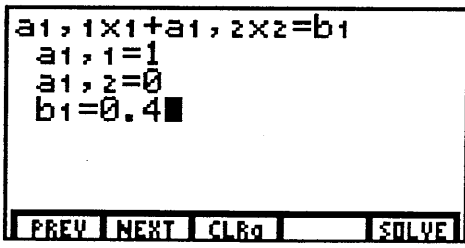
$$e^{-0}\cos 0 \quad A + \quad e^{-0}\sin 0 \quad B = 0.4$$

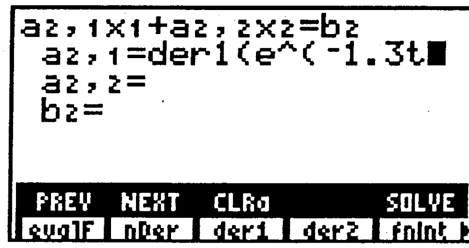
$$(e^{-1.3t}\cos(3.2496t))'A + (e^{-1.3t}\sin(3.249t))'B = 3.6$$

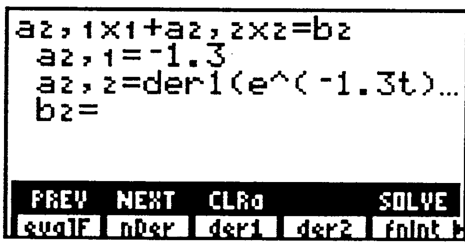
We have two simultaneous equations for A and B, and we simply put the coefficients in as they appear - taking care that we use BODMAS correctly.

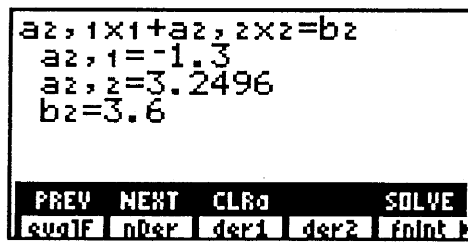


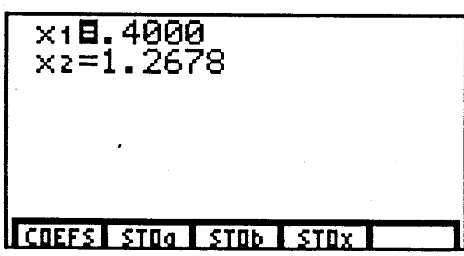








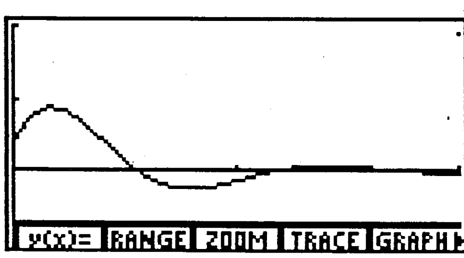




Hence we have A = 0.4 and B = 1.2678, yielding the particular solution :

$$y = e^{-1.3t}(0.4\cos(3.249t) + 1.2678\sin(3.249t))$$

Plotting this solution is always a good idea, providing among other things a check that the boundary conditions are valid!



A variation on this approach when there are complex roots is to use the complex answers directly and only plot the real part of the solution.

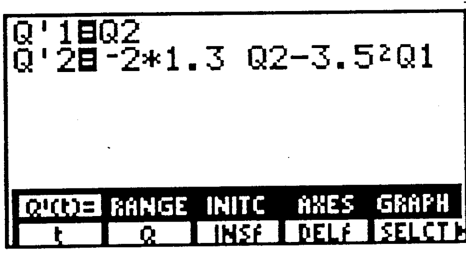
An alternative approach is to use the **differential equation mode in the TI-85 machine** to solve the same question, although this will only provide a graphical and numerical solution. We simply reduce the second order equation to two first order. equations; thus :

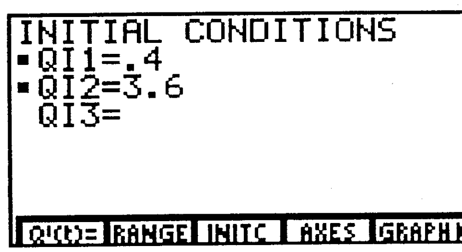
$$\frac{d^2y}{dt^2}+2b\frac{dy}{dt}+k^2y=f(t)$$

becomes

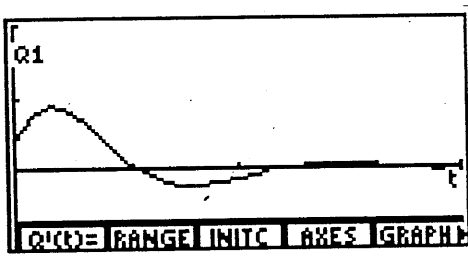
$$Q'1=Q2$$
$$Q'2=-2\times1.3Q2-3.5^2Q$$

As before, let f(t) = 0 and b = 1.3 and k = 3.5 and suppose that the boundary conditions are when t = 1 then y = 0.4 and dy/dt =3.6. Therefore QI1 = 0.4 and QI2 = 3.6. Using the same window and axes gives the following :

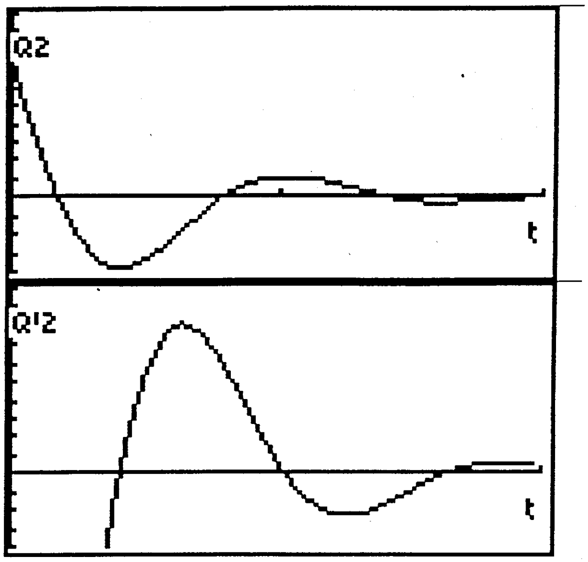




This yields the same answer as above:



Having used this facility, it is then easy to display the velocity (Q2) and acceleration (Q'2) of the system, by changing the variables on the axes:



The ability to solve differential equations more quickly and with less manual work, and the possibility of quick graphical solutions allows the students to concentrate on the new concepts with fewer distractions. The classical approach, with the insights it brings, is not abandoned; however the technology can significantly enhance the students ability to address the concepts without becoming bogged down in heavy algebraic manipulations.

Implications of this approach

Many important issues are raised by the kind of approach being taken here:

- One of the questions to be addressed is that of the impact of the new technology on the processes and content of the traditional engineering mathematics curriculum as presented for example in (4). The curriculum now needs to be reviewed in the light of the new possibilities opened up by recent technological advances. For example we could ask exactly what mathematical skills are required by a modern engineer, or what is the optimum balance of traditional manual and technological skills to enable an engineering student to gain the best value from their education. One might ask for example what manual algebraic skills are required if there is a machine which will solve any equation for you with the SOLVER. There is much to think about here.

- The methods being developed here have implications for the way in which we assess our students' progress. Traditional formal examinations may not be the most appropriate strategy when a technological approach is taken, but coursework of a traditional nature is easy to copy or , if individualised, difficult to mark. Our own favoured approach is to mix the various types of assessment, according to exactly what learning outcome it is that we wish to assess. We have been particularly impressed by some students' achievements in producing formal reports on case study activity. We aim to report on this elsewhere.

Our developments are having an impact far beyond the Mathematics studied by the engineering students, and we have viewed various reactions, particularly with respect to assessment. For example we have heard of one university which has banned graphical calculators from examinations, and at another place the memory banks must be completely cleared. There is obviously much to deal with in this respect.

- Another prime consideration is staff development. Innovations such as the present one are more often than not carried forward by small enthusiastic groups, and it can be difficult to carry other more conservative staff along with the process. Jones(5) reports experience of implementing an innovative approach to mathematics teaching, which sounds all too familiar to our own experience at Sheffield Hallam University, and we have written down some of our own thoughts about implementing change previously (6).

The future

A recent development is the advent of the CBL, which enables real data to be readily collected without the need for expensive laboratory setups. For example the differential equation problem above can be illustrated by real spring-damper systems and the data can be analysed immediately using the TI-85 or TI-82 (3). This portable and comparatively inexpensive equipment gives a large number of other practical modelling possibilities, in areas as diverse as movement, light, temperature, voltage, and even monitoring the stress level of the lecturer by measuring the heartbeat. Going through the entire process from collection of real data, through the modelling of the system to comparing the solutions with the collected data, can give the student understanding and insight into the mathematical processes.

Another consideration is that the technology does not stand still; indeed the next generation of machines have hand-held algebra capability.

To conclude, we note that we are living in the midst of a time of turmoil in mathematics teaching, indeed some have called it a time of paradigm shift. Technology is affecting not only university mathematics teaching, but also what happens in school: some of the new UK 'A' levels, which provide students with one route into university from school, make compulsory use of a graphic calculator. The authors are not ancient, but can recall having to use log tables (8 figures if you were lucky). Things are continuing to develop quickly, and there is a need for educational as well as technical research to make sure that we make the best use of the opportunities presented to us here.

References

(1) H W Gretton: Engineering Education Conference: Increasing student participation, "Teaching mathematics to engineers using graphic calculators", Sheffield Sept 1994.

(2) R Sutherland, S Pozzi: "The changing mathematical background of undergraduate engineers", The Engineering Council, March 1995

(3) C Brueningsen, W Krawiec: "Exploring Physics and Math with the CBL System" Texas Instruments 1994.

(4) "A core curriculum in mathematics for the European engineer", SEFI document 92.1 1992

(5) K S Jones: International Conference on Technology in Collegiate Mathematics Parsippany, "The good, the bad and the ugly: Our experience with requiring graphic calculators in college Algebra", New Jersey Nov 1993.

(6) N V Challis , P Rangecroft: Engineering Education Conference : Increasing student participation, "Who needs the calculus? Quality, standards and change in Engineering Mathematics", Sheffield Sept 1994.

The First Asian Technology Conference in Mathematics
18-21 December 1995

Organised by

Association of Mathematics Educators
in Conjunction with
National Institute of Education &
Radford University, USA

Sponsors

of the

First Asian Technology Conference in Mathematics :

National Institute of Education, Singapore;
the British Council;
the Lee Foundation;
Association of Mathematics Educators,
Singapore; Radford University, USA;
TCI Software Research;
Waterloo Maple Software and
Wolfram Research.

TEACHING OF MATHEMATICAL STATISTICS USING SPREADSHEETS: AN EXPERIMENT IN A DEPARTMENT OF NUTRITION AND FOOD SCIENCE

AKIKO HARADA
HYOGO WOMEN'S COLLEGE
HIRAOKA-CHO
KAKOGAWA CITY
HYOGO 675-01
JAPAN

1. INTRODUCTION

Departments of Nutrition and Food Science in junior colleges are accredited institutions for the training of dietitians. In these departments, computers are used for nutritional calculation and assessment. They are powerful and speedy tools for making calculations. They are also tools to manage information, more powerfully than calculators. In fact, they have become an indispensable part of our society.

Dietitians are expected to serve not only at hospitals and clinics, but also in industries, schools, and in community health organizations. They face challenging opportunities and responsibilities playing a role in the nutritional improvement of the public. The responsibilities of dietitians in the field of health have increased as new scientific discoveries have been made.

Dietitians must collect adequate information in order to fulfill their duties. One of these duties is the management of information with report to our good health. Figure 1 displays one such example of information processing for nutritional improvement and a healthy lifestyle.

Since 1980, objectives of informatics for educating dietitian have been formulated in Japan. The relation between mathematical statistics and informatics in education will be of great importance. The benefits of teaching mathematical statistics have proven greater than what has been previously demonstrated. This paper will present an overview and some of subsequent achievements of how to apply spreadsheets to data management and analysis, derived from an experiment in a junior college program.

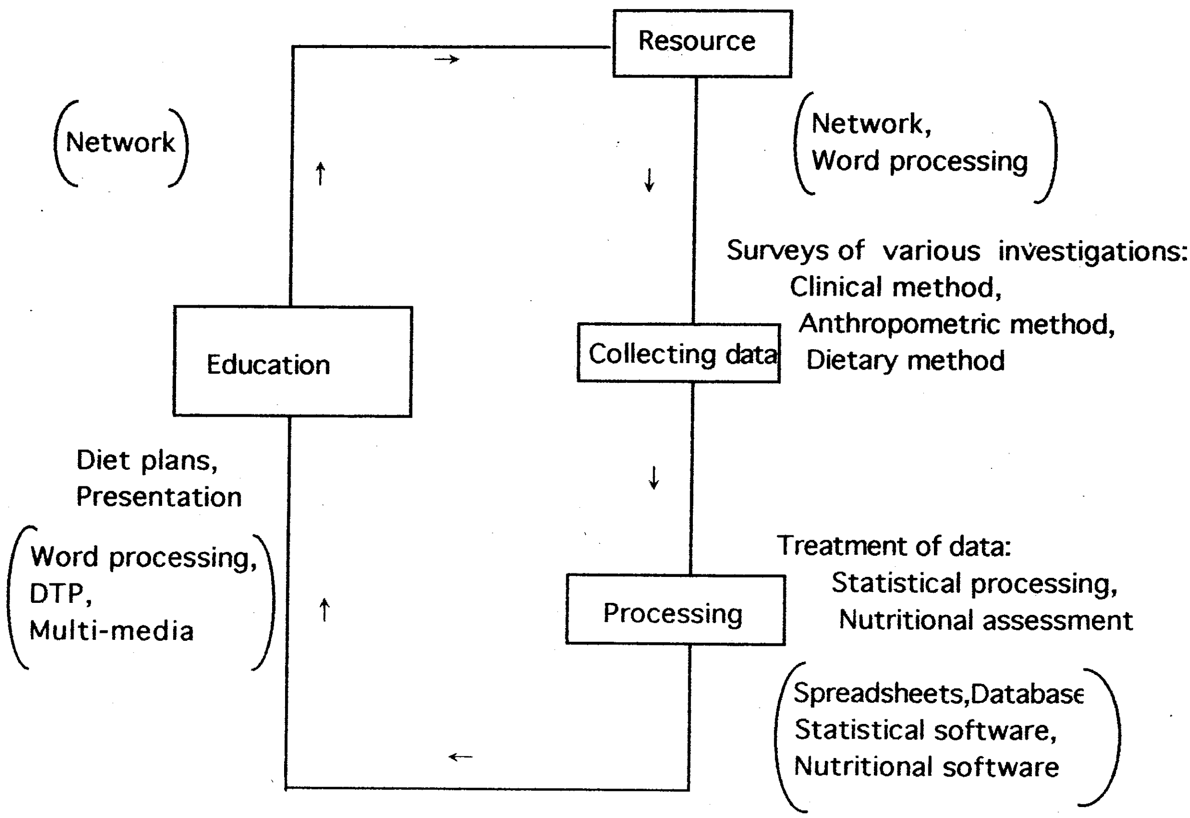


Figure 1. Information Processing for Dietitians

(Brackets mean the use of computers.)

2. TEACHING PROCESS IN THE MATHEMATICAL STATISTICS COURSE

1) DESCRIPTIVE STATISTICS

Continuous variables in question are height and weight of the students, the energy requirement which they determine in the introductory course of nutrition, and the energy intake based on the average of food records of three days .

We can obtain the following output data: body mass index (BMI=weight /height^2) and percentage of total dietary calories in comparison to the requirement by formulas, and the three grades of BMI and the percentage of calories by the nested "@IF function". These are important factors in estimating the desirable weight and food intake.

The frequency distribution is easier to visualize if it is presented graphically. For continuous type variables, a graph called a histogram is commonly used. It may also be used for other types of variables. The grades of BMI and percentage of calories are discrete variables, and these grades (1,2,3: class marks) classify the data into a frequency table. We can calculate percentages of values lying in the class, and draw bar graphs and/or circle graphs. These are obtained by applying functions of "frequency" and "graph", so they are good subjects for learning the skill of spreadsheets, on the basis of continuous variables.

"Numerical @functions" give the sample size, average, standard deviation, variance, and the smallest and largest values of continuous variables (i.e., height, weight, energy requirement and intake, BMI, and the percentage of calories). In the next step, these results and "numerical @functions" give the sample variance, the sample standard deviation, the range, and the three intervals: (x-s,x+s), (x-2s,x+2s), and (x-3s,x+3s). The median is estimated by the "@RANK function" or sorting the set. These values and cells' contents are shown in Table 1 and Table 2.

Table 1. Characteristic Values of Height

Item	Height
Size of the Data	74
Minimum	145.70
Maximum	167.70
Average(m)	157.63
Variance	19.51
Standard Deviation	4.42
Sample Variance	19.78
Sample Standard Deviation(s)	4.45
Range of the Set	22.00
m-s	153.19
m+s	162.08
m-2s	148.74
m+2s	166.53
m-3s	144.29
m+3s	170.97
Median	157.80

Table 2. Cell's Contents for Calculating Characteristic Values

Cell Address	Cell's Contents
A79	Items
B79	Height
A80	Size of the Data
B80	@COUNT(B2..B75)
A81	Minimum
B81	@MIN(B2..B75)
A82	Maximum
B82	@MAX(B2..B75)
A83	Average(m)
B83	@AVG(B2..B75)
A84	Variance
B84	@VAR(B2..B75)
A85	Standard Deviation
B85	@STD(B2..B75)
A86	Sample Variance
B86	+B80*B84/(B80-1)
A87	Sample Standard Deviation(s)
B87	@SQRT(B86)
A88	Range of the Set
B88	+B82-B81
A89	m-s
B89	+B83-B87
A90	m+s
B90	+B83+B87
A91	m-2s
B91	+B83-2*B87
A92	m+2s
B92	+B83+2*B87
A93	m-3s
B93	+B83-3*B87
A94	m+3s
B94	+B83+3*B87

To classify the data into a frequency table one chooses a class interval length, boundaries of classes and class marks. Then we can draw histograms and calculate the percentage of values lying in the class. (Table 3)

The frequency table and histogram of weight, BMI, and the percentage of calories can be drawn in the same way. The histogram of energy requirement and intakes are drawn in one and the same graph.

Table 3 . Frequencies and Z-values of Height

Class Mark	Upper Boundary	Frequency	%	Lower Boundary	Value of Z	Area to the Left of Z(P)	P× Numbers	Frequency	%
146.25	147.49	1	1.35	145.0		0.0000	0.00	0.84	1.13
148.25	149.99	3	4.05	147.5	2.28	0.0113	0.84	2.32	3.14
151.25	152.49	4	5.41	150.0	1.72	0.0427	3.16	5.94	8.03
153.75	154.99	10	13.51	152.5	1.15	0.1230	9.10	11.44	15.46
156.25	157.49	16	21.62	155.0	0.59	0.2776	20.54	15.57	21.04
158.75	159.99	16	21.62	157.5	0.03	0.4880	36.11	15.83	21.39
161.25	162.49	14	18.92	160.0	-0.53	0.7019	51.94	12.02	16.24
163.75	164.99	5	6.76	162.5	-1.09	0.8643	63.96	6.45	8.72
166.25	167.49	4	5.41	165.0	-1.66	0.9515	70.41	2.61	3.53
168.75	169.99	1	1.35	167.5	-2.22	0.9868	73.02	0.98	1.32
		74	100.00			1.0000	74.00	74.00	100.00

(Z: corresponding value for the standardized normal distribution)

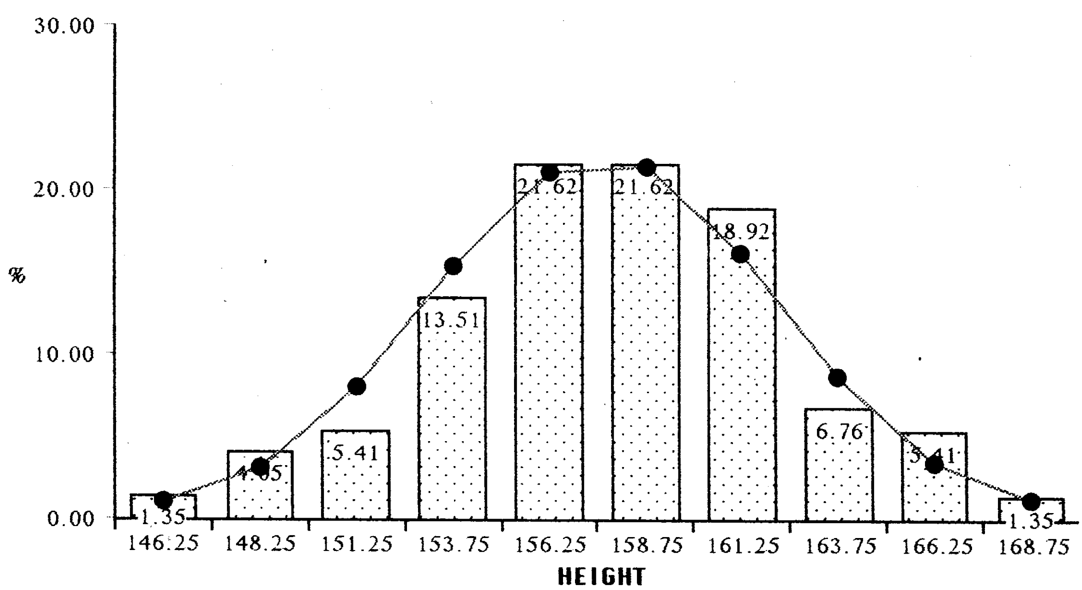


Figure 2. Distribution of Heights. The dotted Curve stands for the Normal Distribution.

The normal distribution is the most useful probability distribution for continuous variables, problems of estimation and tests of hypothesis. Table 3 shows the areas under various parts of the normal curve of heights. Based on the sample size, a frequency table of the normal curve can be obtained. The bar frequency graph and the normal curve of the height-distribution as shown in Figure 2, indicate whether a distribution of height is normal or not. Then, the average, the median, and the mode of the set are drawn on the same graph; these three are almost the same: this fact is the characteristic of the normal distribution.

In data which are associated with two variables, the result of the regression analysis, such as in Table 4, shows the relationship between height and weight. The correlation coefficient is the root of the coefficient of determination, R^2. The rank correlation coefficient is calculated by the ranks of height and weight, using the "@RANK function".

The investigation of the relationship of two variables (x and y) usually begins with an attempt to discover the approximate form of the relationship by plotting data points in the x,y plane (a scatter diagram). By means of it, one can quickly discern whether there is any pronounced relationship, and whether the relationship may be treated as approximately linear.

Usually one studies the relationship between two variables in the hope that any relationship that is found can be used to make estimates or predictions of a particular one of the variables. Thus, in studying the correlation between height and weight, the intention obviously is to use the relationship to predict weight from height. Methods that have been designed to handle prediction problems are known as regression methods. For this purpose, it suffices to draw a sample regression line and two lines parallel to this straight line which are shown in Figure 3. Here, these two lines show an

Table 4. Result of Regression Analysis of Height and Weight

y-Intercept	-75.23
Standard Errors of Estimate	5.13
R^2	0.33
Number of the Set	74.00
Degrees of Freedom	72.00
x-Coefficient	0.81
Standard Error of x-Coefficient	0.13
Correlation Coefficient	0.58
Rank Correlation Coefficient	0.97

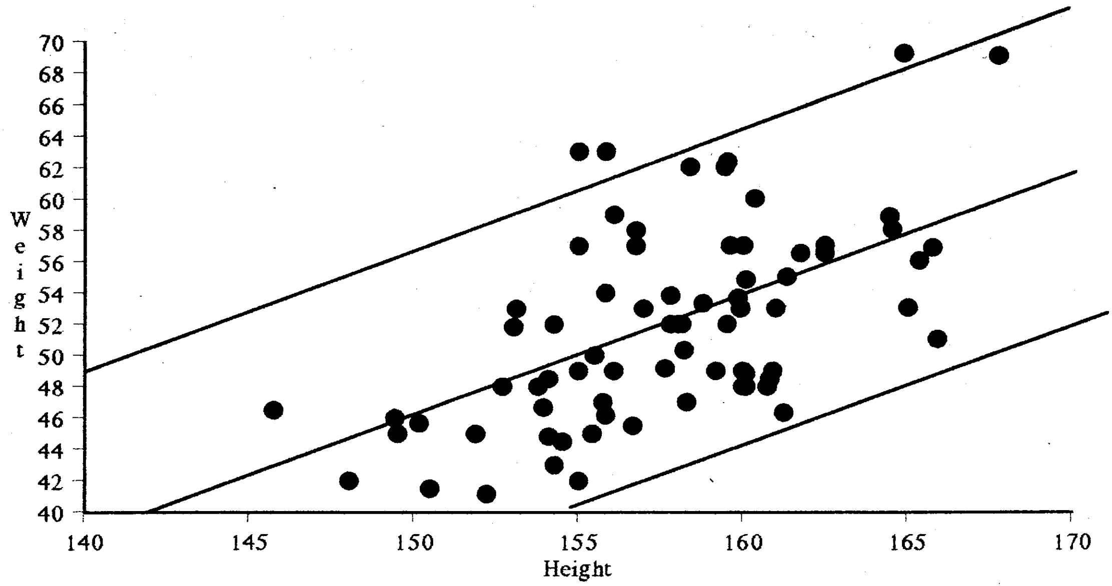


Figure 3. Weight as Function of Height

(Scatter diagram and the regression line for weight on height)

approximate 68 percent prediction bands, under the assumption of the normal distribution.

In our problem,we have scatter diagrams and sample regression lines for each of the pairs of variables selected from height, weight, BMI, and the percentage of calories.

Thus the descriptive part of statistics begins by considering properties of the sample numerically and graphically. We then go on to the consideration of the statistical inference, drawing conclusions about the source of the data.

2) STATISTICAL INFERENCE

The hypothesis that the sample distribution is normal is tested by the value of "chi-square", by making a worksheet such as Table 3.

Correlation coefficients are tested by looking up the relevant tables.

Other tests and inferences are done with electronic hand calculators.

3. AN EXPERIMENT WITH THE METHOD

The part of descriptive statistics which is concerned with collecting and summarizing data is mainly accomplished by drawing graphs and by calculating many different kinds of quantities for an arithmetical description of distributions. To draw graphs and to calculate characteristic numbers may be done by electronic hand calculators, spreadsheets, or various statistical packaged-software.

To make one's own computer programming for the above purpose requires designing of algorithms and the knowledge of computer languages. The question has been raised concerning the ability of certain students in this area. The use of a calculator requires too much time and is prone to error. The use of spreadsheets or a statistical packaged-software do not require designing of algorithms and computational skills. This has been particularly true in the introductory course which most students are not mathematically mature. But the use of a statistical packaged-software tends to attach importance to outputs alone. On the other hand, by the use of spreadsheets one is able to learn various mathematical ideas.

Unfortunately mathematics is one of hated courses. But even students, who can not understand the meaning of the summation symbol Σ, find no difficulties in adding the numbers in the range by using the "@SUM function".

Sometimes, software activities require unessential manipulative skills of students in order to attain the goal. However, in my experimental course, students can now see the usefulness of spreadsheets as more than just manipulative skills or means of solving problems. For example, when they "copy" a formula, they can see cell addresses change automatically. But, when they want to "copy" a formula without changing certain cell addresses, they need to specify "absolute references". In this way, they can understand visually the notion of the "absolute reference".

With the help of spreadsheets, students are able to see immediately what happens, so they learn how to do meaningful work, and at the same time, their desire to learn increases.

Moreover, this is a nice way both to animate all the students in a class and to give them the thrill of discovery. The use of computer graphics or functions can dramatically enhance students' understanding of the main concepts and, for example, methods of normal distribution and linear regression.

While all the students in my course learned statistical inference with hand-held calculators, some of them proceeded to figure out how to use spreadsheets for testing statistical hypotheses.

Figure 4 shows the number of students who enrolled and completed the course from 1988 to 1992. Students from 1988 to 1989 used only hand-held calculators, because we hadn't spreadsheets. Students from 1990 to 1992 used spreadsheets and hand-held calculators by the above mentioned method. The number of students who used spreadsheets increased rapidly. This indicates that using spreadsheets creates motivation for enrollment and completion of the course.

In the introductory course of nutrition, a class was divided into several groups, each of which contained students who had previously studied mathematical statistics, and they used spreadsheets for treating numerical data. The facts that the information processing is useful for the nutritional course and that it uses data from their major subject are some factors for increasing the number of students who completed the course.

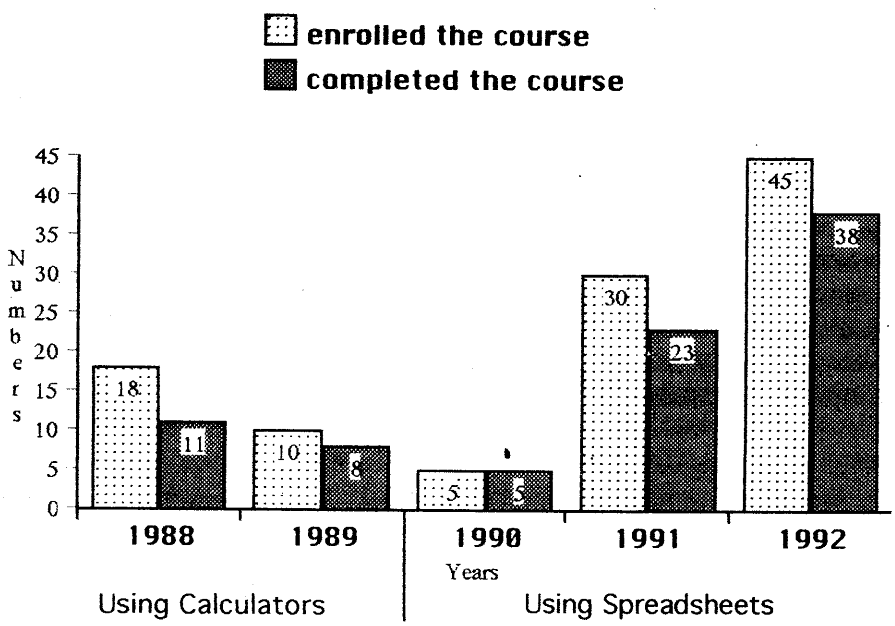


Figure 4 . Number of Students in the Mathematical Statistics Course

(The total number of students in the department is 88.)

4. CONCLUSION

The descriptive statistics require considerable calculations. Students learn how to calculate characteristic values and correlation coefficients, etc., and so they spend a lot of time to do calculations. But the use of spreadsheets makes them to think more about whether the answer makes sense, and less about the correctness of the calculation. Especially graphics and built-in functions are useful.

As we take concrete examples from their major subject, they are interested in learning, and motivated to attend classes. Willingness to participate actively is the key to success in learning. The use of spreadsheets will be helpful not only to learn statistics but also to do calculations of nutrition and presentation as dietitians.

USING THE COMPUTER TO FACILIATE PROBLEM SOLVING IN SECONDARY SCHOOL MATHEMATICS

IAN ISAACS
NORTHERN TERRITORY UNIVERSITY
Darwin, NT 0909, Australia
Eudora: isaacsi@darwin.ntu.edu.au

The simplest and easiest heuristic to teach students learning the techniques of problem solving is that of " Trial and Error". Because it is so natural most mathematics teachers quickly pass it over to go on to more sophisticated strategies such as " Working Backwards" and "Looking for Patterns".

Trial and Error

This strategy is a development of random (or wild) guessing. It differs in that the problem solver goes about it in a systematic manner making a conjecture based on the situation at hand. The student them tests his guess against the data and conditions given in the problem ("The Trial")to see how close or distant the suggested outcome is to the goal ("The Error") and consequently adjusts the next guess in the light of this discrepancy. In other words the student is consciously trying to learn from his/her mistakes.

For example if the following problem is given to a class of students which is about to start the study of how to solve quadratic equations many of them might use a trial and error strategy to find a solution:

Find the Number

Find the number such that three times its square
is the same as six less than eleven times the number.

Some likely approaches you might see are: *Random Trial & Error* - the student picks a (small) positive integer tries it . It doesn't work; tries another which is apparently unrelated to the first.
Systematic Trial & Error - the student makes a table of values starting at 1 (or zero)for the unknown number and begins to systematically check each of them to see if they satisfy the conditions of the problem.
Translating the Problem into an Algebraic Equation- the student gets a correct equation for the problem and tries to solve it using a method for solving simultaneous linear equations; or the student gets an

incorrect equation (say $3x^2 = 6 - 11x$) and also tries to solve it using linear simultaneous methods.

If we focus on the systematic trial and error approach we can produce a table in which is listed x, $3x^2$, $11x - 6$, and a "Comments" column:

x	$3x^2$	$11x - 6$	COMMENTS
0	0	-6	$3x^2 > 11x - 6$
1	3	5	$3x^2 < 11x - 6$
2	12	16	$3x^2 < 11x - 6$
3			
4			

The students will soon discover that for $x = 3$, $3x^2 = 11x - 6$. They will be inclined to stop at that point. You might ask them individually, "Are there any other numbers for which these conditions hold?" They are likely to be puzzled by this question. Some will suspect that the obvious answer ("No") is not correct but cannot think of a possible other number. Others will promptly say "No" without a moment's hesitation. You draw their attention to the relative change in sizes of $3x^2$ and $11x - 6$ as x goes from 0 to 1. It seems reasonable to assume that somewhere between 0 and 1, $3x^2$ is equal to $11x - 6$.

Using a spreadsheet program (for example EXCEL) this conjecture can be explored for values of x between 0 and 1, at intervals of 0.1:

x	$3x^2$	$11x - 6$
0	0	- 6
0.1	.03	-4.9
0.2	.12	-3.8
0.3	.27	-2.7
0.4	.48	-1.6
0.5	.75	-0.5
0.6	1.08	0.6
0.7	1.47	1.7
0.8	1.92	2.8
0.9	2.43	3.9
1	3	5.0

If the rest of the class is still working towards finding a number you can ask those who have finished to plot the graphs

$y = 3x^2$ and $y = 11x - 6$ on the same axes using ANU GRAPH (say). Then from their graphs they can determine the other (approximate) value of x at which the curves intersect (see Figure 1 below).

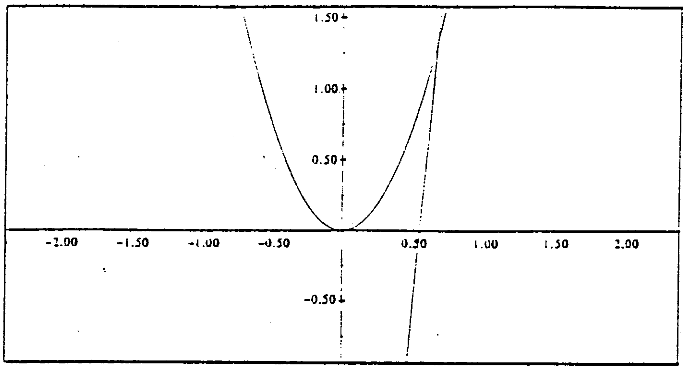


Figure 1: Graphs of $y = 3x^2$ and $y = 11x - 6$

As a first approximation they should get $x = 0.6$ Or 0.7. However at $x = 0.6$, $3x^2 > 11x-6$ and at $x = 0.7$, $3x^2 < 11x - 6$. This suggests that $3x^2 = 11x-6$ between $x = 0.6$ and $x = 0.7$. Using values of x between 0.6 and 0.7 at intervals of 0.01 in the spreadsheet program gives the results shown in Table 1 below:

	A	B	C
1	x	3*x*x	(11*x)-6
2	0.6	1.08	0.6
3	0.61	1.1163	0.71
4	0.62	1.1532	0.82
5	0.63	1.1907	0.93
6	0.64	1.2288	1.04
7	0.65	1.2675	1.15
8	0.66	1.3068	1.26
9	0.67	1.3467	1.37
10	0.68	1.3872	1.48
11	0.69	1.4283	1.59
12	0.7	1.47	1.7

Table 1: Values of $3x^2$ and $11x-6$ for x between 0.6 and 0.7 at intervals of 0.01

From this table it seems that for some value of x between 0.66 and 0.67, $3x^2 = 11x - 6$. A directed discussion led by the teacher at this point should get the students to conjecture that $x = 2/3$ is the number we are looking for. A numerical check will show this to be so. The ensuing discussion should highlight the points that using a spreadsheet program will speed up the process of trial and error immensely. That the results will be expressed in decimal notation which will not always be easily identified with well known rational or irrational numbers such as 1/3, 1/7, √2, √3, etc. That drawing a graph or graphs (preferably with a computer program or graphic calculator) will quickly give an estimate of the required value(s).

The well known problem of finding the shortest path from the farmer to his cow illustrates the speed with which the spreadsheet approach can suggest the solution to this problem.

> The Shortest Path from the Farmer to his Cow
>
> A farmer lives near a river which is 90m from his home (H). He has a cow (C) which is tied 100m from the house and 150m from the river. Every morning the farmer takes a bucket to the river, fills it, and then takes it to his cow. Where is the point (P) on the river bank such that the path from the house to the river and then to the cow is the shortest possible route.
> [Assume that the river bank is a straight line.]

Most students are able to draw a diagram (see Figure 2) to illustrate the problem. They

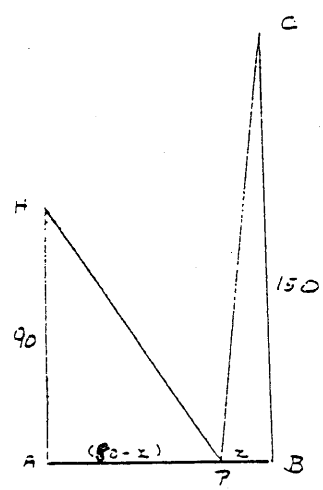


Figure 2: The path of the farmer to his cow

also deduce that AB = 80m, and that HP+PC can be expressed in terms of x, where BP $=x$ metres. The algebraic expression for d (= HP+PC) is quite daunting :

$$d = \sqrt{[150^2 + x^2]} + \sqrt{[90^2 + (80-x)^2]}\ .$$

Most students who attempt to find the minimum value of d using the calculus soon find that they are faced with some even more complex expressions which they cannot, or are unable, to handle. Using a systematic trial and error approach it is possible to tabulate various values of x between 0 and 80 and calculate the corresponding values of d. A fairly tedious task using paper and pencil and a calculator. Using the EXCEL spreadsheet the values for d are quickly calculated as shown in Table 2:

	A	B	C	D	E	F
1	x	150*150+(A1	90*90+((80-A	(B1)^(1/2)	(C1)^(1/2)	D1+E1
2	0	22500	14500	150	120.415946	270.415946
3	10	22600	13000	150.332964	114.017543	264.350506
4	20	22900	11700	151.32746	108.166538	259.493998
5	30	23400	10600	152.970585	102.956301	255.926887
6	40	24100	9700	155.241747	98.488578	253.730325
7	50	25000	9000	158.113883	94.8683298	252.982213
8	60	26100	8500	161.554944	92.1954446	253.750389
9	70	27400	8200	165.529454	90.5538514	256.083305
10	80	28900	8100	170	90	260

Table 2: Values of d for the corresponding values of x from $x = 0$ to $x = 80$

By examining the entries in column F it is seen that the minium value of d (= 252.98) occurs when x = 50. If the students accept this value as a possible solution one can then discuss the relation between the two triangles PBC and PAH when PB = 50m. It is them possible to develop a proof that x = 50 is the desired solution. The ANU graph of d = f(x) also supports the conjecture that d has a minimum value at x = 50 (see Figure 3).

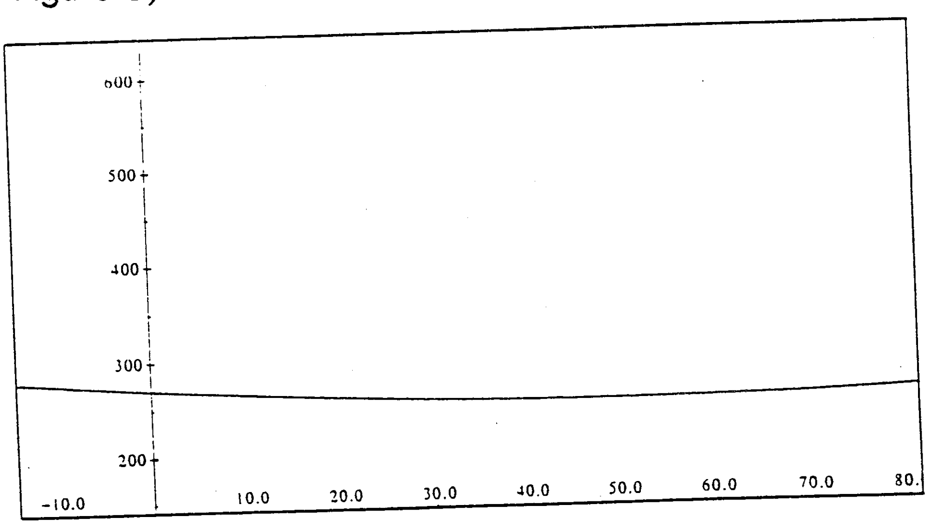


Figure 3: Graph of d =f(x)

As a final example of the power of the trial and error strategy when used with a spreadsheet program we will attempt to find the dimensions of a triangle when we are given its perimeter and area.

The Dimensions of a Triangle given its Perimeter & Area

Find the lengths of the sides of a triangle whose area is $64cm^2$ and perimeter is 36cm.

The analogous problem for a rectangle is fairly straightforward but it soon becomes obvious that there are too many possibilities for the triangle. However if we focus on the fact that these triangles have a constant perimeter and impose the condition that they have a base of fixed length then it is easy to see that the locus of the apex (i.e. the vertex opposite the base) is an ellipse (see Figure 4 below). From which it is easy to show that the isosceles triangle on that base with perimeter 36cm has the maximum area for all triangles for that base.

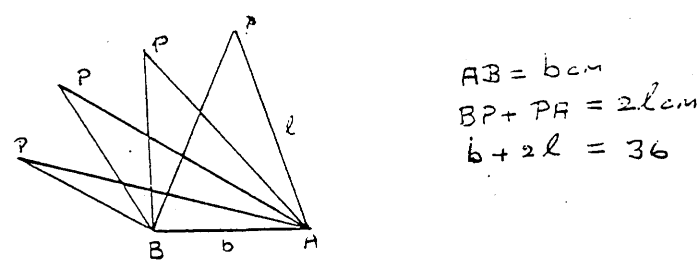


Figure 4: Locus of P for triangles APB on fixed base AB and fixed perimeter.

One can then use a spreadsheet program to examine the areas of isosceles triangles with bases of different lengths (*b* cm) all of which have a constant perimeter of 36cm as shown in Table 3:

	A	B	C	D	E	F
1	b	L	0.5*A1	(B1)^2-(C1)^	(D1)^(1/2)	C1*E1
2	2	17	1	288	16.9705627	16.9705627
3	4	16	2	252	15.8745079	31.7490157
4	6	15	3	216	14.6969385	44.0908154
5	8	14	4	180	13.4164079	53.6656315
6	10	13	5	144	12	60
7	12	12	6	108	10.3923048	62.3538291
8	14	11	7	72	8.48528137	59.3969696
9	16	10	8	36	6	48

Table 3: Areas of isosceles triangles of perimeter 36cm

From the table it is seen that the equilateral triangle of side 12cm has the maximum area of 62.35 cm^2. If one accepts the theorem that for triangles of a given perimeter the equilateral triangle has the maximum area then it follows that there is no solution to this problem.

Conclusion

There are many heuristics which can be used when solving problems of which trial and error has been conventionally regarded as the most basic and least elegant. As a result most mathematics teachers have tended to pooh-pooh its use by their more mature students. However Polya pointed out in *How to Solve It* that many problems fall naturally into two parts: first, find a solution or solutions to the problem then, secondly, justify that the solutions are valid. Spreadsheet programs provide us with a fast and powerful aid in finding such solutions.

HP-48 INTELLIGENCE IN THE MATHEMATICS CLASSROOM AND LABORATORY AN INTERNATIONAL PERSPECTIVE

Richard D. Järvinen, Ph.D.
Winona State University
The Minnesota State Universities
Department of Mathematics and Statistics
Winona, Minnesota 55987
USA
rdjarvinen@vax2.winona.msus.edu

Abstract

The Hewlett-Packard HP-48 handheld computer serves more than one categorical pedagogic role in the mathematics classroom or laboratory. In support of this statement the author draws upon his experience in teaching college mathematics in both the United States and, recently, Japan. More specifically, he draws upon his experience over the past seven years where the HP-48 or its predecessors have been a required and integrated component in the calculus courses he has given. Examples are provided to show various classroom or laboratory roles of the HP-48 as 1) "supercalculator" to produce computations at stages in a solution procedure, 2) "photographer" to capture full portraits of algorithms for later viewing or implementation, 3) "expert" to manifest the presence of an artificial intelligence (expert system) in the context of concept engineering, and 4) "experimenter" to investigate, as well as to help create, conjectures. The author indicates how in each of these four settings the HP-48 can support a threefold approach to doing mathematics: analytically, numerically, and graphically. Perspectives of the author are seasoned by his recent formal training and teaching in the adjacent field of computer science as well as through interactions with students and faculty involved with the use of technology in mathematics.

HP-48 Intelligence in the Mathematics Classroom and Laboratory An International Perspective

The Hewlett-Packard HP-48 handheld computer can serve many roles, categorically different pedagogical roles, in the mathematics classroom or laboratory. The intent of this paper is to give credence to the latter statement. Specifics in the paper provide substance from which a talk can be formed to show the various roles that the HP-48 can assume. These roles are identified, and examples with the HP-48 in these roles are provided.

The reader will find here the use of a matrix to describe the different roles of the HP-48 (hereafter simply called the HP) as it is cast in various classroom or laboratory settings. The roles and associated examples are organized into a matrix of four rows and three columns for ease of subsequent reference.

Row 1 identifies the HP in the role of "supercalculator", able to perform simple to very complex computations as essential, single steps in the solution of a problem. Row 2 identifies the HP in the role of "photographer" of mathematics, a mathematical memory machine, technology to capture the full portrait of an algorithm for later viewing and/or implementation. Row 3 views the HP as "expert", to render expert mathematical advice as in the branch of artificial intelligence known as expert systems. The user of this expert might be viewed as a system analyst or concept engineer, designer of a master plan to bring about the solution of a problem, calling upon an expert system to provide mathematical expertise/advice/information at various stages in a solution scheme or network. Finally, Row 4 sees the HP in the role of "experimenter" or researcher. In this role the HP is employed to test conjectures or to help create them.

While cast as experimenter, the HP might simultaneously take on another role as well. For example, the HP might then also take on the role of supercalculator, or photographer, or expert, all within the same problematic context. The four roles identified in the above paragraph, therefore, should not be assumed to be mutually exclusive (pairwise disjoint, as mathematicians might say).

The HP can be seen to support three methodologies for doing mathematics, namely, the analytical, numerical, and graphical methods. Moving along further to construct the aforementioned matrix of roles of the HP, let us identify Column 1 with the analytical method, Column 2 with the numerical method, and Column 3 with the graphical method.

The matrix of roles of the HP for the purposes of this paper is thus fully described. To elucidate, let us observe that the second row and second column matrix position portrays the HP as holder of an algorithm designed to carry out the details of a numerical method. We use the terminology 'Role 2,2' to identify this classroom/laboratory role.

Adding to an earlier remark made when just four roles by that point in the paper had been identified, we now mention that the twelve roles now identified in the matrix are also not presumed to be mutually exclusive.

In summary, we see the matrix of roles of the HP having the appearance depicted in the matrix below.

Following the matrix display of the roles of the HP are details of examples or indicators of examples that can be used to illustrate and help clarify the different roles for which the HP can be employed in the mathematics classroom or laboratory. How the HP might wear its different "hats" is revealed.

COLUMN / ROW	analytical	numerical	graphical
supercalculator, no programming required	Role 1,1	Role 1,2	Role 1,3
photographer of mathematical algorithms, memory machine	Role 2,1	Role 2,2	Role 2,3
expert, artificial intelligence expert system	Role 3,1	Role 3,2	Role 3,3
experimenter, researcher, fact finder	Role 4,1	Role 4,2	Role 4,3

Matrix of Roles of the HP-48

Here, then, are examples and indicators of the use of the HP cast in each of the twelve roles.

Role 1,1: To illustrate this role we use the HP to calculate values of the normal probability density function

$$f(x) = \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}$$

for various values of x. The context might be to help analyze the specific features of this widely used probability density function. For example, for what value of the random variable x is the probability density half of what its maximum value is? Or, again, one may wish to compare the values of the normal probability density function to the values of some other probability distribution. The NDIST key found under MTH PROB can be used to compute the values of this function. No programming is required.

Role 1,2: In this role the HP might be used to find the fixed point of the probability density function given above. The numerical method known as Picard Iteration can be used to find the fixed point. The iterative formula in Picard Iteration is given by

$$x_{n+1} = f(x_n)$$

Iterates might be generated until there is no change in the 4th decimal place, should that level of accuracy be desired. The iteration is easy to carry out on the HP, for all one need do is repeatedly press the NDIST key to produce the iterates. A "xerox" copy of each iterate can be made by pressing the ENTER key before generating a new term, consequentially saving all terms on the HP stack for subsequent viewing.

Role 1,3: Employ the HP to graph y = x and y = f(x) for f as above to depict the location of the fixed point. Also graph the function $|f'(x)|$ in a neighborhood of the fixed point to see that its values are less than one, a condition stated in a well known theorem which ensures that Picard Iteration will be successful in creating iterates that converge to the fixed point. Graphically seeing the condition that ensures convergence is especially helpful.

Roles 2,1; 2,2; and 2,3 can be illustrated in terms of the context of the three episodes described above. Role 2,1 would have the HP store the probability density function f(x) as a user defined function. Role 2,2 would have the HP store a computer program that carries out Picard Iteration. The program would be called upon to find the fixed point of f. And, finally, Role 2,3 would have the HP store the graph that shows the fixed point of f as well as the graph that shows that Picard Iteration will work. The graphs can be called up and shown at any time in the future.

A far more dramatic use of the HP in Roles 2,1-2,3 is seen in the following.

Role 2,1: We first describe a rather general setting. Have the HP store a function (as an HP user defined function) that needs to be called upon frequently for evaluation in the context of doing an analysis. Assume the context is one where, by observing the behavior of the functional values one is able to suggest a mathematical conjecture, which per se is a decision of a qualitative nature and, thereby, analytical in nature. The HP is thus seen in the role of an assistant for conducting an analysis.

Specifically, the context might be that of linear programming in operations research with the function in focus being that of the objective function whose values one wishes to maximize. Here, have the analysis be sensitivity analysis. Question what happens to the values of the objective function as the column coefficients for a nonbasic variable, for instance, are altered. How much can the column coefficients for a nonbasic variable be changed before a basis becomes suboptimal, resulting then in a new optimal solution and a new set of basic variables?

In this context the HP can be utilized to help students learn the very nature of sensitivity analysis. How sensitive is the value of the objective function to changes in values of the coefficients/parameters in the objective function or expressions for the constraints? How sensitive is the set of basic variables (to remain as such) to the same latter changes?

Role 2,2: Keep the context that of sensitivity analysis as in the example immediately above. Write an algorithm, then stored in HP memory and coded in the computer language of the HP, that will diagnose what the limits of change in the column coefficients of a nonbasic variable can be so that no change is induced in what the basic variables are. Then employ the latter program in the context of a numerical method where incremental step size is used to step through changes in a given coefficient of a nonbasic variable within the specified limits. As the iteration occurs, it is noted what the new value of the objective function becomes in the cases where the basis does not change.

Role 2,3: Keep the context as in Role 2,1 and Role 2,2. Graph the objective function as a function of change in selected coefficients in the column of a nonbasic variable. Store the graphs for future reference. For instance, the graphs can be shown during class at a timely moment in the explanation of sensitivity analysis.

Note, for those that are interested, the author can provide a handout which he has prepared which conveys a specific example in the context of sensitivity analysis as described above, and which also includes a computer program written for the HP that decides if the set of basic variables will or will not change.

Roles 3,1 and 3,2: In this setting give the HP a definite integral to compute. The HP is "smart" and will render an expert decision in the following sense. Using its very own expertise, the HP will know whether or not it can evaluate the integral by using the Fundamental Theorem of Calculus (FTC). To have the HP make this decision, issue the command EVAL after entering the definite integral into the HP. If it can use the FTC, that is, if the HP knows an antiderivative of the integrand, it will evaluate the integral using the FTC. In this sense of decision making as well as by fact of using the FTC, the HP does analysis; it performs in the Role 3,1. The integral below could be used as an example in this case.

$$\int_0^1 \cos \pi x \, dx$$

On the other hand, if the HP does not know an antiderivative of the integrand, then it instead computes the integral using a numerical integration method. The probability density function for the normal distribution could be used as an example in this case. One might compute the following definite integral.

$$\int_0^1 \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \, dx$$

The HP thus behaves as in Role 3,2. We add that the HP can be forced into performing numerical integration for any definite integral simply by issuing the →NUM command.

If the integrand is graphed, the HP can graphically depict the region in the plane associated with the value of the definite integral and the user can graphically control the limits of integration, which illustrates the HP in the Role 3,3. No programming is required in this setting. The HP is expertly "wired" to generate, for descriptive purposes, these graphics.

In recent research done by the author involving the optimization of a nonlinear function of several variables subject to linear constraints in which both the number of variables as well as their values must be decided, the following function was encountered:

$$A_T = 2\pi N \sigma_x \sigma_y \ln \frac{N p q^{\frac{N-1}{2}}}{1 - P(A) - q^N}$$

Here, N is the variable (a discrete variable in context of the problem) and all other expressions on the right side represent constants that are probability values between 0 and 1, with the exception of the sigma which stand for standard deviations. The question arises whether or not this function in N is monotonic in N for all possible values of the different probabilities. If one thinks of N as a continuous variable for the moment and then computes the derivative with respect to N, the resulting expression is seen to be difficult to analyze by "hand" methods.

In this context of research, one might have the HP assist in the analysis by having it produce the derivative with respect to N as well as values of the derivative, Role 4,1. Is the derivative always positive or always negative? What conjecture might the HP suggest?

Or, consider the two functions that arise when N = k and then when N = k + 1. Form a ratio of them. How does the ratio compare to 1 as k increases? A numerical iterative method can be employed in this context, producing terms of a sequence for the purpose of making a conjecture regarding the monotonic nature of the given function. We thus see the HP in Role 4,2.

Yet, the function in N could be graphed on the HP for different assignments of constants to the probability parameters. We see, therefore, the HP in the context of Role 4,3.

Having used the HP to help diagnose what might be the monotonic character of the function, one could then move to pursue making a proof of what is, in fact, true. The HP can indeed serve as research assistant.

This concludes our tour of the various roles of the HP as conveyed in the Matrix of Roles of the HP-48 displayed above. We hasten to add that in a given problematic circumstance, one may employ the use of the HP in many different ways involving many of its roles.

It has been the experience of the author over the past seven years to observe that generally the students he has taught, both in Japan as well as in the United States, have eagerly accepted the use of the HP in the calculus courses the author has given in which the HP has been required and integrated. Their eagerness seems to stem from their feeling of being empowered by having the HP technology at their command. They have again and again been impressed with the depth and scope of this computational device.

Indeed, student attitude is vitally important in the context of using technology to do mathematics. The author wishes to emphasize the likewise vital importance of explaining to students what the role--what mathematical healthy role--technology is to have in the given course they are taking and in their overall mathematical education.

At his home university, for instance, each mathematics major upon graduation must have acquired skills to do mathematics with the use of technology. At least a dozen other desired characteristics of the graduating mathematics major have been identified as well, including that of acquiring collaborative, team-working skills. In his calculus courses, the author requires students to learn how to be mathematically effective both with and without the use of the HP technology, and examinations reflect this. Examinations are given in two parts: Part I is to be done without the use of technology; Part II allows the use of technology, and for most problems in Part II, technology is a necessity, with the nature of the problem requiring its use.

The author sees the discipline of mathematics as taking a leading role among all disciplines in its use of technology to carry out the methods of its subject. In particular he sees the discipline taking a leadership role in the use of artificial intelligence in the form of expert systems, systems designed to expertly carry out the methods of its subject and playing roles such as described in the body of this paper. There is presently widespread use of mathematical expert systems in the mathematics community in the United States, with Mathematica and Maple as prime examples, to mention just two. These are hardware-software combinations that render expert mathematical advice across a wide spectrum of mathematical topics. Surely the day is to come when all disciplines are served by expert systems designed with their discipline and their special decision making in mind.

Perhaps the most gratifying response amongst all those evaluative statements students have made on course evaluations in courses the author has given in which an HP has been required is this by a calculus student. "When I combine the use of the HP with the calculus I know, I feel there isn't a problem I can't solve." Certainly there are problems that this student cannot solve. But it is this kind of spirit that makes a difference, that makes things happen, that brings people forward to make the serious effort it takes to educate themselves. Let us keep in mind as well the very positive impact instructors have on students when they enthusiastically engage their discipline and their students...with or without technology.

FINDING THE n^{th}-ROOT OF A REAL NUMBER: A COMPUTER-ASSISTED ALGORITHM

by

O. Jayanth and R. Sivaramakrishnan

Department of Mathematics, University of Calicut,
Calicut, India 673 635.

A B S T R A C T

The division-method of finding the square-root or cube-root of a real number is extended for finding the n^{th}-root ($n \geq 2$) by a computer-assisted algorithm which is fast enough for computation. The flow-chart is also exhibited.

FINDING THE n^{th}-ROOT OF A REAL NUMBER: A COMPUTER-ASSISTED ALGORITHM

by

O. Jayanth and R. Sivaramakrishnan
Department of Mathematics, UNIVERSITY OF CALICUT, INDIA 673 635

1. INTRODUCTION:

The division-method of finding the square-root or cube-root of a real number is known and widely accepted. The purpose of this note is to point out a new computer-assisted algorithm based on the division-method to find the n^{th}-root ($n \geq 2$) of a real number N which is of the form

a) N a perfect n^{th}-power

b) N not a perfect n^{th}-power, but the n^{th}-root is a non-terminating decimal

c) N not a perfect n^{th}-power, but the n^{th}-root is a terminating decimal.

The advantage of the algorithm is in reducing the number of iterations needed for fixing a decimal place of the n^{th} root. The algorithm is explained in section 2.

2. THE ALGORITHM:

We are given the number N whose n^{th}-root is to be found out.

The technique of the algorithm is based on the following rules:

Rule 1: Grouping the digits of N into sets of n digits:

Writing the number N in decimal notation (base 10), suppose

$$(2.1) \qquad N = a_m \, a_{m-1} \, \ldots\ldots \, a_2 \, a_1 a_o \, . \, b_o \, b_1 \, b_2 \, \ldots \, b_k$$

The digits of N are to be grouped into sets of n digits, to the left of the unit's place in the case of the integer part of N and to the right of b_o in the case of the decimal part.

We denote the groups to the left of a_o by j_o, j_i, .. j_r. The group j_r in the extreme left may contain 1, 2 ... or n digits. We call the numbers of these sets as J_o, J_1, ..., J_r. The group on the extreme right in the case of the decimal part must contain exactly n digits. This is possible if we add sufficient number of zeros to the extreme right.

Rule 2: Finding a divisor:

(i) The first digit on the extreme left of the n^{th}-root is that greatest one-digit number whose n^{th}-power is less than or equal to the number formed by the set of digits on the extreme left, that is, in J_r.

When the grouping is done as per rule 1, the first divisor is the $(n-1)^{th}$ power of that single digit obtained by rule 2(i).

(ii) The other divisors are calculated from the formula

$$(2.2) \quad \text{divisor} = \sum_{k=1}^{n} \binom{n}{k} (10 \text{ x quotient})^{n-k} x^{k-1} \quad (0 \le x \le 9)$$

where x is the number that will determine the 'digit' occurring in the n^{th}-root.

When $x \neq 0$, we could use the alternate formula

$$(2.3) \quad \text{divisor} = \frac{(10 \text{ x quotient} + x)^n - (10 \text{ x quotient})^n}{x}$$

when x = 0, the divisor is written as

$$(2.4) \quad \text{divisor} = n\,(10 \text{ x quotient})^{n-1}$$

which is the limiting form of the right side of (2.3) as $x \to 0$. We write

$$(2.5) \quad x_o = \text{quotient at step 1}$$

x_o is obtained by rule 2(i).

We observe that the quotient at step 0 is 0. Then, we obtain the x_i's in successive steps (by iteration). Using the successive values of x_i's, we get the successive quotients which determine parts of the digits of the n^{th}-root one by one.

The quotient at step 2 is given

$$(2.6) \quad (\text{quotient})_{(2)} = 10\,x_o + x_1$$

where x_1 is to be determined from the formula for

(2.7) dividend $= \left\{ J_r - (x_0 x \text{ first divisor}) \right\} 10^n + J_{r-1}$

where J_r and J_{r-1} are already defined in rule 1 and x_1 is the greatest single-digit number obtained from

(2.8) $$\frac{\text{dividend}}{n\ (10 \times \text{quotient at step } 1)^{n-1}}$$ (see (2.5))

subject to the condition that the product of the divisor obtained from (2.2), (2.3) or (2.4) and the corresponding digit of the n^{th}-root should not be greater than the corresonding dividend shown in (2.7). If this condition is not satisfied, we decrease x_1 (of (2.6) by 1 and test for the condition again. The process is repeated till the condition is satisfied. This is to make the algorithm faster.

In general, the quotient at the i^{th}-step ($i \geqslant 2$) is obtained from the following recurrence relation of successive steps:

(2.9) quotient at the i^{th}-step = 10(quotient at the $(i-1)^{th}$ step) $+ x_{i-1}$

The dividend at the i^{th} step is given by

(2.10) dividend at the i^{th}-step = (dividend at the $(i-1)^{th}$-step − divisor at the $(i-1)^{th}$ step x x_{i-1})$10^n + J_{r-i}$

Then, the quotient at the i^{th}-step is obtained from the formula

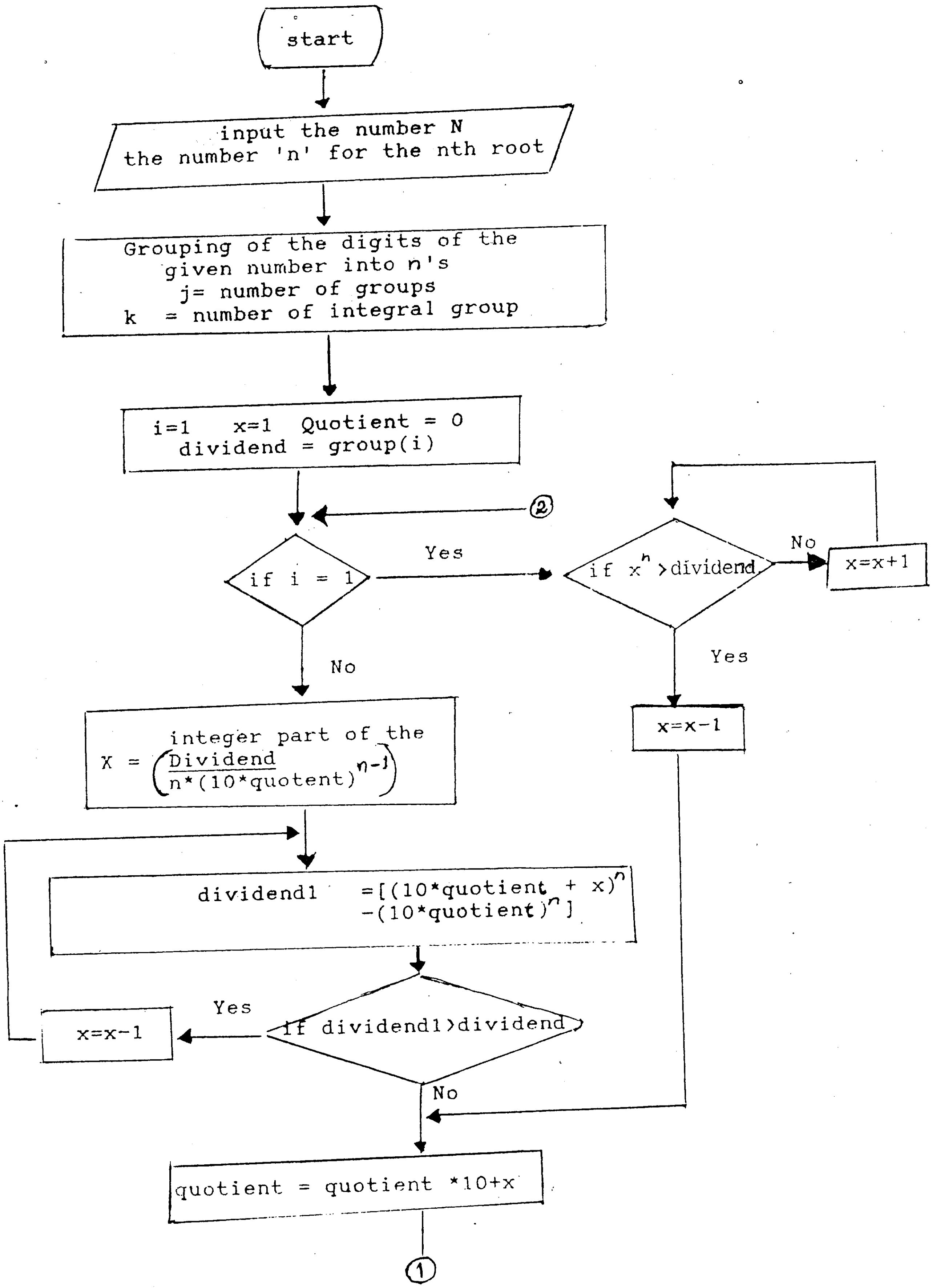
(2.11) $$\frac{\text{dividend at the } i^{th}\text{-step}}{n\ (10 \times \text{quotient at the } (i-1)^{th} \text{ step})^{n-1}}$$

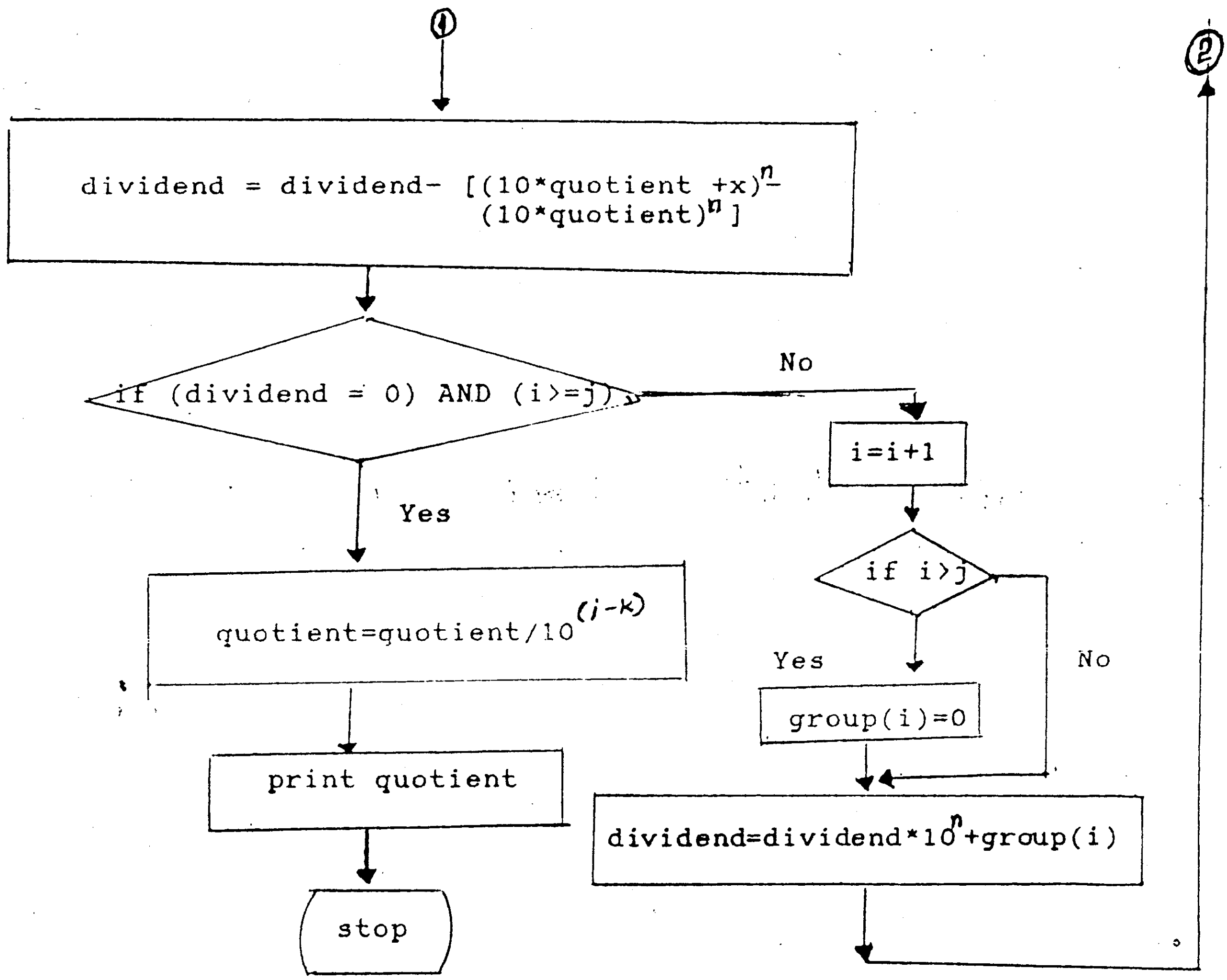
subject to the condition as laid down for formula (2.8).

Rule: 3 If the number N is not a perfect n^{th} power, we put a dot (.) on the right of the unit's place followed by groups of n zeros.

The iteration terminates in a finite number of steps. Then, either the dividend becomes zero or the number of groups of n digits in N is exhausted by Rules 1, 2(i) and 2(ii). It is also desirable to stop the process of iteration if the arrive at sufficient number of digits as required for the n^{th}-root. Further, the n^{th}-root is written by fixing the decimal point at the appropriate position of the quotient at the $(i+1)^{th}$-step (say) which is noted by observing the number of groups r in the integer part of N.

The process of iteration is exhibited in the flow-chart shown below:


start
input the number N
the number 'n' for the nth root
Grouping of the digits of the
given number into n's
j= number of groups
k = number of integral group
i=1 x=1 Quotient = 0
dividend = group(i)
2
if i = 1
Yes
No
if x^n >dividend
No
x=x+1
Yes
x=x-1
integer part of the
X = (Dividend / n*(10*quotent)^(n-1))
dividend1 =[(10*quotient + x)^n
-(10*quotient)^n]
Yes
x=x-1
if dividend1>dividend
No
quotient = quotient *10+x
1



1
dividend = dividend- [(10*quotient +x)^n- (10*quotient)^n]
if (dividend = 0) AND (i>=j)
No
Yes
i=i+1
if i>j
Yes
No
group(i)=0
dividend=dividend*10^n+group(i)
quotient=quotient/10^(i-k)
print quotient
stop
2

3. ILLUSTRATIONS:

First, we illustrate the method of finding the fourth-root of the number N = 7311616.

As per rule 1, we group the digits of N as 731,1616. Since $5^4 = 625$ and $6^4 = 1296$ greater than 731, the first digit (from the left) of the fourth-root is obtained as 5.

The divisor at step 1 is $125 = 5^3$ as per rule 2(i). Subtracting 625 from 731 be get 1061616 as the dividend in step 2. This is obtained by placing the digits 1616 to the right of 106.

Using (2.3), the divisor at the second step is

$$\frac{(5\times 10+x_1)^4 - (5\times 10)^4}{x_1} \qquad \text{(see (2.3))}$$

which is the same as

$$4 \times (10\times 5)^3 + 6\times (10\times 5)^2 x_1 + 4 \times (10 \times 5) x_1^2 + x_1^3$$

and x_1 is the greatest single-digit number given by

$$\left[\frac{1061616}{4\ (10 \times 5)^3}\right] = 2$$

Here [x] denotes the greatest integer not exceeding x.

Therefore, the divisor at the second step is

$$4\times (10 \times 5)^3 + 6\times (10 \times 5)^2\times 2 + 4\times (10\times 5)\times 4+8 = 530808$$

The divisor at the second step multiplied by x_1 = 1061616, as x_1 = 2. This gives the quotient at the second step as 5x10+2 = 52. The dividend in step 3 is zero and the process of iteration terminates. The fourth-root of 7311616 is 52.

The actual working is shown below.

Fourth-root of 731,1616 is 52.

```
                 52
           ------------
          |731,1616
    125   |625
          |------------
 530808   |1061616
          |1061616
          |------------
          |     0
```

The second example is about finding the fifth root of 82.12345. The working is as shown below:

```
                     2.41
               ------------
              |82. 12345
          16  |32
              |------------
     1190656  |50 12345
              | 4762624
              |---------------
 16727617201  |  24972100000
              |  16727617201
              |---------------
              |   8244482799
```

2.41 is the fifth-root of 82.12345 upto two places of decimals.

February 15, 1995.

ON SOME COMPUTER AIDED RESEARCH STRATEGIES

Zhuhan JIANG
Department of Mathematics, Statistics and Computing Science
University of New England, Armidale NSW 2351, Australia
Email: zhuhan@neumann.une.edu.au

We examine some of our strategies on integrating the up front computer technology with the academic research, the tertiary student needs analysis and the self teaching-evaluation. We shall first examine the main features and the applicability of symbolic calculation and computer simulation through the analysis of several practical cases, with stress on the hidden pitfalls and the relevant tips. We shall then propose a simple mechanism, based on the logging of online help retrievals during students' computer lab sessions, to determine students' underlying understanding of the subjects and thus to self-evaluate and improve the teaching quality of a particular course through the computer aided student feedbacks.

Keywords: research, teaching, symbolic manipulation, computer simulation, needs analysis, teaching self-evaluation

1. Introduction

Modern technology, particularly that of computers, has brought about great advance in almost all aspects of the society. Many impossible feats only a short while ago in history have become commonplace. Millions of small and micro computers practically made many people's office or home a mini computing center. Mathematics as one of the oldest subjects in science [JS] has not been spared from the ever increasing technology fever either. In the educational sector in particular, many state of art equipment and computer software have made the teaching and learning more efficient, more fulfilling and often more fun to all parties concerned. It is the indispensable role of computer technology in today's research and teaching of mathematics that we wish to pursue in this paper.

There are mainly two categories in the utilization of modern computers. The first and perhaps the most eye-catching one is the knowledge presentation, with which versatile knowledge data are arranged in the most glamorous and easiest digestible ways. Methods in this category target the efficient transmission of knowledge or teaching materials to typically a group of students. In contrast to the first category, which often depends much more heavily on the *all-purpose* technology than the methods that use it, the second category contains the general study and research through the use of often more crude form of computational power. It is the methods and strategies within this second category that we shall study with some varieties and depth. Our main purpose is to demonstrate the integration of the *computing* power with the general mathematical research and with the research that would lead to teaching improvement. To be precise, we shall be concerned with the following aspects.

I. Symbolic calculation as crude manipulation force. Many mathematical software packages are designed for this purpose – MAPLE, REDUCE, MUMATH and DERIVE to name a few, and some are developed over such platforms [MF,CM]. For this feature we shall look into a Painlevé integrability test [BJ] on a physically important Three-Wave Interaction (3WI) system. The main point we wish to address here is that in symbolic calculations of great complexity, the programming methodology could make a huge difference on the demands on the computing hardware. This is also an example in which the hardware was not powerful enough in the first few attempts until a more sensible algorithms was introduced. More technical details, along with those to be described below, will be expounded in later sections.

II. Symbolic proof to non-rigorous results. In certain research fields such as mathematical or theoretical physics, a great deal of theory and formulas are derived without the rigor that would satisfy a mathematician. Though such theories or formulas would eventually be validated by certain physics phenomena, a direct mathematical proof without any physical experiments, using symbolic packages to verify the proposed solutions for instance, would be advantageous. For this purpose we shall demonstrate with another piece of recent research work of ours, in which some proposed results [JB,JH] had to be first derived via a non-rigorous approach and then explicitly validated through the use of a computer.

III. Simulation to guide and validate a theory. We first note that in some research areas, a great proportion of the work is empirical – which often requires sufficient first hand data to provide insight into the nature of the research topics and their theoretical prospects. Moreover, simulation can also be used to validate certain proposed theory. We shall thus examine another research case in the field of Cellular Automata (CA). We shall show how our devised extensive computer simulation shed lights on the related theory while disproving certain ill-conceived conjectures. This is an example where the use of modern technology could prevent researchers from wasting unnecessarily their precious time on proving or disproving an untested proposition.

IV. Online help, student needs analysis and teaching self-evaluation. The objective here is two fold: the first is to to provide instant helps in classroom or computer laboratory with lecture notes and extra examples etc, while the second is to log all the help retrieval sessions as feedbacks to the course coordinators. The concept of classroom assessment by the use of computers is not new [W], but our proposed scheme is quite different from the traditional methodology. We note that the device of such online help system, though important and valuable, fall into the first category that we defined earlier on, and is itself not of our main interest as we have already limited ourselves to the study of the second category. In other words, our profound interest here lies in the analysis of the underlying feedback data through the online help system. The aim of this mechanism is to provide a lecturer with a type of evaluation of students' interests and needs, so that lecture topics and time allocation could be readjusted to the maximum advantage to the students. The strategy that we wish to address later on is based on the subject logging and time-use logging of the students' help requests. We shall thus propose a simple and yet quite efficient method to self-evaluate the teaching quality and to provide hints to adjusting the direction of the future offering of the same course.

All the above features will be dwelt upon later on, along with the supporting practical examples. We shall also address some technology-related aspects in the use of computers, and thus give only the minimum scientific background materials for the related examples. In fact, this paper is organized as follows. In section 2, we first present and analyze some strategies used in performing heavy symbolic calculations, along with the study on the possible hidden pitfalls. The example for this purpose will be the Painlevé integrability test of the 3WI system. We then demonstrate with another simpler example the machine validation role often played in the research of mathematical physics. Section 3 is designed to show the visualization and intuitiveness provided by a computer simulation, and how they could also lead researchers along the right track of theoretical pursuit. A model of solitonic cellular automaton is examined for the various gains induced by the corresponding simulation. In section 4, we move to a research more related to the teaching improvement. We propose and study in some details a method based on the logging of student online help. We then show how we can use this method to evaluate the student needs and the subject difficulty distribution, and provide a means of self teaching-evaluation and improvement. Finally, a brief summary is presented in section 5.

2. Critical power of symbolic calculations

In this section, we shall be concerned with the proper use of symbolic calculations in the general mathematical research so as to maximize the underlying computing power. We shall present two simple case studies to demonstrate the different aspects involved. The Painlevé integrability test on the 3WI system, as the first example, will be used to display a few typical features and hidden pitfalls that often exist in *heavy* or huge task of symbolic calculations, along with the proposed solution measures. These features and pitfalls, in contrast to those discussed in [MT], are mainly due the heaviness of the calculations. The second example on the explicit solutions of two space dimensional nonlinear Schrödinger (NLS) equations, on the other hand, will be used to show the theoretical proof or validation role played by symbolic computer algebra packages.

The 3WI system is a set of physically important partial differential equations

$$\partial_t w_{jk} = \alpha_{jk}\partial_x w_{jk} + \beta_{jk}\partial_y w_{jk} + \gamma_{jk} w_{j\ell} w_{\ell k} \tag{1}$$

with parameters $\alpha_{jk} = \alpha_{kj}$, $\beta_{jk} = \beta_{kj}$ and $\gamma_{jk} = -\gamma_{kj} \neq 0$ for all $j \neq k$, where indices j, k and ℓ permute 1,2 and 3. The main part of the Painlevé test of 3WI system, to cut the long definition short, is to show [BJ] that eqs.(1) admit series solution in the form of $w_{jk} = \sum_{m=0}^{\infty} u_{jk}(m,y,t)[\phi(x,y,t)]^{c_{jk}+m}$, where $\phi(x,y,t) \equiv x + \psi(y,t)$, c_{jk} are integers and $\psi(y,t)$ is an arbitrary function. The insertion of w_{jk} into (1) thus gives

$$\sum_{m=0}^{\infty} \left\{-\partial_t u_{jk}(m-1,y,t) + \beta_{jk}\partial_y u_{jk}(m-1,y,t) + (c_{jk}+m)\varphi_{jk}(y,t)u_{jk}(m,y,t)\right\}$$

$$\times\left[\phi\right]^{c_{jk}+m-1} + \sum_{m=0}^{\infty}\left\{\gamma_{jk}\cdot\sum_{s=0}^{m} u_{j\ell}(m-s,y,t)u_{\ell k}(s,y,t)\right\}\cdot\left[\phi\right]^{c_{j\ell}+c_{\ell k}+m} = 0 \tag{2}$$

where $u_{jk}(m,y,t) \equiv 0\ \forall m < 0$ and $\varphi_{jk} \equiv \varphi_{jk}(y,t) = \alpha_{jk} + \beta_{jk}\psi_y(y,t) - \psi_t(y,t)$. By equating the coefficients of the same powers of ϕ, a set of equations for iteratively deriving u_{jk} can be obtained, and the equations from the leading orders imply $c_{jk} = -1$. Since the iteration relations in (2) can be rewritten as a vector equation $\mathbf{M}(m)\mathbf{U}(m) + F_m[m, \mathbf{U}(0),..,\mathbf{U}(m-1)]=0$ where $\mathbf{U}(m)= [u_{12}(m), u_{13}(m), u_{23}(m), u_{21}(m), u_{31}(m), u_{32}(m)]^T$, and $\mathbf{M}$ and $\mathbf{F}$ are matrix and vector determined by (2). The remaining task is to show that at the finite number of resonances m, at which (i.e. when det($\mathbf{M}$)=0) the iteration breaks down temporarily, the compatibility is satisfied. This property is essentially what is known as the Painlevé property and is a working indicator of the system's solvability. We are now left with verifying the compatibility at the easily derivable resonances at m=0, 2 and 3. In other words, we need to derive from (2) explicitly *all* the $u_{jk}(m,y,t)$ for m=0,1,2 and then check that $\mathbf{P}(m)F_m$=0 for *any* $\mathbf{P}(m)$ satisfying $\mathbf{P}(m)\mathbf{M}(m)$=0, with m being 0, 2 and 3.

This task is in fact a *huge* symbolic calculation due to the explosiveness on the number of terms in the expansion process. It was first performed on a 486 PC with 15M memory, with the program written in MAPLE V. We ran into constant difficulties of running out of memory until reasonable ways were later found to artificially break down the calculation into smaller pieces. These difficulties would perhaps often characterize a problem of large symbolic manipulation. Our main methods and experiences with such a problem can be summarized as follows.

Maximize automation. Almost all symbolic packages provide interactive processing environment and such environment is very useful for quick solutions or small problems. When a problem becomes sufficiently complex, in the sense that many intermediate non-standard manipulations are required, such method is often clumsy to say the least. The point we are making here is that we may start, if necessary, with interactively operating a symbolic calculation, but once the difficulty is overcome, the process upto that point should be immediately made automatic, i.e. no further human interface should be needed at least upto that point whenever the program is re-run. This will save a great deal of time in the end, as many (incomplete) program runs are often needed before a final solution is obtained.

A masked pitfall. If a program fails miserably due apparently to the lack of hardware capability, the first thing to do is not necessarily looking for a better equipment. Instead, the solution often lies in an improved algorithm. This is especially true for the use of run time memory. A simple increase on the hardware capacity often turns out insufficient anyway.

Notational simplification. In the course of deriving $\mathbf{U}(m)$ iteratively for the above example, time and again we are faced with manipulating or simplifying a huge collection of algebraic terms containing various derivatives such as `D[1,2,3,3,3](u)(x,y,t)` (representing $\partial_x\partial_y\partial_t^3u$). Obviously, when one is only manipulating such terms algebraically, a derivative like `D[1,..](u)(x,..)` is treated no different from a normal variable. However, derivative terms normally occupy more machine memory space and also slow down the algebraic calculation or simplification. Hence it can be much more economical if one first transforms all such derivatives into new simple variables, e.g. denoting `D[1,2,3,3,3](u)(x,y,t)` by just `u102033`, before further iterations or other algebraic manipulations take place. Even if a final result is expected to be very simple, the chances exist (as did in the study of this example) that a machine's capacity is exceeded before the final very simple form is reached from the simplification of a bulk of terms. Of course, we may transform the new variables back to their originals should we have to perform further differentiations or other non-algebraic operations. Before the adoption of such measures in our example, incidentally, we could not proceed to find even $\mathbf{U}(3)$ without hitting the memory limit of 15 megabytes.

Distributed simplification. Suppose at the end of the notational simplification, a further algebraic simplification is to be performed on a huge collection of terms. Then there is still a chance that a machine's limit is reached before the simplification reaches its end. This happened when we tried to simplify the compatibility condition at m=3, which was expected to reduce to just 0. The basic idea of distributed simplification is to break down a huge expression into smaller sections at a cost of more computational time due to the possible overlappings.

The reasons are that when a machine tries to expand a complicated expression by inserting still more complex expressions so as to present it in a type of 'irreducible' form, the computer memory breaks down often long before the simplification reaches there. Often the increase on the demand of such as computer memory and/or time is exponential, hence looking for an improvement on hardware is often not sufficient, as the hardware quality improves only linearly in general.

To exemplify our proposed distributed simplification, let us look at the following trivial example. Suppose a *long* and complicated expression δ is a polynomial of $u(x)$, $v(x)$, $w(x)$ and $\partial_x w(x)$, whose coefficients are other function mixtures, and suppose a global simplification of the expression δ breaks down due to insufficient computer memory. Then we could instead calculate the coefficients of $u(x)$, $u(x)^2$, $u(x)v(x)$ and so on *separately*, so that the memory requirement is reduced to a minimum. However this type of fragmenting the calculation should not be overdone, as too much of it could increase the overall computational time so drastically that the overall calculation breaks down due to the over consumption of computing time! This is exactly where human interface is appreciated. Experience shows that one should break the calculation problem at a slow pace, as it takes usually much shorter time to hit, if any, the memory or other hardware limit.

In this example of Painlevé test, the method we used to cut down the memory consumption is simple: just verify *terms by terms*, selectively as explained above, rather than as a whole. It has slowed down the simplification considerably, but eventually achieved the task by overcoming the problem of hardware 'insufficiency'. We note that no matter how advanced a computer is, there will still be practical problems which make the hardware somewhat 'incapable'. Hence the above considerations will always be needed from time to time.

Nowadays symbolic algebras are becoming increasingly important in applied mathematics and mathematical physics. One of the reasons is that many results in such areas are nonrigorous, and many of them need to be validated in physical reality or via other means. Symbolic calculations go a long way in helping out with some of the validation tasks. The purpose of the rest of this section is to present a simple practical application and demonstrate how effectively computer algebras could come to our rescue.

To illustrate our point more concretely, consider the following example of two space dimensional NLS equations

$$iq_t = q_{xy} - 2q\partial_x^{-1}\partial_y(qr), \quad ir_t = -r_{xy} + 2r\partial_x^{-1}\partial_y(qr) \tag{3}$$

The NLS equations in one space dimensions model many physical phenomena including optical fiber transmission, and the explicit solitary traveling wave solutions (solitons) are of special interests to the industrialists. The NLS equations thus also received intensive mathematical analysis.

Under very stringent assumptions, some explicit solitons can be obtained explicitly by the *formal* use of the Inverse Scattering Transforms [JB]. For clarity, we give here a simplest solution

$$q(x,y,t) = -\frac{2i\eta(f(y-(\xi+i\eta)t))^* e^{-i\xi x}}{e^{\eta x} + \mid f(y-(\xi+i\eta)t) \mid^2 e^{-\eta x}}, \quad \xi,\eta \in R, \eta > 0 \tag{4}$$

where f is an arbitrary analytic function. Although one could in principle verify if this is indeed a solution of (3), the explosiveness on the length of the expanded expression makes it almost impossible for human to contemplate directly (not to mention other more complicated solution forms). This is where the power of symbolic calculation moves in again, this time as a means of proof or validation. For the above example, by using $f_y = \tilde{f}$, $f_t = -(\xi + i\eta)\tilde{f}$, $f_y^* = \hat{f}$ and $f_y^* = -(\xi - i\eta)\hat{f}$ due to the meromorphism of f, we could machine-verify directly that solution (4) satisfies (3). We have thus illustrated a simple case of validation, in which computer algebra *as well as* certain theoretical understandings are unified.

3. Intuition and validation by computer simulations

The vast machine power gives more wings to the imagination of the researchers: they could think faster and bigger, and let the machine 'judge' some of their primitive thoughts before they decide if they will pursue those primitive ideas further on. This helps to conduct more efficient research. The symbolic calculation in the previous section, though important, offers only one branch of direct application of computing technology. In this section, however, we shall consider another such branch — computer simulation. In an example of solitonic cellular automation to be given below, we shall show how computer simulation could verify, propose and induce many interesting properties, some of which may be proven rigorously. This is a typical case of computer power helping find a sensible research direction.

The cellular automaton (CA) that we are now using to exemplify our point is the time evolution of the following binary pattern

$$a^t = ..0..0a_0^t..a_i^t..a_L^t0..0.., \quad L < \infty, \quad a_i^t \in \{0,1\} \tag{5}$$

where $a_0^t = a_L^t = 1$. Their evolution was determined by [J]

$$a_i^{t+1} = \begin{cases} 0 & T_i = 0 \\ a_i^t & T_i \ \ odd \\ \overline{a_i^t} & T_i \neq 0 \ \ even \end{cases}, \qquad T_i = T_i(a^t) \equiv \sum_{j=1}^{r} a_{i-j}^{t+1} + \sum_{j=0}^{r+1} a_{i+j}^t, \tag{6}$$

where $\bar{0}$=1, $\bar{1}$=0, $r > 2$ is an integer and $a_n^{t+1} = 0$ is assumed for n sufficiently far to the left. CAs are not only mathematically interesting, but also often closely related to world of electronics. Before going further, we note that clusters of nonzero bits separated by at least $(r+1)$ zero bits are termed particles. Thus the evolution of the bit patterns becomes the evolution of particles and the particle collisions. The theoretical purpose is to find out if and when the collisions are solitonic, i.e. particles re-emerge after collision, and what quantity (energy) is conserved during the evolution. Let us now look at this example and see how computer technology influenced the course of the research of this model.

First we recall [J] the followings without going into details. (i). The simulation first shows that no non-zero particles disappear at any time, implying general conservation of some quantity. This quantity of energy is *then* theoretically shown to be $E(a^t) = \sum_{i=-\infty}^{\infty} \mid a_i^t - a_{i-r-1}^t \mid$. (ii). Simple particle collisions are solitonic, as can be seen from simulation results, conforming to the theoretical results. (iii). In the exhaustive simulation of non-simple particle collisions under a given pattern width, there are always some non-solitonic exceptions. Hence there is no need to contemplate a proof of solitonic collision among *all* types of non-simple particles. (iv). There is an empirical phenomena

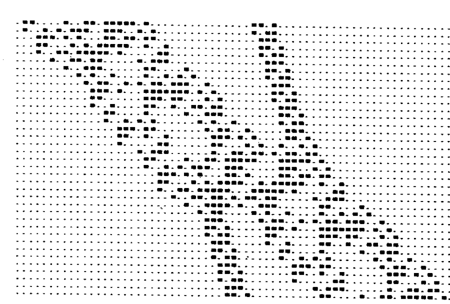


Fig 1: particles re-emerge after collision with left-moving frame speed=r=3

that if a collision is solitonic then, the exact pattern will reappear at the same time. This holds for all simulated collisions, see Fig 1 for example, but is still beyond a rigorous proof so far. (v). Finally a strong belief, due to the observation of other similar CAs, that a slow particle can not be shifted backward after a collision, was shown to be false due to the very rare cases picked up by the exhaustive collision simulations.

These observations turned out to be the backbone in the study the CA model. They helped verify existing results, remove wrong conjectures and derive or prove the related rich properties. To start with, computer simulations can not only verify some existing propositions, but also provide often correct insights [J] into a research subject before enormous time is invested by the researchers. For instance, while the first observation (i) prompted one to define a proper energy, (ii) is the theoretical result of another similar model and the simulation results provided enough confidence to the pursuit of its rigorous proof in the end. The simulation result (iv) also posed a theoretical challenge for its proof — an direct example of an conjecture induced by the computer power. Finally, a particle shift phenomenon (v) after collisions that was strenuously observed [PST] empirically in a similar but different model by PST, is successfully eradicated for the new model (6) by the counter examples (see e.g. Fig 2) picked up by the simulation program automatically.

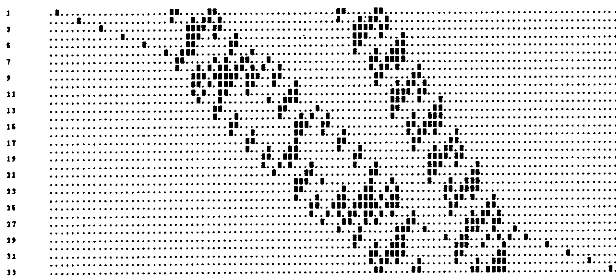


Fig 2: rare occurrence of backward shift after collision with left-moving speed=r=4

We note again that if a computation task fails due to the apparent inferiority of computing hardware, then the solution is most likely to come from improved algorithm rather than upgrading the hardware. In the study of our CA (6), for example, the two particle collision simulation and statistics were first written and run on a 486 PC, and then re-run on a UNIX work station. We discovered, agreeing with our point here, that the maximum particle seeds width r with which UNIX can handle without reaching its hardware limit does not significantly improve on a good PC configuration.

4. Needs analysis through the use of online help system

In this section, we shall be considering another type of research, the research on student needs and course-teaching self-evaluation. There are a variety of ways to conduct student evaluation, see e.g. [BB,W,WW], including the so-called *action research* [W]. The main objective here, however, is to devise a simple online help retrieval system that also logs the user sessions for later analysis.

First let us explain briefly a typical implementation. Suppose a student in a computer session class needs help on some specific subjects. Then he could immediately shell out and call a online help dispatch program. The dispatcher will start clocking the time immediately. Every time the student changes a minor subject, the help dispatcher records the new subject and the total time spent on the previous subject. When the student finally leaves the online help, the local log file, containing the solicited help subjects and the corresponding times used, is appended to the (remote) master log file. The online helps may be organized in many different ways and different forms, though we find that presenting notes in graphic GIF files serves almost all purposes. This type of implementation, though not as fancy in looks as those dedicated window based help systems, requires only a *minimum* from the hardware or parent software, and can be used on almost any existing packages without running into problems such as memory shortage. In other words, such systems are highly portable to different application packages.

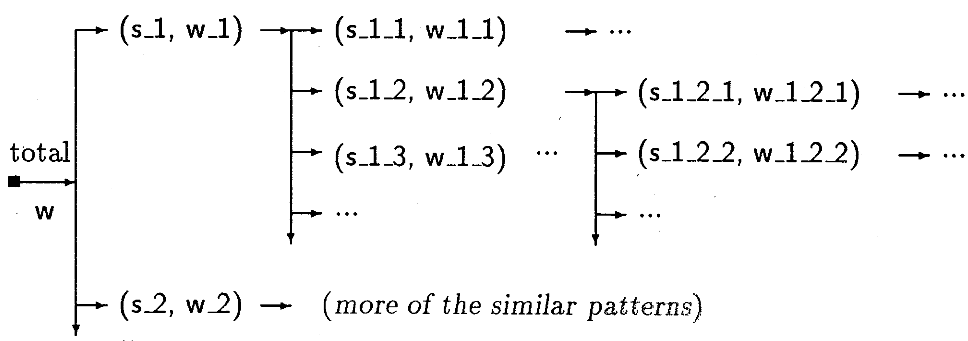
Let us now concentrate on analyzing the data of the help retrieval log. Typically the log file will each line contain essentially one help request entry in the form of

help_subject	*time_used*	*time_started*

where the first field *help_subject* could also be just a page number of the main help file, the second field *time_used* is the total time a user spent on reading the helps concerning the particular subject, and the last (optional) field *time_started* contains the date and time at which the help on the subject is requested. Our purpose is thus to use the above log data to establish a needs chart, and use it determine the new time allocation on each individual topic. Our proposed strategy is as follows.

We first introduce a weight function to measure the amount of time for each help retrieval. The weight function $w(t)$ is a nonlinear function of (help-retrieval duration) time. Currently the weight function w is still very much to be experimented. However, we propose to choose $w(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx$, with a *normal distribution* integrant, as the weight function. This is because we believe that the initial help time should gain highest weight rank, while each help session should have a weight upper bound. Of course, better weight functions will have to come from more practical experiences. Moreover, we for convenience assume that the weight function $w(t)$ is date invariant in the sense that it is independent of the third field *time_started*.

By adding up all the weights of each help retrieval and sorting out the hierarchy list, we obtain a student need weight table in typically the following form



(7)

where, for instance, s_1_2 is a subject name whose help is requested for a duration of time whose weight is w_1_2. The above is thus a distribution table of student's need for extra helps. We can also make a similar table, with the time replacing the weight, of the current allocation of teaching time on each of the subject. Hence our following task is to provide a recommendation for the increase or decrease on the time allocated to each teaching subject.

First we need to determine how much help weight is the average, so that a subject needs more lecture time or attention if its *difficulty ratio* is above the average. Because the total lecturing hours are fixed, one could either take out some time from the subjects that have difficulty ratio below the average, or drop some subjects from the syllabus. Suppose a maximum of 25% of change is allowed on any particular subject, then a new time allocation table can be calculated immediately by the following method

$$
\begin{aligned}
C &= O + (W - RO)/R\,, \\
N &= \begin{cases} 1.25O & \text{if } C > 1.25O \\ 0.75O & \text{if } C < 0.75O \\ C & \text{if otherwise} \end{cases}\,, \\
T &= \sum_{\text{all subjects}} N_{\text{subject}}
\end{aligned}
\tag{8}
$$

where, for each given subject, O is old allocated time, $R{=}O/total\_weight$ is the average difficulty ratio, W is the weight of the help time requested for the subject, C is the crude recommendation for the new time allocation induced by the help request distribution table, and N is the final recommended new time allocation for the subject. We note that a limit of such as 25% on the change of time allocation may not be necessary in practice, and in such circumstances the crude recommendation C will be adequate enough.

If the estimated new total lecturing time T is significantly different from that of the targeted one, we could add a few new subjects or drop several existing ones. If all subjects are to be retained, then an iteration would have to be performed until the final estimated number of hours is close enough to the fixed total allocation. However, we shall not elaborate further on this particular situation.

We note that the above procedure can be applied to each main or sub branch of a course or a subject. We could also absorb proper minor subjects into relatively major titles so as to facilitate the calculations. These strategies can enable one to determine either macro or micro reallocation of time on each subject unit.

To illustrate the need of such subject absorption in the analysis, we consider for instance the following set of subjects in a course of numerical analysis.

- a branch
  - Integration
    - Trapezoidal
    - Simpson
    - Gaussian
    - Error estimate
  - Differential equations
    - Euler
    - Taylor
    - Rung-Kutta
    - Multi-step
      - AM method
      - AB method
    - Error analysis

Since the time allocated to teach some subjects is likely to be quantum, proper absorption of minor subjects will be needed to enable each entry of the final table to have a multiple of a fixed amount of time such as an hour. For example, in an hour's lecture of multi-step method, we might decide to cover together two topics, the *AM method* and the *AB method*. This means that the branching of *multi-step* in the above table will be removed. Similarly we might need also to absorb in the table the *Trapezoidal method* and the *Simpson method* into a single teaching unit. Hence it will be more appropriate if one shrinks the hierarchical subject (time allocation) table so as to reduce the effect of quantum allocation of time.

To summarize, we in this section have proposed a simple algorithm that will help determine students needs and adjust the course materials for the future offering. This approach of feedbacks via online help should be very useful, particularly for courses that require computer practicals. While lecturers will have to eventually 'adjudicate' the information provided by such a mechanism, this approach gives a powerful indicator of what particular subjects that will have to be addressed more thoroughly. Moreover, such a third party teaching evaluation is often most objective and confidential, and is therefore a great companion.

5. Conclusion

In this work, we addressed a number of the roles played by the use of modern computer technology in the academic research, as well as in the research of students needs and teaching self-evaluation. We examined the applicability of symbolic calculation and computer simulation, with stress on the possible pitfalls and the recommended solution measures. Finally we proposed a working method based on the retrieval of online help during students practical sessions, to determine how a same or similar course could be improved when it is offered again. All these features would be unachievable without the ever sophisticated modern computer technology. And once again, it proves that modern technology can help a great deal, even in the oldest subject mathematics.

References

[BB] Balla J and Boyle P, *Evaluation of student performance: a framework for improving practice*, Asses. & Eval. in Higher Educ. **19** (1994)17.

[BJ] Bullough R K and Jiang Z, *A note on some features of the Painlevé analysis*, Physica Scripta, In press 1995.

[CM] Clarkson P and Mansfield E L, *Symmetry reduction and exact solutions of shallow-water wave equations*, preprint, University of Exeter, 1994.

[J] Jiang Z, *An energy- conserved solitonic cellular automaton*, J. Phys. A **25** (1992)3369.

[JB] Jiang Z and Bullough R K, *Integrability and a new breed of solitons of an NLS type equation in 2+1 dimensions*, Phys. Lett. A **190** (1994)249.

[JH] Jiang Z and He X, *NLS type equations in 2+1 dimensions: new type of solutions and non-isospectral problems*, Theor. Math. Phys. **99**, (1994)382.

[JS] Johnston S, *Choosing mathematics, 'you need it even if you don't want to do it'*, Austr. J. Educ. **38** (1994)233.

[MF] Mansfield E L and Fackerell E D, *Differential Gröbner bases*, Macquarie University, **92/108**, 1992.

[MT] Mitic P and Thomas P G, *Pitfalls and limitations of computer algebra*, Compu. & Educ. **22** (1994)355.

[PST] Park J K, Steiglitz K and Thurston W P, *Soliton-like behaviour in automata*, Physica D **19** (1986)423.

[W] Wolff D, *Computers in classroom research*, Compu. & Educ. **23** (1994)133.

[WW] Wade B D and Walker C, *Assessing the effect of various academic parameters on the academic performance of honors students at Southern University—BR*, Education **115** (1994)63.

TEACHING A LABORATORY BASED LINEAR ALGEBRA COURSE

Thomas W. Judson
University of Portland
5000 N. Willamette Blvd
Portland, Oregon 97203
Email: judson@hood.uofport.edu

With the increased availability of inexpensive computing and the introduction of sophisticated software such as MATLAB^{TM}, applications involving linear algebra have become more important during the past decade. Computations that once required a mainframe can now be easily accomplished with a relatively inexpensive desktop computer or even a handheld calculator. For this reason, linear algebra has become one of the most important mathematical tools in the sciences, engineering, economics, and computing. The introduction of technology in the teaching of undergraduate mathematics makes the subject of linear algebra one of the most interesting in the current reform movement in mathematics education. The ability to solve large systems of linear equations, compute powers of stochastic matrices, find eigenvalues, and realize the importance of numerical analysis when applying mathematics to real world problems makes linear algebra an ideal subject to be taught as a laboratory course.

Linear Algebra and the Educational Reform in Mathematics

When compared to calculus and differential equations, the study of linear algebra is a relatively recent development. It was James Joseph Sylvester who first used the term "matrix" to denote an $m \times n$ array of numbers. Together with Arther Cayley, he laid the foundations of modern matrix algebra. Linear algebra also developed along abstract lines under the influence of mathematicians such as Emmy Noether. In modern linear algebra courses, we see the influence of both matrix algebra and abstract vector spaces.

In the 1940s, linear algebra began to find its way as a course in graduate school. When undergraduate texts began to appear in the 1950s, linear algebra soon became a standard undergraduate course that was taught at the junior or senior level; however, for the next twenty or thirty years linear algebra was taught primarily to mathematics majors. During this era, a course in linear algebra was either viewed as necessary preparation for abstract algebra or served as a sequel course to multivariate calculus that would provide the necessary background for doing calculus on $\mathbf{R}^n$ in the framework of modern analysis. The role of linear algebra has changed in the last ten or fifteen years. At the present time the course is commonly taught as the first course after calculus or multivariate calculus. Many students taking linear algebra come from disciplines other than mathematics. Students have become more interested in the applications of linear algebra.

The reform movement in undergraduate mathematics education can trace its roots to the "Lean and Lively Calculus" conference at Tulane University in January, 1986. There was a perceived need that the calculus curriculum for the previous ten years had not been meeting the needs of the students and that the mathematical experience that students received from calculus was too limited. Also during the 1980s, inexpensive technology was began to make its first appearance. These two events have driven calculus reform to its present state. While educational reform in mathematics has not met with universal acceptance and there is no widespread agreement on the exact directions that education reform in calculus should take, there is general agreement that reform is here to stay in some form.

The ATLAST Project and the LACSG Recommendations

It is natural that the reform movement in mathematics would spread beyond calculus. Since matrix algebra is one of the most computationally intense areas in all of mathematics, the need to exploit technology has led to educational reform in linear algebra. Two of the most important developments in way that linear algebra is now being taught in the United States are the ATLAST project and the recommendations of the Linear Algebra Curriculum Study Group (LACSG). These groups, sponsored by the National Science Foundation and aided by inexpensive technology, have set the foundation for the reform movement in the linear algebra curriculum.

In 1990, The LACSG met at Williamsburg College to examine the current practices that prevailed in teaching linear algebra. The group presented their initial findings at the annual meeting of the American Mathematical Society and the Mathematical Association of America in San Francisco in January,

1991. Since this initial announcement, the LACSG recommendations have undergone modification after receiving a tremendous amount of input from the mathematical community. Some of the current LACSG recommendations are listed below.

- A first course in linear algebra must respond to the needs of client disciplines such as science, engineering, economics, and computing.
- Serious consideration should be given to making the first course in linear algebra a matrix based course. This does not mean that the theory of linear algebra should be neglected, but that it should be taught in terms of matrices whenever possible and not in terms of abstract vector spaces.
- Faculty should consider the needs and interests of students as learners.
- The utilization of technology should be encouraged in the first linear algebra course.
- A more advance course in linear algebra and matrix theory should receive high priority in every mathematics curriculum.

The LACSG has presented a core curriculum centered around these recommendations [**C2**].

The ATLAST project was cosponsored by the International Linear Algebra Society and the National Science Foundation. The goal of the project was to **A**ugment the **T**eaching of **L**inear **A**lgebra through the use of **S**oftware **T**ools. Part of the project was centered around multiday workshops, where were participants learned to use MATLAB™ to construct laboratory based exercises to incorporate technology into linear algebra courses. People who attended the ATLAST workshops designed computer exercises that were class tested and then incorporated into a database. A selection of these exercises was to be published in the public domain.

Teaching Linear Algebra using Laboratory Exercises

Until recently most instructors took an axiomatic approach to teaching linear algebra. Lectures most often consisted of proofs, definitions, and examples.

Very few applied examples were included, since instructors were either unfamiliar with them or such examples were not readily available in the literature. Until a few years ago, working examples in the classroom of a nature that one might actually encounter in a real life application or assigning such problems as homework was an impracticality due to prohibitive computations.

The introduction of sophisticated calculators and computer software such as MATLAB™, DERIVE™, Mathematica™, and Maple™ has revolutionized the teaching of undergraduate mathematics during the last few years. Many mathematics courses are moving away from more traditional formats at the lower levels towards laboratory based courses. With the introduction computer classrooms and laboratories, it is now possible for both instructor and student to do extended computations quickly and easily. Certainly, the calculations that one is required to perform in a typical linear algebra course are equal to or surpass computations that are required in other courses. Computer exercises allow the study of interesting and realistic problems. Tedious computations can be avoided allowing the students to focus on concepts. Properly guided, students can develop their mathematical intuition and make new discoveries via experimentation. They can investigate topics in laboratories that are not covered in the lectures.

Solving Linear Systems and Finding Matrix Inverses

Using software such as MATLAB™, students can examine the efficiency as well as the accuracy of the algorithms that are used in linear algebra. For example, the difference between solving systems of linear equations using Gaussian elimination and Gauss-Jordan elimination is easy to see using MATLAB™, since one can count the number of floating point operations needed to do a specific matrix computation. Students quickly convince themselves that Gaussian elimination is the fastest way to solve a linear system.

Students can also judge the accuracy of a computation. For example, if A is an $n \times n$ matrix, the MATLAB™ command `inv(A)` will return A^{-1}. If A is singular or "very close to being singular," MATLAB™ will print a warning message. If we try to calculate the inverse of the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix},$$

the following message is returned.

```
Warning:  Matrix is close to singular or badly scaled.
          Results may be inaccurate.  RCOND = 2.055969e-18

ans =

       1.0e+16 *
        -0.4504    0.9007   -0.4504
         0.9007   -1.8014    0.9007
        -0.4504    0.9007   -0.4504
```

Of course, the matrix A is singular. However, we do not have to change A by very much to get MATLAB™ to respond with a good message. If we ask MATLAB™ to compute the inverse of

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9.0001 \end{pmatrix},$$

then the reply is

```
ans =

       1.0e+04 *

         0.9998   -1.9999    1.0000
        -1.9999    4.0000   -2.0000
         1.0000   -2.0000    1.0000
```

The matrix B is not nearly as bad as A. However, since B is only slightly different that A, the inverse of B might not be as accurate as we think. Every nonsingular matrix has a number associated with it called the *condition number.* The condition number of a matrix measures how sensitive the solution of a corresponding system of linear equations is to errors in data. The condition number of an $n \times n$ matrix A is defined by $\text{cond}(A) = \|A\|\|A^{-1}\|$, where

$$\|A\| = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}.$$

Although the actual definition of condition numbers requires a knowledge of matrix norms, students can use condition numbers as a "black box" quite early in a first course in linear algebra to determine the accuracy of numerical solutions. As a general rule, if the condition number of a matrix is approximately 10^k, then the last k digits of A^{-1} cannot be trusted. The condition number of the matrix B given by MATLAB™ is `1.0109e+006`. This means that the last six digits of the inverse of B are questionable if we display B^{-1} to the full 16 digit precision available in MATLAB™.

Stochastic Matrices

In another laboratory assignment students were asked to investigate stochastic matrices. Given the matrix

$$A = \begin{pmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{pmatrix},$$

with MATLAB™ students can quickly and easily convince themselves that

$$A^n \to \begin{pmatrix} 0.4 & 0.4 \\ 0.6 & 0.6 \end{pmatrix}$$

as $n \to \infty$. Students could also see that $A^n\mathbf{x}$ has the same limit as $n \to \infty$ whether $\mathbf{x}$ is $(1000\ 500)^T$ or $(750\ 750)^T$. After doing several such experiments, students can be directed towards answering more theoretical questions.

Spanning Sets, Linear Independence, and Bases

Students always seem to have difficulty with the concepts of spanning sets, linear independence and bases. Although MATLAB™ cannot fully answer questions about linear independence, in special cases it does quite well. Given a set of vectors $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ in $\mathbf{R}^n$, it is difficult for students to decide if S is a linearly independent set or if S is a spanning set for $\mathbf{R}^n$. Ultimately, this is answered by solving a system of linear equations. If students are forced to solve complicated systems by hand, many of them become quickly discouraged and are unwilling to do all of the required calculations. Further laboratories can explore the the idea of rank of a matrix A. Students can be lead to discover the relationalship between the fundamental subspaces that are associated with an $m \times n$ matrix A. That is, they can discover the fact that

$$\operatorname{rank}(A) + \dim N(A) = n.$$

Eigenvalues and Differential Equations

Laboratory exercises can be used to explore topics that might not be covered during lectures. A laboratory exercise on systems of linear differential equations illustrate the importance of eigenvalues. From calculus students know that the functions that satisfy the differential equation $x'(t) = ax(t)$ are those of the form $x(t) = ce^{at}$. In a well written laboratory exercise, we can demonstrate the theory behind finding solutions to systems of first order linear differential equations of the type

$$\begin{aligned}
x_1'(t) &= a_{11}x_1(t) + \cdots + a_{1n}x_n(t) \\
x_2'(t) &= a_{21}x_1(t) + \cdots + a_{2n}x_n(t) \\
&\vdots \\
x_n'(t) &= a_{n1}x_1(t) + \cdots + a_{nn}x_n(t),
\end{aligned}$$

where the a_{ij}'s are scalars. It is not difficult to get students to believe that the solutions involve exponential functions. We show that solving the system $X'(t) = AX(t)$ involves finding the eigenvalues of A. Of course, in the examples and exercises, it is necessary to choose the system with some care. Systems like

$$\begin{aligned}
x_1'(t) &= 2x_1(t) - 3x_2(t) \\
x_2'(t) &= 4x_1(t) - 5x_2(t)
\end{aligned}$$

work well, since

$$A = \begin{pmatrix} 2 & -3 \\ 4 & -5 \end{pmatrix}$$

has distinct real eigenvalues. Since MATLAB™ can easily calculate eigenvalues for large systems, interesting applications of systems of linear applications can be incorporated into the laboratory.

The Singular Value Decomposition

One of the most important results from linear algebra is the singular value decomposition. For any $m \times n$ matrix A, there exists an $m \times m$ orthogonal matrix U, an $n \times n$ orthogonal matrix V, and an $m \times n$ matrix

$$\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix},$$

where D is the $r \times r$ diagonal matrix

$$D = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{pmatrix},$$

such that

$$A = U\Sigma V^T.$$

The values $\sigma_1, \sigma_2, \ldots, \sigma_r$ are the singular values of A. The singular values have the property that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

The singular value decomposition has very nice applications to digital imaging. Any image such as a photograph can be digitized and stored as a matrix. Each entry in the matrix corresponds to a pixel in the image. The various numerical values of the entry tell what the color of each pixel should be. Generally, the color of neighboring pixels are very close to the color of a given pixel.

If we are clever, we can reduce the storage that is required for an image. Suppose that we digitize a photograph a store it in a matrix A. If $U\Sigma V$ is the singular value decompositon of A, then we can write A as

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_n \mathbf{u}_n \mathbf{v}_n^T,$$

where $\sigma_1, \sigma_2, \ldots, \sigma_n$ are the singular values of A. If we truncate this sum after k terms, then

$$A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$$

gives a good approximation of the original photograph. Often, only 80 percent of the storage required to store the original matrix is needed.

Since MATLAB™ not only has the ability to find the singular value decomposition of a large matrix but is also capable of excellent graphics, it is the ideal vehicle for doing a laboratory that applies the singular value decompostion to digital imaging.

Where Do We Go from Here?

There are many difficulties in implementing a laboratory based linear algebra course. First of all, new software must be learned. While MATLAB™ is an excellent numerical linear algebra program, it lacks the ability to do symbolic computation.[1] Computing facilities and computer classrooms are expensive to maintain and are not readily available at many universities, although sophisticated calculators such as the HP-48G and TI-85 may help solve this problem. Often times, there is some resistance to computing from students (and faculty). If open-ended laboratory assignments are required, even more resistance can be expected. Finally, there are not a lot of computer exercises available for linear algebra. Developing good laboratory exercises for linear algebra is difficult and time consuming.

If linear algebra curriculum reform is to be successful, many problems need to be addressed including some of the following.

- Faculty must be persuaded that linear algebra curriculum reform is worthwhile.
- More laboratory exercises must be developed.
- More texts that integrate technology must be written.
- Courses must remain courses in mathematics and not become button pushing courses. The theory of linear algebra is still very important.
- The Internet and the World Wide Web must be exploited.

In conclusion, students are often critical of the need for linear algebra especially when a linear algebra course is taught in an abstract fashion with few meaningful applications. A matrix oriented laboratory approach to teaching linear algebra can successfully motivate students to learn linear algebra.

[1] Mathworks has introduced a Maple™ Toolbox, which should help solve this problem.

References

[C1] D. Carlson, Teaching Linear Algebra: Must the Fog Always Roll In? *College Mathematics Journal*, 1 (1993) 29–40.

[C2] D. Carlson, et. al., The Linear Algebra Curriculum Study Group Recommendations for the First Course in Linear Algebra, *College Mathematics Journal*, 1 (1993) 41–46.

[J] T. Judson, *Introduction to Linear Algebra*, PWS Publishing Company, Boston, (In preparation).

[L] S. Leon, *ATLAST: Augmenting the Teaching of Linear Algebra through the use of Software Tools.* Notes from the June 17–19, 1993, Michigan State University ATLAST Workshop.

[S] G. Strang, *Linear Algebra and Its Applications*, third edition, Harcourt Brace Jovanvich, Orlando, Florida, 1988.

SYMBOLIC COMPUTING IN PROBABILITY AND STATISTICS [1]

Zaven A. Karian
Department of Mathematics and Computer Science
Denison University
Granville, Ohio 43023 U.S.A.
Karian@Denison.edu

Statisticians were among the first mathematical scientists to use computers in their teaching and in various statistical analyses. Initially, when computation was confined to numeric calculations, it was natural to use computers to calculate sample statistics. With the advent of methods for the generation of random samples from a variety of distributions, computer-based statistical simulations developed into a powerful technique with applications to industry and education (see, for example, [2]). More recently, sophisticated graphic methods have allowed us to observe patterns by visualizing attributes of theoretical distributions and empirical data. In short, we have come a long way.

This paper deals with the use of symbolic computation—yet another type of computation—in probability and statistics. The manipulation of mathematical symbols, something that is done regularly in probability and statistics, somehow has not yet found its way into standard statistical packages such as MINITAB and SPSS. Moreover, the statistics packages that are bundled with symbolic computing systems such as *Mathematica* and *Maple* are very limited in scope and provide only small fraction of the functionality that is necessary for statistical analyses. There seem to be three, somewhat obvious, remedies to this situation:

- add symbolic computational features to existing statistical packages;
- make the statistical subsystems of symbolic computing systems much more comprehensive;
- merge a statistics package with a symbolic computing system in such a way that users have complete access to all symbolic and statistical features and results from one of these subsystems can be communicated to the other in a reasonably transparent way.

Given the effort and costs associated with the development of a comprehensive symbolic or statistical software package, the third alternative listed above seems to be the most reasonable one in the short run.

[1]Supported by grants from the Fund for the Improvement of PostSecondary Education (FIPSE grant P116B30079) and the W. M. Keck Foundation.

This paper will focus on the advantages of using symbolic computation in teaching probability and statistics. The examples that follow use the *Maple* symbolic system with a supplement of about 130 statistical procedures. (A detailed description of the supplement and the manner in which it can be used in probability and statistics courses is given in [3].) A general narration is given for each example; the details of the *Maple* interaction, including some graphic output, are given in the appendices.

1. Simple Simulations

Almost all statistical packages (e.g., MINITAB or SPSS) have built-in features that allow users to extract random samples from well-known distributions. What about sampling from arbitrary discrete or continuous distributions? If the statistical procedures were embedded within a symbolic computing environment, as is the case here, we could easily pass an expression representing a probability density function (p.d.f.) as a parameter to a procedure and have the procedure return a random sample from the specified distribution.

The interaction given in Appendix A illustrates different ways of generating random samples. First, the statistical routines are read in; next, `A := Die(6,8);` is used to simulate 8 repeated rolls of an 6-sided die followed by `Freq` which gives the frequencies of $1, 2, \ldots, 6$ in the sample. `DiscreteS([0, 1/2, 1, 1/2], 8);` simulates 8 flips of a coin by randomly sampling from the distribution of a random variable that assumes values 0 and 1, each with probability 1/2.

The two commands `pdf := x/10;` and `Sample := DiscreteS(pdf, 1..4, 5);` define and then generate a random sample from the distribution with p.d.f. $x/10$, $x = 1, 2, 3, 4$. The last part of Appendix A generates a larger sample (of size 300) from this distribution and gives its histogram (dashed lines) with a superimposed probability histogram of the distribution.

2. Relationships Between Distributions

In almost any undergraduate probability or statistics course students are taught about the circumstances under which $B(n,p)$, the binomial distribution with parameters n and p, can be approximated by the normal distribution, $N(\mu, \sigma^2)$ where $\mu = np$ and $\sigma^2 = np(1-p)$. One way of illustrating this relationship is to look at the graphs of various binomial p.d.f.s with the appropriate normal distribution superimposed and see the quality of the approximation for different values of n and p.

The *Maple* interaction given in Appendix B first animates the probability histograms of $B(n,p)$ for $p = 0.25$ and $n = 2, 4, \ldots, 32$. In the second loop, the p.d.f. of $N(\mu, \sigma^2)$ is superimposed on the binomial probability histograms. Since animation cannot be easily shown on a printed page, four of the 16 frames of the animation corresponding to $n = 4$, 8, 12, and 16 are shown. By a simple modification of what is shown in Appendix

B one could fix n and produce an animation based on changing values of p (see the paper by Karian in [1] for another approach).

3. Understanding an Important Theorem

The Central Limit Theorem (CLT) is the most important result that students encounter in introductory probability and statistics courses. Generally, illustrations (not proofs) of the CLT consist of taking sums or averages of random samples from some distribution and showing that the sums or averages indeed tend to normality as the sample size increases. Because of the constraints of the commonly used statistical packages, the distributions used in such illustrations tend to be "bell-shaped." Thus, making it difficult to underscore the fact that the tendency to normality prevails even in pathological, say "U-shaped" distributions, provided the distribution has a finite mean and a finite variance.

In the *Maple* interaction of Appendix C, the U-shaped continuous p.d.f.

$$f(x) = (3/2)x^2, \quad -1 \leq x \leq 1$$

is used. First, the mean and variance are determined and then, through `ContinuousS` random samples (first of size 3 and then of size 400) are produced. The first graph shows that there is a reasonably good fit between the sample and the distribution. The statement following the graph produces 100 samples, each of size 4, from this distribution (the first of these, `A[1]`, is shown). Next, the means of these 100 samples are calculated and eventually a histogram of the means is plotted together with the approximating normal and (for emphasis) the original distribution. It is clear that for samples of size 4, the assertion of approximate normality is not easy to justify. However, experimenting with different sample sizes, such as size 16 (shown next) rectifies this situation.

4. The Power Function of a Statistical Test

In statistical testing situations it is often difficult to compute the probability of type II error, β. Even though it is quite valuable to know β for various simple alternate hypotheses, this is not considered in most courses because of the sheer computational difficulties.

Appendix D shows how β or equivalently the power function of a test can be examined by the use of symbolic computation. The specific illustration considers the null hypothesis $H_0 : \sigma^2 = 100$ against the two-sided alternative $H_1 : \sigma^2 \neq 100$ for a normally distributed random variable. First the sample size, n, is set to 21 and the probability of type I error, α, is set to 0.05. For this choice of n and α, `L` and `U`, the lower and upper bounds of the critical region are computed. The percentiles of the χ^2 distribution needed for this are obtained through `ChisquareP(n-1, alpha/2);` and `ChisquareP(n-1, 1-alpha/2);`.

The power function for this test,

$$K(\sigma^2) = 1 - \beta = \int_L^U \frac{x^{(n-1)/2-1}e^{-x/2}}{\Gamma((n-1)/2)2^{(n-1)/2}}\,dx$$

$$\approx \int_{959.0777392/\sigma^2}^{3416.960690/\sigma^2} \frac{x^9 e^{-x/2}}{9!\,2^{10}}\,dx,$$

is computed through the use of ChisquareCDF which uses numeric integration to evaluate cumulative distributions. What a symbolic mess!

The statements `P1:=plot(K, sigma=0..20):` and `P1;` give a graph of the power function—something very difficult to do with ordinary statistical packages. But wait, we can do much more! How does $K(\sigma^2)$ change as we vary α? We choose additional cases with $\alpha = 0.005$, 0.01, 0.025, 0.1, and 0.2 and animate (or display simultaneously) all 6 power functions. This tells us how β is influenced by α. We can also consider the impact of n on β. The last graph shows a collection of power functions with α fixed at 0.05 and $n = 5, 9, 13, 17, 21,$ and 25.

References

[1] Karian, Z. A. (Ed.), *Symbolic Computation in Undergraduate Mathematics Instruction*, MAA Notes Number 24, The Mathematical Association of America, Washington, D.C., 1992.

[2] Karian, Z. A. and Dudewicz, E. J., *Modern Statistical, Systems and GPSS Simulation*, W. H. Freeman, 1991.

[3] Karian, Z. A. and Tanis, E. A., *Probability and Statistics Explorations with Maple*, Prentice-Hall Inc., Englewood Cliffs, 1995.

Appendix A: Simple Simulations

```
> read(`a:stat.m`):

> A := Die(6, 8);
```

$$A := [4, 3, 4, 6, 5, 3, 6, 3]$$

```
> Freq(A, 1..6);
```

$$[0, 0, 3, 2, 1, 2]$$

```
> A := Die(4, 400):

> Freq(A, 1..4);
```

$$[101, 107, 87, 105]$$

```
> Coins := DiscreteS([0, 1/2, 1, 1/2], 8);
```

$$Coins := [1, 1, 0, 1, 1, 0, 0, 1]$$

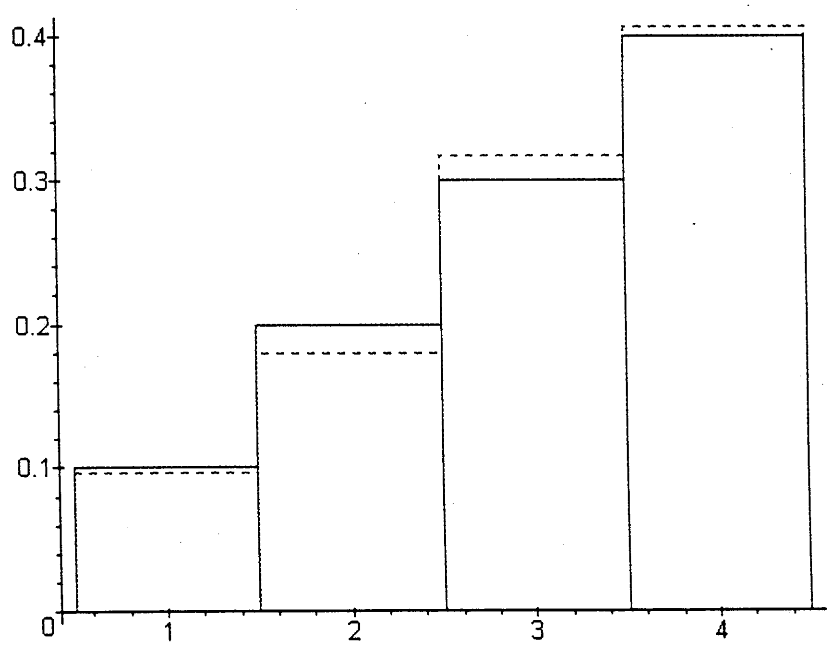
```
> pdf := x/10;
```

$$pdf := \frac{1}{10}x$$

```
> Sample := DiscreteS(pdf, 1..4, 5);
```

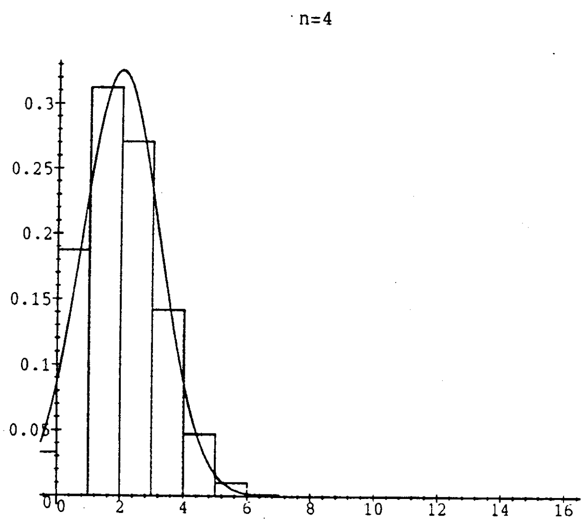
$$Sample := [3, 4, 1, 1, 1]$$

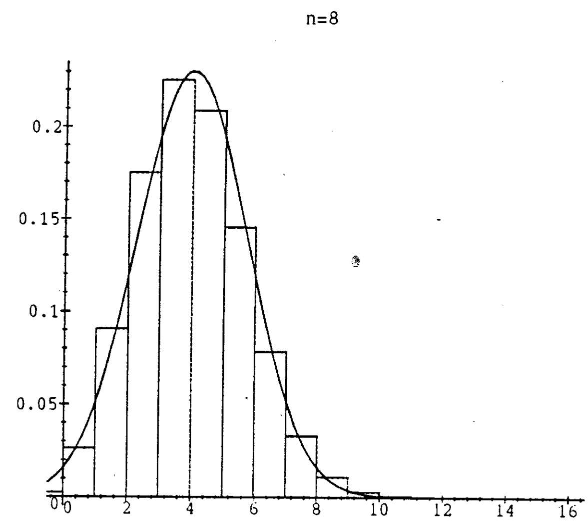
```
> Sample := DiscreteS(pdf, 1..4, 300):
> EH := Histogram(Sample, 0.5..4.5, 4):
> TH := ProbHist(pdf, 1..4):
> plot({EH,TH});
```

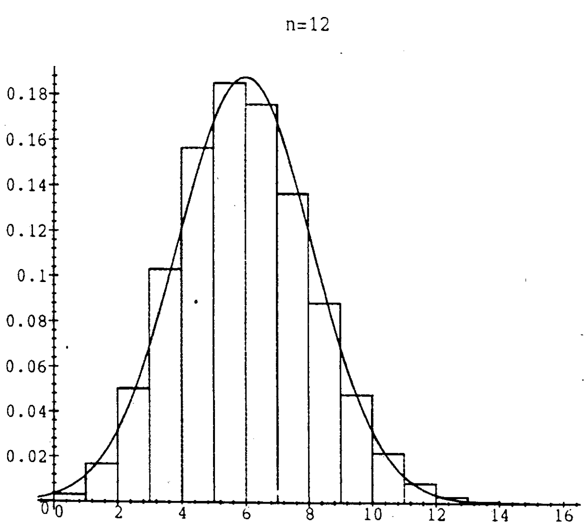


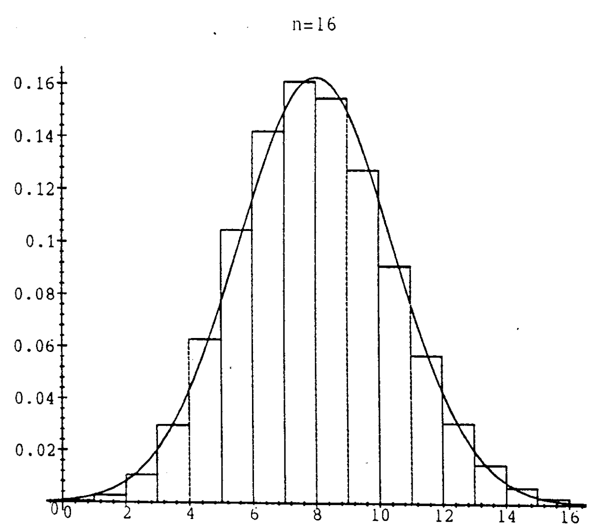
Appendix B: Approximating the Binomial

```
> read(`a:stat.m`): with(plots, animate, display):
> for i from 1 to 16 do
>     H.i := ProbHist(BinomialPDF(2*i, 0.25, x),  -0.5..16.5, 17):
> od:
> display([seq(H.i,  i = 1..16)], insequence=true);

> for i from 1 to 16 do
>     H := ProbHist(BinomialPDF(2*i, 0.25, x), -0.5..16.5, 17):
>     N := plot(NormalPDF(i/2, 3*i/8, x), x = -0.5..16.5):
>     P.i := plot( {N,H} ):
> od:
> display([seq(P.i, i = 1..16)], insequence=true);
```









Appendix C: Central Limit Theorem

```
> restart; read(`a:stat.m`);

> f := x -> (3/2)*x^2;
```

$$f := x \to \frac{3}{2}x^2$$

```
> int(f(x), x=-1..1);
```

$$1$$

```
> mu := int(x*f(x), x=-1..1);
```

$$\mu := 0$$

```
> var := int(x^2*f(x), x=-1..1) - mu^2;
```

$$var := \frac{3}{5}$$

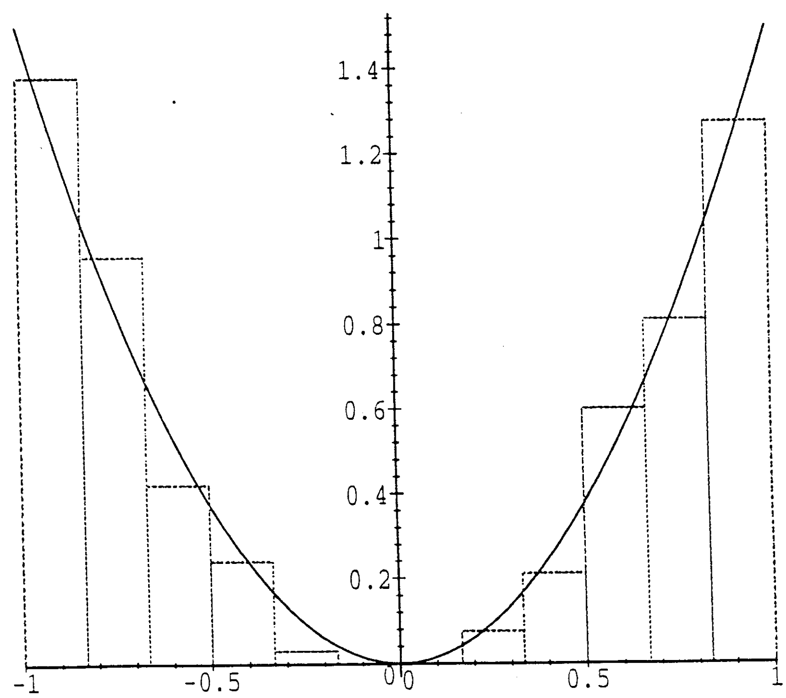
```
> F := int(f(t), t=-1..x);
```

$$F := \frac{1}{2}x^3 + \frac{1}{2}$$

```
> A := ContinuousS(f(x), -1..1, 3);
```

$$A := [-.5242038871, -.7095721400, -.6776578358]$$

```
> A := ContinuousS(f(x), -1..1, 400):

> H := Histogram(A, -1..1, 12):
> P := plot(f(x), x=-1..1):
> plot({P,H});
```

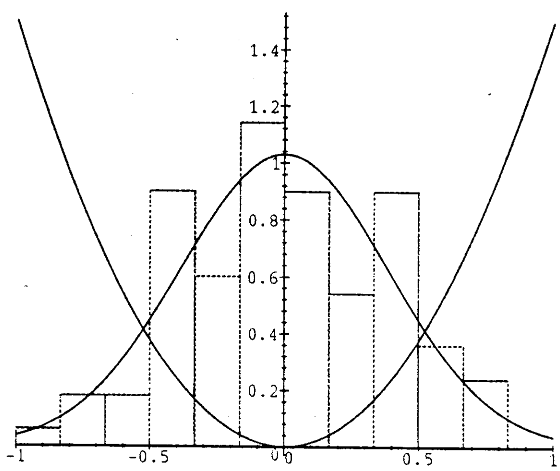


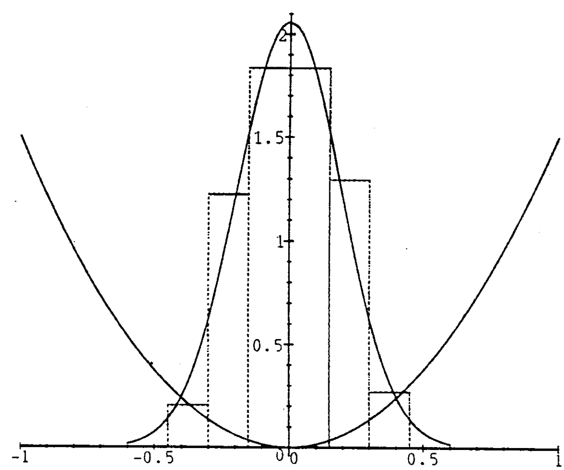
```
> A := [seq(ContinuousS(f(x), -1..1, 4), i=1..100)]:
> A[1];
```

$$[-.5262355326, -.6564311711, -.6818391032, .8768027182]$$

```
> M := [seq(Mean(A[i]), i=1..100)]:
> H := Histogram(M, -1..1, 12):
> n := NormalPDF(0, var/4, x):
> N := plot(n, x=-1..1):
> plot({P,H,N});
```

```
> A := [seq(ContinuousS(f(x), -1..1, 16), i=1..100)]:
> M := [seq(Mean(A[i]), i=1..100)]:
> H := Histogram(M, -0.6..0.6, 8):
> n := NormalPDF(0, var/16, x):
> N := plot(n, x=-0.6..0.6):
> plot({P,H,N});
```





Appendix D: Power Function of a Statistical Test

```
> read(`a:stat.m`); with(plots, display):
```

1. First we fix the sample size n at 21 and vary alpha.

```
> n := 21; var:=100;
```

$$n := 21$$

$$var := 100$$

2. Specifically for alpha=0.05, we calculate L and U, the left and right end-points of the confidence interval associated with this test.

```
> alpha:=0.05;
```

$$\alpha := .05$$

```
> L := var*ChisquareP(n-1, alpha/2)/sigma^2;
```

$$L := 959.0777392 \frac{1}{\sigma^2}$$

```
> U := var*ChisquareP(n-1, 1-alpha/2)/sigma^2;
```

$$U := 3416.960690 \frac{1}{\sigma^2}$$

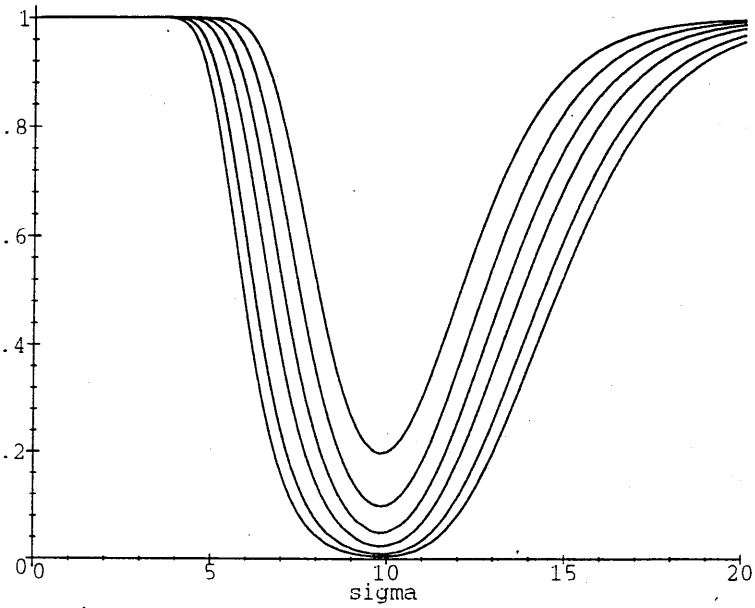
```
> K := 1-ChisquareCDF(n-1,U) + ChisquareCDF(n-1,L);
```

$$K := 1. + .5382288911\ 10^{-53}\ e^{\left(-1708.480345 \frac{1}{\sigma^2}\right)} \big(.3344971066\ 10^{75}\ \sigma^2 + .6595712237\ 10^{65}\ \sigma^{10} + .2253728944\ 10^{68}\ \sigma^8 + .1544229000\ 10^{63}\ \sigma^{12} + .3174263540\ 10^{57}\ \sigma^{16} + .2711583434\ 10^{60}\ \sigma^{14} + .1857945600\ 10^{54}\ \sigma^{18} + .6417419339\ 10^{70}\ \sigma^6 + .6349797023\ 10^{77} + .1566290687\ 10^{73}\ \sigma^4\big)/\sigma^{18} - .2589076543\ 10^{-59}\ e^{\left(-479.5388696 \frac{1}{\sigma^2}\right)} \big(.2678706355\ 10^{77}\ \sigma^2 + .8510187505\ 10^{69}\ \sigma^{10} + .8161931392\ 10^{71}\ \sigma^8 + .7098642504\ 10^{67}\ \sigma^{12} + .1852161810\ 10^{63}\ \sigma^{16} + .4440917903\ 10^{65}\ \sigma^{14} + .3862380981\ 10^{60}\ \sigma^{18} + .6523272256\ 10^{73}\ \sigma^6 + .1427270908\ 10^{79} + .4468803720\ 10^{75}\ \sigma^4\big)/\sigma^{18}$$

```
> P1 := plot(K, sigma=0..20): P1;
```

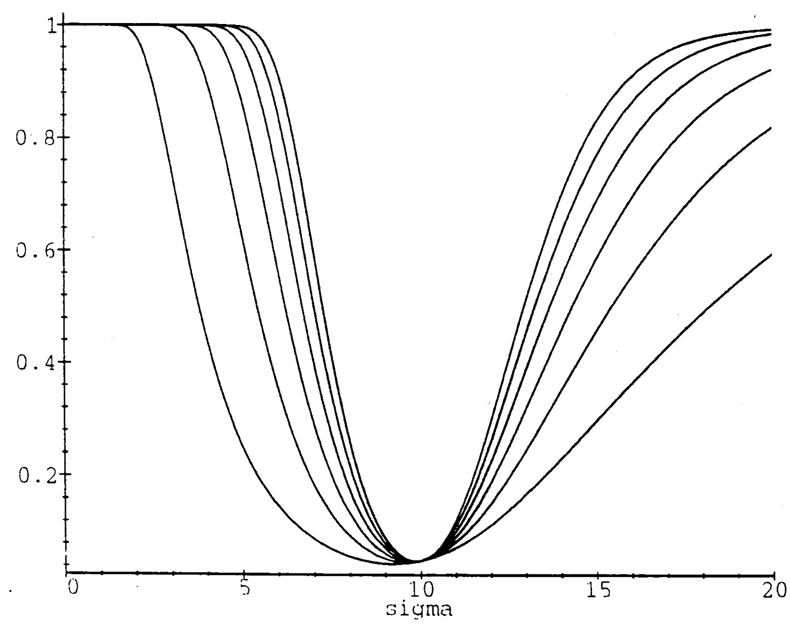
3. Repeat this process for alpha = 0.005, 0.01, 0.025, 0.1, and 0.2.

```
> display([P1,P2,P3,P4,P5,P6], insequence=true);
> plot({P1,P2,P3,P4,P5,P6});
```



4. We can also study the power function when alpha is fixed at 0.05 and n assumes a variety of values (n=5, 9, 13, 17, 21, 25 in the following).

```
> alpha:=0.05;
> for i from 1 to 6 do
>    n := 1+ 4*i:
>    L:=var*ChisquareP(n-1, alpha/2)/sigma^2:
>    U:=var*ChisquareP(n-1, 1-alpha/2)/sigma^2:
>    K:=1-ChisquareCDF(n-1,U)+ChisquareCDF(n-1,L):
>    Q.i:=plot(K, sigma=0..20):
> od:
> display([Q.1, Q.2, Q.3, Q.4, Q.5, Q6], insequence=true);
> plot({Q.1, Q.2, Q.3, Q.4, Q.5, Q.6});
```



Building An Information Superhighway Between Student and Teacher

Jonathan Lewin,

Kennesaw State College, P.O. Box 444, Marietta, GA 30061.
jlewin@kscmail.kennesaw.edu

Abstract

In today's academic environment, many students are denied the sort of quality student-teacher interaction that was taken for granted when their courses were designed. As a result of this lack of communication, academic standards have suffered, particularly in disciplines such as mathematics.

This article suggests a way of opening up a line of communication between student and teacher using a combination of e-mail and *Scientific WorkPlace,* a spectactular new scientific word processing system and front end for *Maple.* Our intention is to enable students to submit questions, assignments and tests to their teachers and to receive handouts and lecture notes from their teachers — all by e-mail.

This article also displays some of the features of *Scientific WorkPlace* as a front end for Maple and reveals the simplicity with which mathematical operations can be performed in *Scientific WorkPlace.*

1 The Communication Dilemma

The reality of today's college environment, particularly in a commuter school like the one in which I teach, is that many students maintain their own households and support themselves by working long hours at their jobs. To accommodate the needs of such students many institutions have been compelled, in recent years, to reduce sharply the number of lectures that are given in each course and to make each lecture correspondingly longer. These longer lectures impose a grave burden on students who simply cannot stay alert for so much time. Furthermore, it is quite common for students in this kind of course to arrive on campus for their classes and then to leave as soon as soon as they can to return to their places of employment. This makes it hard for student and teacher to engage in the one-on-one dialogue that students may need in order to study successfully.

Unfortunately, the classical format of a typical college mathematics program is designed for the traditional full-time student who spends most of his/her time on campus and has ample access to the teacher. We need to face the fact that one of the root causes of student failure is our own failure to recognize that our classical teaching technique was designed to fit an environment that no longer exists at many schools.

2 A Software Solution of The Problem

2.1 The Role of The Software

The purpose of this article is to suggest a solution to the communications dilemma described above. I shall suggest a possible use of modern technology that enables us to give our students individual care even when we cannot meet them face to face. Students will be able to talk to us, to receive lecture notes, to question us, to submit assignments to us and even to be tested by us when they are not in our physical presence. In other words, I shall suggest an efficient way in which students of science and mathematics can engage in what has come to be called **distance learning.** For the establishment of an adequate line of communication between teacher and student, we need two basic ingredients:

- Both teacher and student should have access to a personal computer. Ideally, they should also be able to communicate by e-mail. If e-mail is not available then they should be able to send one another diskettes quickly and efficiently.

- Both teacher and student should have access to special software that enables almost effortless reading and writing of complex mathematical material at the computer display. The documents should, however, be represented by plain text files that can be sent back and forth by e-mail.

It is the second of these two ingredients that has, until now, been most illusive. None of the standard word processing products that I have seen is sufficiently versatile or easy to use to serve as a medium of communication. I am confident, however, that this void has been filled amply by a new product known as *Scientific WorkPlace.*

2.2 *Scientific Word* And *Scientific WorkPlace.*

Scientific Word is a strikingly new kind of word processing utility that works by front ending the LaTeX typesetting system. It is quite unlike any other word processing product on the market and is the word processor of choice for most kinds of scientific or mathematical writing. *Scientific Word* is also sold bundled together with Maple[1] V, Release 3 and this bundle is known as *Scientific WorkPlace.* In *Scientific WorkPlace, Scientific Word* acts as an extremely user friendly front end for Maple and, in this way, it provides the ultimately user-friendly symbolic manipulator and graphing utility.

Using *Scientific WorkPlace,* students can send and receive mathematical documents very easily by e-mail. Even when they are unable to meet with the professor

[1] Maple V is a product of Waterloo Maple Software.

personally, they can submit questions, homework assignments, projects and tests in the form of *Scientific WorkPlace* documents. By using e-mail to answer many of the questions that must today be answered during their office ours, professors will be able to enhance the level of contact they have with students and to make much more efficient use of their office hours. Instead of simply waiting in their offices for students who may or may not turn up, professors will be able to work at the computer screen communicating with their students. Even a student who is out of town at the time of a given lecture can receive a perfect set of lecture notes by providing his/her e-mail address. This is the sense in which *Scientific WorkPlace* can help us to set up an information superhighway between student and teacher.

2.3 Some Added Benefits

One of the most difficult lessons to convey to students of mathematics is that true learning does not take place unless the student has learned to write mathematics in such a way that the sentences are complete, meaningful and readable. All too often, college students are victims of the poor teaching (often coupled with multiple choice testing) that they experienced in their prior schooling. All too often, their work consists of a bunch of meaningless symbols that are dotted aimlessly around the page together with an "answer" that is gift wrapped in a big box.

In order for a mathematics course to be truly successful we have to break this habit. In my opinion, the most fundamental principle of mathematics teaching is that the process of learning mathematics must be seen as being identical to the process of learning to teach mathematics. Whatever mathematics a student writes should be presented as if the student were teaching it to whomever will be reading it. In becoming a good teacher, a student is becoming a good teacher to him or her self. I mention all this because I have noticed among my own students that, the moment they start writing their mathematics on the computer screen they start taking pride in the appearance of their work and in the quality of their mathematical writing. As their writing improves, some of their lifelong bad habits begin to disappear and they become better problem solvers. I believe that the learning of mathematical word processing should be part and parcel of the training that students receive in a college mathematics curriculum.

2.4 Making The Software Available

In order for *Scientific WorkPlace* to serve as a educational tool, it is necessary for students to have access to it. The simplest way to do this is to set up a lab with computers on which *Scientific WorkPlace* is installed or is available over a network.

Ideally, the computers that run *Scientific WorkPlace* should also allow students to send and receive e-mail. Some students, who have their own personal computers, may find it worthwhile to acquire their own personal copy of *Scientific WorkPlace* which is available in a student version at a sharply reduced price.

2.5 Longer Term Goals

I believe that the availability of *Scientific WorkPlace* will encourage teachers to write up their lecture notes and to send them by e-mail to all their students. The advantages of this compiling of lecture notes are as follows:

- Teachers who prepare such notes may give more organized and polished lectures.
- Students who are ill or out of town will still have the course materials at their fingertips.
- Students who lack note taking skills will still have a proper set of lecture notes.
- Courses may come to depend less on expensive (and often worthless) textbooks.
- Teachers will be encouraged to make their lecture notes available to other teachers in their departments. In my own department, for example, our network includes a logical drive H: that we all share. I can copy files to this drive and I can see files that have been placed there by my colleagues. Eventually, a bank of material will begin to accumulate. This material will enable teachers to communicate more closely with each other. It can serve as a basis for faculty discussion and it can be used to guide new faculty and part time faculty.

3 *Scientific WorkPlace* as A Word Processor

In this section I shall provide a brief discussion of the way in which I believe a mathematics course can benefit from the word processing features of *Scientific WorkPlace.* In the next section I shall discuss its capabilities as a friendly front end for Maple.

3.1 Getting The Students Started

Starting in Fall 1995, every one of my classes will begin with two or three lab sessions in which I shall instruct my students in the fundamentals of using the word processing features of *Scientific WorkPlace.* I shall make no attempt to teach my students everything that *Scientific WorkPlace* can do. Instead, I shall teach them exactly what they need to know: how to open a document, how to preview, how to print a document, and how to write in a document that has already been opened.

After those initial sessions, the students will be able to use the lab computers at any time when they are not being used by another class. I shall encourage those students who have their own computers to acquire the student version of *Scientific WorkPlace* if they can afford to do so.

Part of the essence of *Scientific WorkPlace* is that it is so simple to use in this context. The students need know nothing about such things as styles, tab stops, fonts, margins, hyphenation, pagination, space to be left between lines and other similar time-consuming visual activity that is, unfortunately, the bread and butter of writing with other word processing systems. Each student will be provided with a diskette containing a blank document. He or she will be told: "If you want to send me material, open this document, use the operation save as to give it a name and then write what you want into it. If you print the document it will look perfect. If you e-mail the document to me I will be able to read it and, if necessary, print it.

If I e-mail my students a homework assignment then all they have to do is open the *Scientific WorkPlace* document I have sent them and insert their solutions directly into it. If they decide to print their work, the printed copy will be automatically hyphenated, paginated and formatted and will have the appearance of a publication quality document. Alternatively they can e-mail back the *Scientific WorkPlace* document for me to read on screen.

At worst, if e-mail is not available, then the teacher can make the file available on diskette and students can return the diskette with their completed homework. This is still vastly superior to handing in homework that has been written with pencil and paper.

I am confident that I will be able to make my students proficient users of *Scientific WorkPlace* and that even students who are not computer or keyboard experts will learn to feel comfortable writing mathematics this way. After a little practice with *Scientific WorkPlace,* they will be able to write complicated mathematical expressions as fast on the computer screen as one would write by hand. In fact, an experienced user can write more rapidly using *Scientific WorkPlace* because repeated parts of a mathematical expression can be produced in a flash using *Scientific WorkPlace's* drag and drop features. For example, in order to type

$$\sqrt{\frac{1-\cos x}{1+\cos x}}=\sqrt{\left(\frac{1-\cos x}{1+\cos x}\right)\left(\frac{1-\cos x}{1-\cos x}\right)}=\frac{1-\cos x}{|\sin x|}$$

there is no need to type the fraction $\frac{1-\cos x}{1+\cos x}$ more than once and there is no need to type the fraction $\frac{1-\cos x}{1-\cos x}$ at all.

3.2 Workshops for Faculty

In order for the students to benefit properly from a knowledge of *Scientific WorkPlace,* it is important that as many as possible of their courses make use of it. Therefore, we need to convince as many as possible of the professors in the department to take the trouble to become competent *Scientific WorkPlace* users. In my own department we have acquired a faculty site license that allows us to install *Scientific WorkPlace* into our office computers and we conduct workshops in order to train our faculty to become efficient users of the system.

As soon as the system is accepted as a word processor within the department, members of the department should start exchanging lecture notes, homework assignments, test papers and other material that has been written in *Scientific WorkPlace* documents. If the department computers are linked by a network then it would be a good idea to provide a special place in the network for this shared material.

Members of the faculty should also be reminded that they can send mathematical material to one another using e-mail. This even applies to more complicated documents that contain pictures because, although the pictures themselves may be binary files, *Scientific WorkPlace* allows us to store the document in a "wrapped" form in which the entire document, pictures and all, is a plain text file.

4 Using The Maple Operations

In this short article, I can't even begin to describe all of the Maple operations. I shall demonstrate just a few of them in order to whet the appetite for more and to exhibit their intuitive simplicity when they are performed in *Scientific WorkPlace.* This simplicity is the key difference between *Scientific WorkPlace* as a front end for Maple and any other symbolic manipulator that I have seen. In order to use one of those other products it is necessary to type an instruction in its native language requesting the solution to the given problem. This is not merely inconvenient because, in order to be able to write in that language, one needs to have learned it in the first place. One also needs to be able to remember it. Thus, in order to use any of those products with my students I would have to begin by teaching them how to write commands in language that the product understands.

In sharp contrast, all of the symbolic manipulations that are performed by Maple in *Scientific WorkPlace* can be obtained by a couple of simple mouse clicks on mathematical expressions that appear on the screen in exactly the same form in which we would have written them down on paper. It takes me only a few minutes to make my students feel like expert users. To show this contrast I have included with some of the examples that follow a brief description of my attempts to solve the same problem using Mathematica[2] Version 2.2 for Windows.

4.1 Some Simple Evaluations

We begin with the limit

$$\lim_{n\to\infty}\frac{(n!)^2 e^{2n}}{n^{2n+1}}.$$

To evaluate this limit in *Scientific WorkPlace*, one simply has to click on the Maple menu with the mouse and then click on the option evaluate. This will produce the answer:

$$\lim_{n\to\infty}\frac{(n!)^2 e^{2n}}{n^{2n+1}}=2\pi.$$

I tried several times to evaluate this limit using Mathematica but I never did succeed. For my fifth try I typed

Limit[(((Gamma[n+1])^2)*(Exp[2*n]))/n^(2*n+1),n->Infinity]

which yielded several complaints about essential singularities. Then I gave up. Presumably there is a way to ask Mathematica to find this limit but it is beyond me and I dread the thought of teaching it to students.

[2]Mathematica is a product of Wolfram Research Inc.

Although I have ragged Mathematica a little in this article I must emphasize that Mathematica is a first class high end symbolic manipulator. There are many important types of application for which Mathematica is a *sine qua non.* No doubt, my failure to produce results using Mathematica is the result of my own incompetence. The point is, however, that though I am equally incompetent to use Maple in *Scientific WorkPlace* I can nevertheless perform any operation I like.

Some other expressions that I evaluated in *Scientific WorkPlace* with a couple of mouse clicks are the following:

$$\lim_{x\to 0} \frac{e - ex/2 - (1+x)^{1/x}}{x^2} = -\frac{11}{24}e,$$

$$\sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{1}{90}\pi^4,$$

$$\begin{aligned}\frac{d^3}{dx^3}x^4e^x\cos 5x &= 24xe^x\cos 5x + 36x^2e^x\cos 5x - 180x^2e^x\sin 5x - \\ &\quad 288x^3e^x\cos 5x - 120x^3e^x\sin 5x - 74x^4e^x\cos 5x + \\ &\quad 110x^4e^x\sin 5x\end{aligned}$$

4.2 Evaluating An Integral

To evaluate a simple integral such as

$$\int_0^1 \sqrt{1-x^2}dx$$

in *Scientific WorkPlace,* all one has to do is type the expression, place the cursor there, click on Maple and evaluate. This yields

$$\int_0^1 \sqrt{1-x^2}dx = \frac{1}{4}\pi.$$

I asked Mathematica to find this integral as well. After a couple of tries I typed

Integrate[Sqrt[1-x^2],x,0,1]

and I received output that looked something like

Pi
— .
4

I returned to *Scientific WorkPlace* and typed the harder integral

$$\int_0^1 \sqrt[3]{1-x^2}dx.$$

A click on Evaluate gave

$$\int_0^1 \sqrt[3]{1-x^2}dx = \frac{2}{15}\left(\sqrt{\pi}\right)^3 \frac{\sqrt{3}}{\Gamma\left(\frac{2}{3}\right)\Gamma\left(\frac{5}{6}\right)}.$$

(I decided that I didn't want to see the latter expression in Mathematica.) Clicking on Evaluate Numerically yielded

$$\int_0^1 \sqrt[3]{1-x^2}dx = .84130926319527255670501144743.$$

4.3 Solving An Ordinary Differential Equation

I am no expert on differential equations but I thought I should show one anyway. So I began with the equation

$$x^2y'' - xy' + y = 0. \tag{1}$$

When I placed the cursor in this equation clicked on the Maple option Solve ODE, I obtained,

$$\text{Exact solution is}: y(x) = C_1x + C_2x\ln x.$$

Since Equation (1) is a simple Euler equation I decided to spice it up a little. So I tried a bunch of variations. For each of these I clicked on the Solve ODE option in *Maple.* The results included Bessel functions, hypergeometric functions and various other special functions that I have never heard of. Then, quite by accident I tried the equation

$$\left(1 - x^2\right)y'' - xy' + y = 0 \tag{2}$$

which yielded

$$\text{Exact solution is}: y(x) = C_1x + C_2\sqrt{(-1 + x^2)}.$$

I don't know what it is about Equation (2) that distinguishes it from equations like

$$\left(1 - x^2\right)y'' + xy' + y = 0$$

whose solutions involve special functions. But this experimentation with *Scientific WorkPlace* taught me something that I could never have discovered on my own. Finally, I tried some nonhomogeneous variations of Equation (2). Among these, the equation

$$\left(1 - x^2\right)y'' - xy' + y = x$$

gave me

$$\begin{aligned} y(x) &= -\frac{1}{2}\frac{-x\sqrt{(-1+x^2)} - \ln\left(x + \sqrt{(-1+x^2)}\right) + \left(\ln\left(x + \sqrt{(-1+x^2)}\right)\right)x^2}{\sqrt{(-1+x^2)}} \\ &\quad + C_1x + C_2\sqrt{(-1+x^2)} \end{aligned}$$

and the equation

$$\left(1 - x^2\right)y'' - xy' + y = x^2$$

gave me

$$y(x) = \frac{2}{3} - \frac{1}{3}x^2 + C_1x + C_2\sqrt{(-1 + x^2)}.$$

I think that the simplicity of *Scientific WorkPlace* makes it an ideal tool for experimentation. I would have liked to try these equations with Mathematica as well but I simply couldn't see how to make Mathematica recognize the symbol y''. I got as far as solving the equation $y' = 3$. For this purpose I had to type

```
DSolve[y'[x]==3,y[x],x]
```

which yielded the output

{{y[x]->3x+C[1]}}

Scientific WorkPlace gives the solution as $y(x) = 3x + C_1$ which I kinda prefer.

4.4 Using Maple as A Writing Tool

Let's say one wants to write

$$\lim_{t\to x} \frac{t^7 - x^7}{t - x} = \lim_{t\to x} \frac{(t-x)\left(t^6 + t^5x + t^4x^2 + t^3x^3 + t^2x^4 + tx^5 + x^6\right)}{t-x}.$$

There is no need to type out the expression on the right. All one needs to do is use drag and drop to repeat the expression on the left, highlight the numerator $t^7 - x^7$, hold down the control key and click on the Maple function Factor.

4.5 Drawing A Graph

I shall include just one graph just to emphasize that graph drawing is included in all the remarks I have made above. In order to draw the graph

$$z = xy \sin x$$

all one needs to do is type $xy \sin x$, open the Maple menu and click on Plot 3D, Rectangular. This yields the graph shown in Fig. 1.This graph is just one of a wide

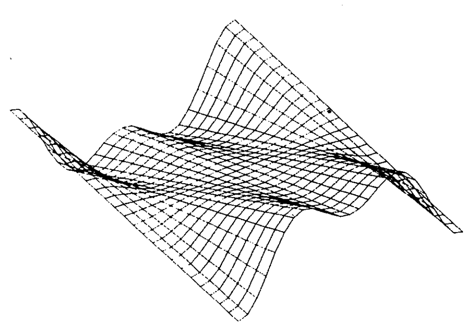


Figure 1: The graph $z = xy \sin x$

variety of possible graph types that are supported by *Scientific WorkPlace*. As with other Maple operations, the drawing of graphs is simple and quick. So is the act of rotating the graphs and zooming towards them or away from them. The entire process is simple enough to encourage students to experiment with a large number of examples, to make them feel like experts and to provoke them to learn from their experience.

5 In Conclusion

The value of *Scientific WorkPlace* as a tool for the teaching of mathematics is twofold:

- *Scientific WorkPlace's* unique word processing capability makes it ideally suited as a communications medium between teacher and student.

- As a front end to Maple, *Scientific WorkPlace* provides simplicity and convenience that is not available with other products. Students using *Scientific WorkPlace* can be taught rapidly and simpiy to perform complicated operations and the process of operating Maple is simple enough to encourage the user to experiment.

I believe that if we take advantage of these unique features of *Scientific WorkPlace* we can take a big step forward in our quest for quality mathematics education.

A THEOREM-PROVER FOR GEOMETRY OVER ELLIPTIC CURVES[1]

R. Padmanabhan[2]
Department of Mathematics
University of Manitoba
Winnipeg, Manitoba
Canada R3T 2N2
e-mail: padman@mira.cc.umanitoba.ca

"The development of mathematics toward greater precision has led, as is well known, to the formalization of large tracts of it, so that one can prove any theorem by using nothing but a few mechanical rules".

-- K. Gödel

Abstract.
This article describes a new geometric theorem-prover based on certain local-to-global principles of the classical algebraic geometry. First we formalize the rigidity of morphisms of a projective curve (defined over an arbitrary algebraically closed field) into certain inference rule for a suitable equational theory, called the rule (gL) here. Then we transform the geometric concepts and statements associated with an elliptic curve into equivalent algebraic ones via a system of equational definitons. Our goal is to prove incidence theorems of projective nature valid for the points on the curve. After transforming these geometric statements into algebra, we formally derive the equational conditions which correspond to the required geometric conclusions. Naturally, these proofs reside completely within the realms of first order logic with equality. We apply these techniques to some classical examples of synthetic geometric constructions given on elliptic curves using conics. To further demonstrate the formal nature of such equational proofs, we use a new version of OTTER, a theorem-prover for the first order logic with equality developed at the Argonne National Laboratory, Illinois, USA in which the rule (gL) has already been implemented. This formal mechanism is pressed into service to 'automatically' prove the incidence theorems without any further reference to the underlying geometry or topology of the curves!

Throughout the human history, developments in the sciences have caused people to constantly change their views of humans and their place in the cosmos. Galileo showed that we are placed on a tiny planet which revolves around the sun; the Darwinian revolution changed our view of human origins; Gödel's incompleteness theorems shed new light on the difference between truth (= what humans can perceive) and provability (= what humans can actually demonstrate). Now the electronic revolution has rekindled the age-old question: can computers learn, think or reason the way we humans do? In other words, can the process of reasoning be automated?

[1]Based upon a lecture given at the First Asian Technology Conference in Mathematics, Singapore (ACTM 95).

[2]This research was supported by an operating grant from the NSERC of Canada (#A8215)

By "automated reasoning" we mean a formal process of drawing conclusions from a given set of facts - i.e. a process which automatically generates new facts by drawing conclusions from the hypotheses or from facts which are already established. New facts are obtained by applying certain specific, context sensitive, rules of reasoning, which are known as *rules of inference*. Let me illustrate this with some classical examples (of the so called "commutativity theorems") in increasing order of sophistication:

(1) In a group, if the square of every element is the identity, then the group is abelian.
(2) If R is a ring in which for every element $x \in R$ $\exists$ an integer $n(x) > 1$ such that $x^{n(x)} = x$ then the ring R is commutative.
(3) Every polynomially defined group law on the plane $\mathbb{R}\times\mathbb{R}$ is commutative.
(4) Every group law definable as a morphism on a complex elliptic curve is commutative.
(5) If Γ is a projective curve and if $\langle\Gamma; f(x,y), e\rangle$ and $\langle\Gamma; g(x,y), e\rangle$ are two algebraic groups with same identity element "e", then $f(x,y) = g(x,y)\ \forall x,y \in \Gamma$.

Of these, the first two properties lie strictly within the framework of first order logic with equality and hence, thanks to the completeness theorem of equational logic, are provable within the relevant equational theory using the famous deduction rules of Garrett Birkhoff (e.g. [WT], p. 381 or [WW], Theorem 29, p.180). For example, the properties (1) and (2) say, respectively, that

$$\{\text{group axioms}, x^2 = e\} \quad |\!=\!(eL)\!= \quad \{xy = yx\}$$
$$\{\text{ring axioms}, \forall x \in R\ \exists\, n(x)\ (x^{n(x)} = x)\} \quad |\!=\!(eL)\!= \quad \{xy = yx\}$$

where "|=(eL)=" stands for the standard (Birkhoff) set of inference rules of equational logic. However, in the last three examples, one works with special models **K** of algebras of a given type, usually richer in structure that the class of all models of that type. For example, it is easy to see that it is impossible to prove the property (5) within the scope of equational logic. Let Γ be the three-dimensional space $\mathbb{R}^3$ and consider the two group laws $f(x,y)$ and $g(x,y)$:

$f((a,b,c), (x,y,z)) = (a+x, b+y, c+z)$, the usual direct product of '+' and
$g((a,b,c), (x,y,z)) = (a+x, b+y+az, c+z)$.

It is clear that both the algebras $\langle\Gamma; f(x,y), e\rangle$ and $\langle\Gamma; g(x,y), e\rangle$ with $e = (0,0,0)$ are even **algebraic groups** having the same identity element, but while f is commutative g is not! Property (5) claims that this cannot happen for group laws on an elliptic curve.

Recall that an elliptic curve, viewed as a plane algebraic curve in the complex projective plane, is given by a non-singular cubic equation. In this project, we bring out the equational properties of the algebraic laws *naturally definable* on cubic curves (i.e.by means of synthetic constructions) from the point of view of equational logic. One may wonder what equational logic - a topic in universal algebra - has got to do with cubic curves - a topic in classical algebraic geometry.

"Elliptic curves" is one of those fascinating subjects which ignore customary interdisciplinary boundaries and where, for instance, rationally defined quasigroup laws, certain combinatorial configurations, formal groups, number theory and incidence theorems all happily coexist. We say that an algebraic curve Γ admits an algebraic law (of arity n), if there exists an n-ary morphism $\mu(x_1,x_2,...x_n)$ on the curve Γ (i.e. $\mu(x_1,x_2,...x_n)$ is a regular function or a rational function on the product set $\Gamma^n = \Gamma\times\Gamma\times...\times\Gamma$). The non-singular cubic curves are pregnant with a number of such algebraic laws all of which are morphisms of the curve. For example, the famous Cayley-Bacharach Theorem of classical algebraic geometry says that every algebraic curve induces a rational operation on cubic curves via a complete intersection cycle (see e.g. the figure 2 for the 5-ary conic process). Without much further ado, let us list a few equational properties enjoyed by such algebraic laws defined on non-singular cubic curves:

1. Any two group laws sharing the same identity are, indeed, equal.
2. Any group law definable on an elliptic curve must be commutative.
3. Every quasigroup law definable on such a curve gives rise to a group law.
4. Every cancellative di-associative groupoid on an elliptic curve must be associative.
5. Every groupoid having a two-sided identity element 'e' is already a group law.
6. Any two binary Steiner quasigroup laws having a common idempotent are equal; i.e. $\{f(x, f(y,x)) = g(x, g(y,x)) = y,\ f(e,e) = g(e,e) = e\} \vdash_{(gL)} \{\forall x,y\ f(x,y)=g(x,y)\}$
7. Any two 5-ary Steiner quasigroup laws having a common idempotent are equal (see the Lemma 7 below for details).

One of the purposes of this research is to develop a new geometric theorem-prover based on certain local-to-global principles of the classical algebraic geometry so that we can formulate and prove properties like the above purely within the realms of first order logic with equality. To this end we need additional *context-sensitive rules of inference.* One such rule is the famous rigidity lemma which is valid for the morphisms of projective curves, in particular for elliptic curves, (see e.g. p. 104 in [JM] or p. 152 in [IS]):

$$\exists y_0 \exists z_0 \forall x\ (f(x,y_0)=z_0) \Longrightarrow \forall x \forall y \forall z\ (f(x, y)=f(z, y)) \quad ....(gL)$$

i.e. whenever the deduction procedure meets the local equality $f(x,y_0) = z_0$ for some word f and some elements y_0, z_0 then it churns out the global multi-variable identity $f(x, y) = f(z, y)$. Multi-variable because here x, y or z could be vectors i.e. $x = (x_1,x_2,..,x_m)$ etc. On a much elementary level, the rule (gL) is clearly valid for the so-called affine algebras: $p(x_1,x_2,..,x_m) = \sum n_i x_i + k$ in any abelian group.

Otter and the Implementation of the Rule (gL).

Otter [see WM, page 19] is a computer program developed at the Argonne National Lab in Illinois, U.S.A. that attempts to prove theorems stated in first-order logic with equality. Here we restrict our attention to its capabilities in equational nd implicational logics. The user inputs axioms and the denial of the goal(s), and Otter searches for a contradiction by working both forward from the axioms and backward from the goal(s). Equational reasoning is accomplished by paramodulation and demodulation. Paramodulation is equality substitution extended with unification: if the two terms in question can be made identical by instantiating variables, then equality substitution is applied to the corresponding instances. Demodulation is the use of equalities as rewrite rules to simplify other equalities. The following example illustrates the interplay between paramodulation and demodulation. Consider {f(x,f(g(x),y))=y, f(u,g(u))=e, f(w,e)=w}, with e nullary; Otter can infer x=g(g(x)) "in one step" by unifying f(u,g(u)) and f(g(x),y)) (which instantiates u to g(x) and y to g(g(x))), replacing f(g(x),g(g(x))) with e, then demodulating with f(w,e)=w.

The rule (gL) is implemented in Otter in two ways that are analogous to paramodulation and demodulation. Let f be the operator to which (gL) applies, and let F[a1,x] represent a term in f that contains subterm a1 at a particular position, with x representing everything else in the term. Given F[a1,x]=F[a2,y], (i.e., a1, and a2 are in corresponding positions), with a1 and a2 unifiable. By (gL) we infer F[z,x']=F[z,y'], where z is a new variable, and x' and y' are the appropriate instances of x and y. For example, from f(f(x,y),f(z,f(x,z)))=f(u,f(y,u)), we can (gL)-infer f(f(x,y),f(z,w))=f(f(x,z),f(y,w)) by unifying u and f(x,z). We also use (gL) as a rewrite rule whenever possible. That is, we rewrite F[a,x]=F[a,y] to F[z,x]=F[z,y] (again, z is a new variable).

Notations. Variables are distinguished from constants by starting with u, v, w, x, y, or z. Proofs are by contradiction, and the denials of the goals contain constants, i.e., objects for which the goal fails to hold. The justification for each step is in brackets and specifies the inference rule and any rewriting that occurs. The inference rules are "para_from" (substitute into the second equality), "para_into" (substitute into the first equality) and "gL". Rewriting is specified with either "demod,..." (simplification with ...) or "gL-id". Also, if Σ is a set of identities and if σ is an identity in the language of Σ, then we write

$$\Sigma \models\!\text{(gL)}\!= \sigma$$

if $\{\Sigma \cup \text{(gL)}\} \models\!\text{(eL)}\!= \sigma$ in the usual equational logic ([WW], Theorem 29, p.180). Whenever convenient, we also say that the axioms Σ "(gL)-implies" σ etc.

Applications to the Projective Geometry of Cubic Curves.

As an example, let us produce a rather "mindless" (i.e.automated) proof of the powerful four-variable median law (x•y)•(z•t) = (x•z)•(y•t) for the classical binary morphism '•' of chord-tangent construction on non-singular cubics just from the relatively weak two-variable Steiner quasigroup laws {x•(y•x) = y, (y•z)•z = y} without any reference to the geometry or the topology of curve.

Theorem 1. {x•(y•x) = y, (y•z)•z = y} $\models$(gL)= {(x•y)•(z•t) = (x•z)•(y•t)}

Human Proof.

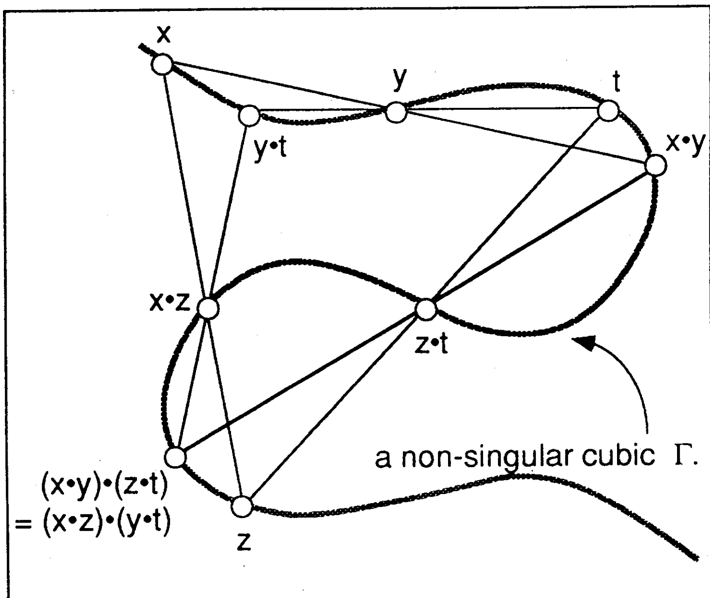
Define the 5-ary composite operation f(x,y,z,t,u) by f:=((xy)(zt))(u((xz)(yt)).

Now we have, by the law x(yx) = y, f(x,c,c,t,d) = d for all x. Thus by the rule (gL), the 5-ary expression f(x,y,z,t,u) does not depend upon x <u>for all</u> y,z,t,u. In paticular, we have

	f(x,y,z,t,u)	= f(x_1,y,z,t,u)	∀x ∀x_1
i.e.	((xy)(zt))(u((xz)(yt))	= ((x_1y)(zt))(u((x_1z)(yt))	∀x ∀x_1
(let now x_1=yz)		= (((yz)y)(zt))(u(((yz)z)(yt))	
		= t(ut) ---- by the Steiner laws	
		= u	
		= ((xz)(yt))(u((xz)(yt))	

and hence one right-cancellation of the common term '(u((xz)(yt))' yields right away the desired median law (x•y)•(z•t) = (x•z)•(y•t).

Comapre the above with an "automated proof" obtained by OTTER (see Figure 1), with even a milder assumption i..e the validity of Steiner law only at 'e' which is, in fact, the spirit of thè rule (gL). This generalization is a new discovery made by the machine!



Lemma. {x.(e.x) =e} =(gL)=> {(xy)(zt)=(xz)(yt)}
----> UNIT CONFLICT
at 7.11 sec ----> 987 [binary,986,5] $F.
Length of proof is 3. Level of proof is 3.
---------------- PROOF ----------------
3 [] f(x,f(e,x))=e.
5 [] f(f(A,B),f(C,D))!=f(f(A,C),f(B,D)).
8 [para_into,3,3,gL-id] f(x,f(z,x))=f(y,f(z,y)).
23 [para_into,8,8] f(f(x,y),f(z,f(x,z)))=f(u,f(y,u)).
986 [gL,23] f(f(x,y),f(z,u))=f(f(x,z),f(y,u)).
987 [binary,986,5] $F.
----------- end of proof ------------
human steps for #23:
(xy)(z(xz))
= (ab)(y(xy)) by #8
=(u(yu)) by #8
This IS a discovery by the machine!!

Figure 1

We now apply this to the geometry of plane cubic curves without any further reference to the underlying geometry or the topology of curves:

Since a non-singular cubic curve is a projective curve and since the chord-tangent law of composition obviously satisfies the Steiner law $x \bullet (y \bullet x) = y$, we do have the following

Corollary 2. The classical binary morphism "•" of chord tangent construction defined on a non-singular cubic curve is medial (Figure 1).

Historical remark. This important identity for cubics is usually proved using some sophisticated machinery like the Bezout theorem, Riemann-Roch theorem or Weierstrass elliptic functions etc. Thus, it was first proved for plane cubic curves by I.M.S. Etherington using the classical Bezout theorem (see [1]). In [7], the first author gave a proof for elliptic curves over an arbitrary algebraically closed field k. See also an elaborate proof of this identity by using the important concept of intersection multiplicities in ([AK], Lemma 3.9, pp. 68 - 70). The associativity of the classical group law is a simple consequence of this identity. Now let us connect this basic operation * of chord-tangent construction with the classical group law on elliptic curves.

Theorem 3. Let + and * be any two binary morphisms on a non-singular cubic curve over an algebraically closed field k let 'e' be a rational point (i.e. a k-point) satisfying the identities $\{x+e=e+x=x, x*(y*x)=y\}$. Then $z+t = ((e*e)*z))*(t*e)$ and also '+' is a group law.

Proof. (as produced by OTTER using the rule of inference (gL)).
3 [] x+e=x.
4 [] e+x=x.
5 [] x* (y*x)=y.
6 [] ((e*e)*a)* (b*e)!=a+b.
20 [para_into,5.1.1.2,5.1.1] (x*y)*x=y.
28 [para_into,20.1.1.1.1,4.1.1] (x*y)* (e+x)=y.
53 [para_into,6.1.1.1.2,3.1.2] ((e*e)* (a+e))* (b*e)!=a+b.
1231 [para_into,28.1.1.1.2,28.1.1,gL-id] (x*y)* (z+x)= (u*y)* (z+u).
1778 [para_from,1231.1.1,20.1.1.1] ((x*y)* (z+x))* (u*y)=z+u.
1779 [binary,1778.1,53.1] $F.------------ end of proof -------------

Notice that in Step #1778, Otter did obtain a 4-variable generalization of what we really asked for. In particular, if * is the binary morphism of chord-tangent construction and if 'e' is chosen as a point of inflexion, we have $z+t = ((e*e)*z))*(t*e) = (e*z)*(t*e) = (z*t)*e$ which is precisely the classical group law construction as defined by Poincaré (see e.g. Theorem 38 in [AK]). Also, $(x+y)+z = ((x*y)*e)*z = ((x*y)*e)*((z*e)*e) = ((x*y)*(z*e))*((e*e)$ by the median law of Theorem 1, which is now symmetric in y and z (again, by the median law) and hence the binary operation '+' is a monoid. One can easily verify that $-x := x*e$ is the group inverse. Moreover, from the above proof we may appreciate the full power of the rule (gL) by noticing that

Corollary 4. <u>Any</u> binary morphism '+' defined on an elliptic curve over an algebraically closed field and admitting a two-sided identity must be an abelian group law on the curve!

The logic of this project in now clear. First we transform the geometric concepts and statements associated with an elliptic curve into equivalent algebraic ones via a system of equational definitions:

Geometric Concepts	Equational Definitions and/or Characterizations
points	elements
lines (chords)	{(P,Q)| P,Q elements}tangents{(P,P)| P an element}
P,Q,R collinear	P*Q = R
Q lies on tangent at P	P*P = Q
inflexion points	{P| P*P = P i.e. idempotent elements.
singular points	{P| P*Q = P for all Q} i.e.P is an absorbing element.
non-singular curve	no absorbing elements (a special case of the rigidity lemma!)
conic	{(P,Q,R,S,T)| i.e. all 5-tuples}
sextic points P	{P|((P*P)*(P*P))*P = P}
an elliptic quartic	a ternary operation satisfying certain equational identities
bitangents PQ	2p + 2q = 0 under the induced group law (for elliptic quartics)

After transforming the geometric statements into the language of algebra, we formally derive the equational conditions which correspond to the required geometric conclusions. Let us start with a simple example of a well-known geometric property and give two proofs: one informal proof using classical ideas and a new formal equational proof using the above dictionary and the identity established in Theorem 1.

Theorem 5. Let P and Q be two point on C, a non-singular cubic curve in the complex projective plane. If the tangent at P passes through Q and the tangent at Q passes through P, then P = Q and moreover it is an inflexion point.

Proof.

traditional proof: If P ≠ Q, then let ℓ be the line PQ. Let us compute the complete intersection cycle of the curve C with the line ℓ:

$$C \cdot \ell = \{ P,P,Q,Q\}$$

since it is both the tangent at P and the tangent at Q. However, C being of degree 3, it cannot have more than three common points, counting multiplicities! In other words, no cubic curve can accommodate a bitangent!! Hence Q = P and since the tangent at P now touches the curve again at P, P is an inflexion point.

equational proof: In the language of the binary morphism "*" of chord-tangent construction, we have P*P = Q and Q*Q =P. Let us compute:

$$P = (P*P)*P = Q*P = Q*(Q*Q) = Q$$

and since P*P = P now, P is an inflexion point!!

Finally, we mention one typical example of a non-trivial construction theorem in classical algebraic geometry which could be proved within the first order logic with equality. Such constructions are special cases of the so-called Cayley-Bacharach Theorem (see Theorem 1.4.19, p. 74 in [MN]).

Theorem 6. Let Γ be a non-singular cubic and let x,y,z,t,u be five points on the curve. Let Q be the unique conic determined by these 5 points. By the celebrated Bezout theorem of classical geometry, we have $|\Gamma \cap Q| = 6$, counting multiplicities. Let now $f(x,y,z,t,u)$ be the 5-ary morphism on G defined by the complete intersection cycle $\Gamma \cap Q = \{x, y, z, t, u, f(x,y,z,t,u)\}$. Then the sixth point $f(x,y,z,t,u)$ can be found by a simple ruler construction as shown in the diagram below (Figure 2).

For a conventional proof of the ruler construction, see [MPW]. Here we indicate the methodology for an automated proof residing totally within the realms of first order logic with equality.

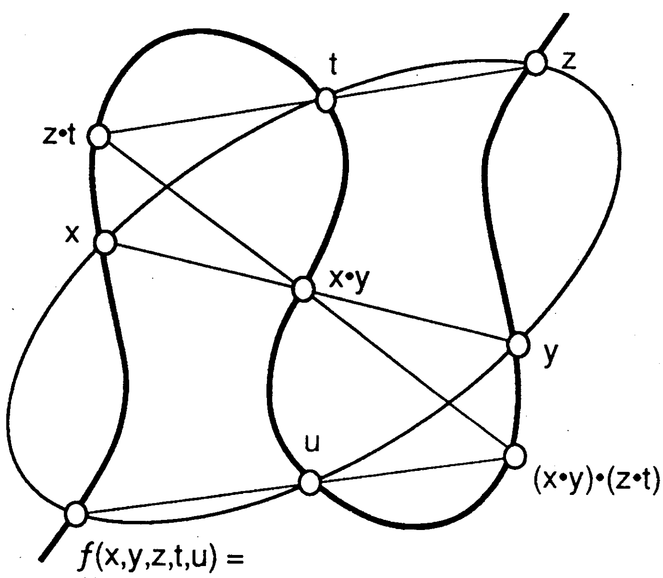


Figure 2

To this aim, let us define g(x,u,z,t,u) = ((x•y)•(z•t))•u where "•" stands for the binary morphism of chord-tangent construction on the cubic. To prove the theorem, we have to show the validity of the identity $f(x,y,z,t,u) = g(x,y,z,t,u)$ for all $x,y,z,t,u \in \Gamma$. The spirit of algebraic geometry implicit here is that the above identity is globally valid if the local validity after specializing z, t and u to say "e", an inflexion point, the traditional zero of the group law on Γ can be formally verified. Here both the hypothesis and the conclusion lie within the framework of first order logic with equality and thus is a most natural candidate for our first order theorem-prover. With this in mind, let us state a technical lemma:

Lemma 7. (Uniqueness of 5-ary Steiner law, see [PM2] or the WWW site mentioned below) Let Σ be the set of identities of type $\langle 5,5,0 \rangle$ defined by $\Sigma = \{\ \mu(e,e,e,e,e) = e$, μ is symmetric, $\mu(\mu(x,y,e,e,e),y,e,e,e) = x$, for $\mu \in \{f, g\}\}$. Then $\Sigma \models$(gL)= $\{f(x,y,z,t,u) = g(x,y,z,t,u)\}$.

The above result on the ruler construction for finding the unique sixth point common follows right away from this Lemma. As a further application to this result, we give a first-order proof of the following theorem of classical algebraic geometry (see exercise #2, page 75 in [MN]).

Theorem 8. Let a,b,c,d be four non-collinear fixed points on a non-singular cubic curve C and let Q be an irreducible conic passing through a,b,c and d. Since C and Q have six common points (counting multiplicities), let C.D = a+b+c+d+p+q, say. Then the line pq always passes through a fixed point of the cubic C.

Proof. By the ruler construction given in Theorem 6, we have

$$((a\bullet b)\bullet(c\bullet d))\bullet p = q$$

as the sixth common point where now "•" is the usual binary morphism of chord-tangent construction. Multiplying both sides by p and using the definition of "•" (i.e. $p\bullet(u\bullet p) = u$), we get

$$(a\bullet b)\bullet(c\bullet d) = p\bullet q$$

Since a,b,c,d are fixed points on the cubic curve C, (a•b)•(c•d) is a also fixed point on C and hence the variable family of lines pq always pass through this fixed point (see Figure 3).

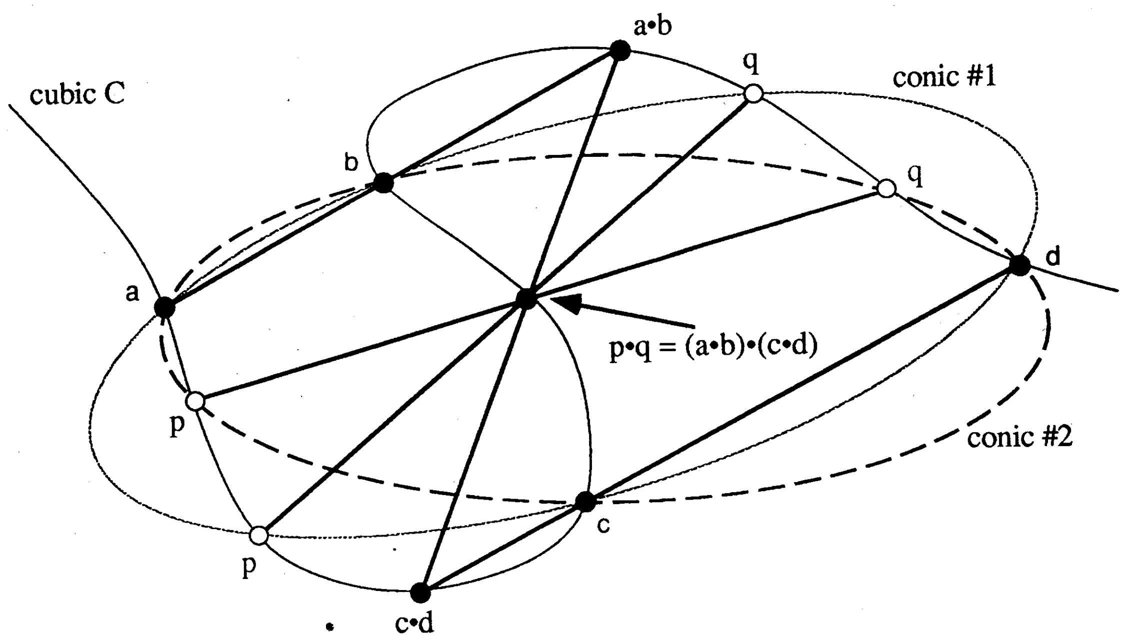


Figure 3

This, in essence, is our joint project [PM3] with Dr. William McCune of the Argonne National Lab in Argonne, Illinois, U.S.A. - a Monograph containing a compendium of about 100 such problems involving identities, implications and closure conditions in the areas of groups, quasigroups, rings, semigroups, lattices and incidence algebras along with the software complete with the necessary input files. More information on open problems, new results, complete Otter proofs, relevant input files, search strategies for single identities etc. can be found in the following worldwide web site http://www.mcs.anl.gov/home/mccune/ar/new_results/index.html.

References

[AK] Anthony W. Knapp — *Elliptic Curves*, Princeton University Press, 1993

[MPW] N.S. Mendelsohn, R. Pdman abhan & B. Wolk — Straight edge constructions on cubic curves, *C.R. Math. Acad. Sci. Canada* 10 (1988) 77-82.

[JM] J.S. Milne — Abelian Varieties in "*Arithmetic, Geometry*", p.103-150, Springer-Verlag, New York, 1986.

[MN] Makoto Namba — *Geometry of Projective Algebraic Curves*, Marcel Dekker, New York and Basel, 1984.

[RP] R. Padmanabhan — Logic of equlity in geometry, *Annals of Discrete Mathematics* 15 (1982) 319-331.

[PM1] R. Padmanabhan and W. McCune — Automatic reasoning about cubic curves, *Computers and Mathematics with Applications*, 29(2), pp. 17-26, 1995.

[PM2] R. Padmanabhan and W. McCune — An Equational Characterization of the Conic Constuction on Cubic Curves, Argonne National Lab Preprint, 1995.

[PM3] R. Padmanabhan and W. McCune — *Equational Logic with OTTER* , a Monograph in preperation.

[WM] W.McCune — Otter 3.0 Manual, Tech. Report ANL-94/6 Argonne National Lab, Argonne, Illinois, 1994.

[IS] I. R. Shafarevich — *Basic Algebraic Geometry*, Springer-Verlag, 1977.

[WT] W. Taylor — Equational Logic, Appendix 4 in G. Gratzer, *Universal Algebra*, Springer Verlag, New York, 1979.

[WW] W. Wechler — *Universal algebra for computer scientists* Springer-Verlag, Berlin, 1992.

Developing the Computer Assisted Materials based on the Student's Misconceptions in Teaching the Calculus.

Shin Hyun-Sung

The intention of study was to develop the computer assisted teaching materials through the investigation of the students' misconceptions in mathematics teaching. Data were collected on the misconceptions of 180 secondary students and 15 college freshmen who study the calculus.

Each student was administered 45 test items with the written and oral forms which fell into five levels of understanding extended from Hart's Scheme.

This study wrote the consistent misconceptions that they had shown in understanding the basic notion of elementary analysis and some ideas related to developing the computer assisted materials which help them overcome those misconception.

Over the past years many research efforts have focused on the study of the mathematics structure and structure of understanding. It was assumed that the students could grasp rather complex math. Topics, provided they were presented in a from appropriate to the students' level of intellectual development.

Piaget(1960) was explicitly concerned with the process and development of thinking. By providing an example of the sum of inner angles in a triangle, he conveyed his notion of the structure of understand. His structure was something actively constructed by the human.

Meanwhile Ginsburg(1990) attempted to situate understanding and his view pictured it as a sense-making procedure. Through the view of understanding the author found there was no discrete point which we could say that it was present or absent.

In this study one research which we can investigate the structure of their understanding is implicitly suggested and their misconceptions in the advanced mathematical thinking are also discussed. Finally the strategies of developing the computer assisted materials bassed on the research results are attempted.

Research problem and design

The intent of study was to investigate the students' misconceptions in learning mathematics and develop the computer assisted materials based on those misconceptions. To do this the following resources were necessary.

(1) Consistent error patterns or inconsistent error patterns in the students' mathematical ideas.

(2) Strategies used to create an understanding.

During the school year 1993 ~ 1994, 180 eleventh graders from 3 classes at 3 high were administered 45 items including both written forms and oral question. Those items fell into five levels of understanding extended from Hart's scheme.

All students were asked to perfrom the written items, which required the students to explain and justify their solution strategy.

The students who finished the written from continued to answer the oral question form, but those who did not give clear answers on the paper were asked to interview. These interviews were conducted by well trained teachers, including the author. All interviews were tape-recorded and completely transcribed afterwards. The oral forms consisted of test items covering topics not easily represented in written from and which needed a variety of answers from student.

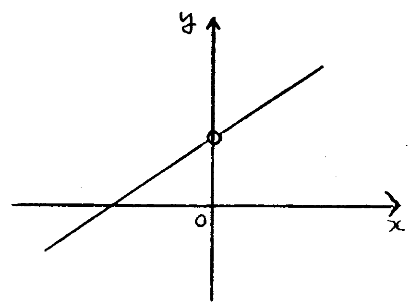
Result and discussion

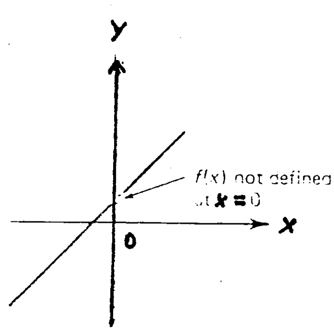
(1) Limit

The major difficulty seemed to be in understanding of the limit process $\lim_{x \to a^+} f(x) = \lim_{x \to a^-} f(x)$ and numerical construction of the sequence $\{f(x_n)\}$ and $\{x_n\}$. That is, they could not connect the nulmerical construction to understand the limit process in concept-formation such as $\lim_{x \to a^+} f(x)$ $\lim_{x \to a^-} f(x)$, $\lim_{x \to a^+} f(x) = \lim_{x \to a^-} f(x)$

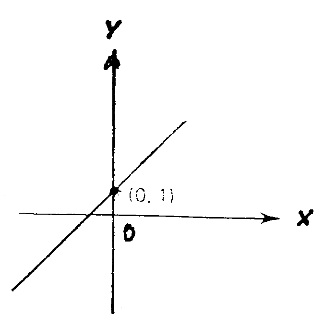
Especially they found limits by substituting x=a into f(x) and guessing the limit process in the figure.

In the figure, many studentsans wered that there was no limit because f(x) can not continue at x=1. Probably the approximating process should be introduced much before some formal representation.





(a) $f(x) = \frac{x(x-1)}{x}$; dom $f = R - 0$



(b) $g(x) = x + 1$; dom $g = R$

x	$f(x)=\frac{x(x+1)}{x}=x+1$	x	$f(x)=\frac{x(x+1)}{x}=x+1$
1	2	-1	0
0.5	1.5	-0.5	0.5
0.1	1.1	-0.1	0.9
0.05	1.05	-0.05	0.95
0.01	1.01	-0.01	0.99
0.001	1.001	-0.001	0.999

It in important to note that for $f(x)=\frac{x(x+1)}{x}$, it is still not permissible to set x=0 because this would imply division by zero. However, we now know what happens to this function as x approaches zero. We can see why it is important that we are not required to evaluate f(x) at x=0 when we calculate the limit as x approaches zero. The example could give some aid to overcome the misconceptions about the limits of discontinuous function at x.

(2) Derivatives

The most difficult problems in the area of the derivatives were some lack of understanding in the following concepts.

the rate of change including △x, △y

differnce between $\frac{f(3+h)-f(3)}{h}$ and $\lim_{h\to 0}\frac{f(3+h)-f(3)}{h}$

since they showed a serious lack of understanding of the rate problems in the physical situations, they could not connect f′(x) to the limiting process and answered f′(3)=f′(x).

The computer could also assist them to overcome those misconception by presenting some appropriate examples related to errors.

Rate of change : suppose that an object is dropped from rest from a given height. The distance s the object has dropped after t seconds(ignoring air resistance) is given by the formula $s=1/2gt^2$ where $g = 9.8\ m/sec^2 = 32ft/sec^2$ is the acceleration due to gravity. We now ask : What is the velocity of the object 2 seconds? To answer this question, it is first necessary to note that velocity can be regarded as a rate of change. Whether measured in meters per second, feet per second, or miles per hour, velocity is the ratio of change in distance(meters, feet, miles) to the change in time(seconds, hours).

Table gives(in column 3) the distance the object has fallen (in meters) after t seconds, tabulated every 0.2 second. Column 4 shows the distance the object has fallen in the previously elapsed o.2 second. Column 5 indicates the average velocity over each 0.2 second interval of time. We have average velocity

$$\text{average velocity} = \frac{\text{distance fallen}}{\text{elapsed time}} = \frac{\text{change in distance}}{\text{change in time}} = \frac{\triangle s}{\triangle t}$$

where in this problem $\triangle t = 0.2$ sec. From the table, we see that the average velocity of the object is constsntly increasing. What is the exact velocity at t = 2? That is, instead of an average velocity we want to know the velocity after precisely 2 seconds. Table will help answer this question. When t = 2, the object is falling faster than at any previous time. Let v(t) denote the velocity at time t. Since 18.62 m/sec is the average velocity for the time t = 1.8 sec to t = 2 sec,

t	t^2	$s=1/2gt^2 = 4.9t^2$(m)	$\triangle s$ = change in s since last measurement(m)	average velocity over last 0.2 sec, $\triangle s/\triangle t$(m/sec)
0.	0.	0.	-	-
0.2	0.04	0.196	0.196	0.98
0.4	0.16	0.784	0.588	2.94
0.6	0.36	1.764	0.98	4.9
0.8	0.64	3.136	1.372	6.86
1.0	1.0	4.9	1.764	8.82
1.2	1.44	7.056	2.156	10.78
1.4	1.96	9.604	2.548	12.74
1.6	2.56	12.544	2.94	14.7
1.8	3.24	15.876	3.332	16.66
2.0	4.0	19.6	3.724	18.62
2.2	4.84	23.716	4.116	20.58
2.4	5.76	28.224	4.508	22.54
2.6	6.76	33.124	4.9	24.5
2.8	7.84	38.416	5.292	26.46
3.0	9.0	44.1	5.684	28.42

we have

$$(\text{velocity at } t = 2) = v(2) > 18.62\text{m/sec}.$$

Similarly, at t = 2, the object is falling more slowly than when t > 2. Therefore

$$v(2) < 20.58\text{m/sec}.$$

which is the average velocity between the time t = 2 sec and t = 2.2 sec. we obtain the estimate

$$18.62 < v(2) < 20.58\text{m/sec}.$$

To narrow this down, we choose a smaller value for $\triangle t$. If $\triangle t = 0.05$ sec, we calculate the average velocity, $\triangle s/\triangle t$, for t between 1.95 and 2.0 sec and for t between 2.0 and 2.05 sec. Then v(2) will lie between these two values. We then have the new estimate

$$19.355 < v(2) < 19.845 \text{m/sec}$$

t	t^2	$s = 4.9t^2(m)$	$\triangle s(m)$	$\triangle s/\triangle t(m/sec)$
1.95	3.8025	18.63225	-	-
2.0	4.0	19.6	0.96775	19.355
2.05	4.2025	20.59225	0.9225	19.845

We continue this process by choosing an even smaller value for $\triangle t$, say $\triangle t = 0.001$.

Then we calculate the values in Table. Therefore

$$19.5951 < v(2) < 19.6049 \text{m/sec.}$$

t	t^2	$s = 4.9t^2(m)$	$\triangle s(m)$	$\triangle s/\triangle t(m/sec)$
1.999	3.996001	19.5804049	-	-
2.0	4.0	19.6	0.0195951	19.5951
2.001	4.004001	19.6196049	0.0196049	19.6049

As $\triangle t$ becomes smaller and smaller, we see that the average velocity $\triangle s/\triangle t$ gets closer and closer to the actual velocity at value $t = 2$.

Derivatives : The following limiting process can give excellent example to correct the misconceptions described above.

The function $f(h) = \frac{(3+h)^2-9}{h}$ $\lim_{h \to 0} \frac{(3+h)^2-9}{h}$

is not defined at $h = 0$ since at $h = 0$,

$$\frac{(3+0)^2}{0} = \frac{9-9}{0} = \frac{0}{0}$$

Which in an undefined expression. In table we can tabulate some values of f(h) for small values of h. It appears that $\lim_{h \to 0} f(h) = 6$ we can show this directly as follows :

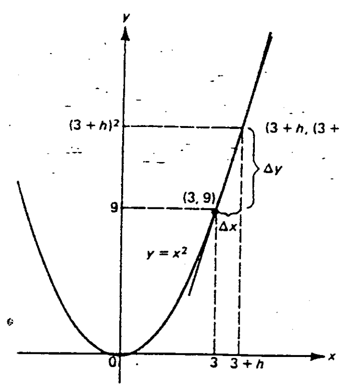
$$\lim_{h \to 0} \frac{(3+h)^2-9}{h} = \lim_{h \to 0} \frac{(9+6h+h^2)-9}{h} = \lim_{h \to 0} \frac{6h+h^2}{h}$$

h	$f(h) = \frac{(3+h)^2-9}{h}$	h	$f(h) = \frac{(3+h)^2-9}{h}$
1	1.75	-1	5
0.5	6.5	-0.5	5.5
0.1	6.1	-0.1	5.9
0.01	6.01	-0.01	5.99
0.001	6.001	-0.001	5.999
0.00001	6.00001	-0.00001	5.99999

Now, for $h \neq 0$, $\frac{h(6+h)}{h} = 6+h$, which is a function identical to f(h) except at $h = 0$. Thus $\lim_{h \to 0} (6+h) = 6$, hence $\lim_{h \to 0} \frac{(3+h)^2-9}{h} = 6$.

This is another example which illustrates the fact that $\lim_{x \to x_0} f(x)$ may exist even though $f(x_0)$ does not exist. Heve $f(0)$ is not defined while $\lim_{h \to 0} f(h)$ exists and is equal to 6.

There is an interesting geometrical interpretation to the result we have just obtained. In figure we have drawn the graph of the function $y = x^2$. When $x = 3$, $y = 3^2 = 9$. If h represents some small number, then $x = 3 + h$ is close to the value $x = 3$. Moreover, $f(3 + h) = y = (3 + h)^2$ is the value of y at $x = 3 + h$. Consider the straight line joining



the two points (3, 9) and $(3 + h, (3 + h)^2)$. We find that the slope of this line is given by

$$m = \frac{\Delta y}{\Delta x} = \frac{(3 + h)^2 - 9}{(3 + h) - 3} = \frac{(3 + h)^2 - 9}{h}.$$

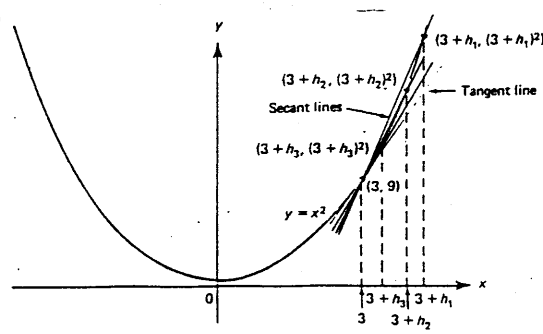
We called this last function f(h). The line joining the points(3, 9) and $(3 + h, (3 + h)^2)$ is called a secant line to the parabola $y = x^2$. Now a glance at figure indicates that as h approaches zero, $3 + h$ approaches 3 and the secant line approaches the line which is tangent to the curve $y = x^2$ at the point(3, 9). This line is called a tangent line. Since the secant lines are approaching the tangent line as $h \to 0$, it is resonable to assert that the slopes of the secant lines approach the slope of the tangent line as $h \to 0$.

But since f(h) represents the slope of the secant line joining the points (3, 9) and $(3 + h, (3 + h)^2)$, we have

slope of the tangent line to the curve $y = x^2$ at the point(3, 9) $= \lim_{h \to 0} f(h) = \lim_{h \to 0} \frac{(3 + h)^2 - 9}{h} = 6.$

Moreover, we can write the equation of the tangent line to the curve $y = x^2$ at the point (3, 9) :

$$\frac{y-9}{x-3} = 6 \quad \text{or} \quad y = 6x - 9.$$



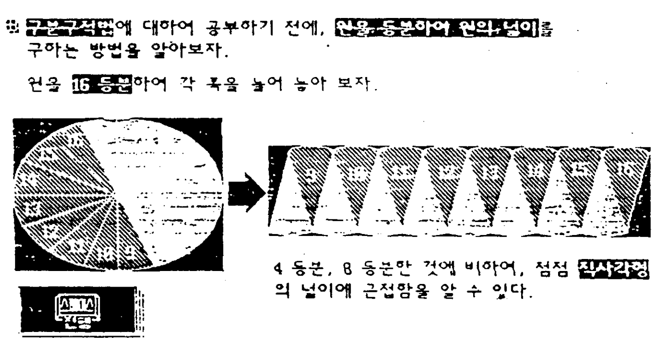
(3) Integration

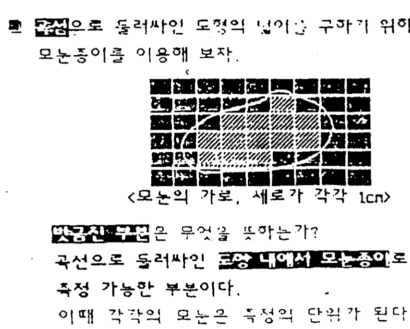
The students showed a serious lack of understanding of the following mathematical ideas.

Partitioning on [a, b], approximating area enclosed by the curve f(x) and the x-axis from x = a to x = b, connecting approximating area $\sum_{i=1}^{n} f(x_i^*)\Delta x$ to $\lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\Delta x$, $x_i^* \in [x_{i-1}, x_i]$

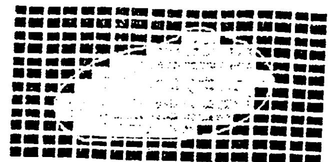
The smooth connection from approximating area $\sum_{i=1}^{n} f(x_i^*)\Delta x$ to $\lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\Delta x$ could be the most difficult problem in constructing the instructional sequence in the calculus. The computer also give some excellent assistence to correct those misconceptions.

The first one is related to partitioning problem shown above.





■ 이번에는 모눈의 가로, 세로의 간격을 계속 좁혀 나갈 경우, 모눈의 크기와 구하려는 도형의 넓이와의 관계를 생각해 보자



<모눈의 가로, 세로의 간격이 좁아짐>

모눈의 가로, 세로 간격이 점점 좁아지면,
도형 내에서 모눈으로 계산할 수 없는 넓이의 오차가 점점 줄어들어 "0"으로 접근할 것이다. ⊡

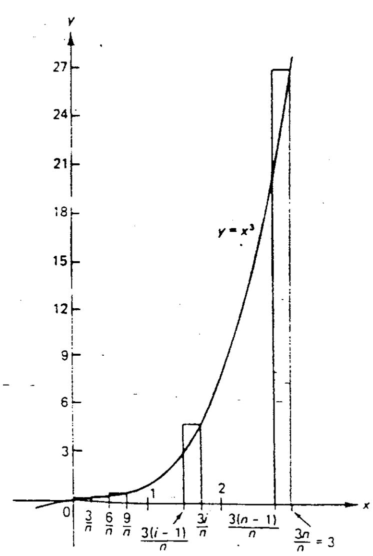
The other is related to approximating the area enclosed by the curve f(x) and the x-axis from x = a to x = b and connecting approximating area $\sum_{i=1}^{n} f(x_i^*)\Delta x$, to $\lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\Delta x,\ x_i^* \in [x_{i-1}, x_i]$. To approximate the area bounded by the curve $y = x^3$ and the x-axis from x = 0 to x = 3. We can divide the interval into n equal subintervals, each having length (b - a)/n = 3/n. The partition points are

$$0 = \frac{0}{n} < \frac{3}{n} < \frac{6}{n} < \frac{9}{n} < \cdots < \frac{3(n-1)}{n} < \frac{3n}{n} = 3.$$

For convenience, we choose $x^*_i = x_i$ (the right - hand endpoint of the subinterval), so that $f(x_i{}^*) = f(x_i) = x_i{}^3 = \left(\frac{3i}{n}\right)^3 = \frac{27i^3}{n^3}$ for $i = 1, 2, \ldots, n$.

Then $A_0^3 \approx \sum_{i=1}^{n} f(x_i)\,\Delta x = \sum_{i=1}^{n} \frac{27i^3}{n^3}\cdot\frac{3}{n} = \frac{81}{n^4}\sum_{i=1}^{n} i^3$.

Yes, there is a formula for the sum of the first n cubes.



we obtain $\quad A_0^3 \approx \frac{81}{n^4} \cdot \frac{n^2(n+1)^2}{4} = \frac{81}{4} \cdot \frac{(n+1)^2}{n^2}.$

Some of these approximations are tabulated in Table. It seems that as $n \to \infty$, $A_0^3 \to 20.25 = 81/4$. Indeed,

$$\lim_{n\to\infty} \frac{81}{4} \cdot \frac{(n+1)^2}{n^2} = \frac{81}{4} \lim_{n\to\infty} \frac{n^2+2n+1}{n^2} = \frac{81}{4} \lim_{n\to\infty}\left(1+\frac{2}{n}+\frac{1}{n^2}\right) = \frac{81}{4}.$$

n	n^2	n + 1	$(n+1)^2$	$(n+1)^2/n^2$	$81/4 \cdot (n+1)^2/n^2$
1	1	2	4	4.0	81.0
5	25	6	36	1.44	29.16
10	100	11	121	1.21	24.5025
100	10,000	101	10,201	1.0201	20.657025
1,000	1,000,000	1,001	1,002,001	1.002001	20.29052025
10,000	100,000,000	10,001	100,020,001	1.00020001	20.2540502

We will put these calculations on a more formal basis in the next definition and shall show how the techniques we have introduced here can also be used to solve problems that do not involve areas.

Let a < b and suppose that $\lim_{\Delta x\to 0} \sum_{i=1}^{n} f(x_i^*)\Delta x = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\frac{b-a}{n}$

exists and is independent of the way in which the number x_i^* are chosen. Then the integral of f(x) on the interval [a, b], written $\int_a^b f(x)\,dx$

$$\int_a^b f(x)\,dx = \lim_{\Delta x\to 0} \sum_{i=1}^{n} f(x_i^*)\Delta x = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\frac{b-a}{n}$$

If the limit exists, then f is said to be integrable on the interval [a, b] we can introduce the computer program to make them understand those concepts.

```
10  REM F(X)=X^3의 정적분
20  CLS
30  INPUT LX, RX
40  PRINT "범위 : ("; LX; ", "; RX; ")" : PRINT
50  RANGE=RX-LX
60  FOR I=1 TO 10
70    N=RANGE/(I*10)
80    SUM=0
90    X=LX
100   FOR J=1 TO I*10
110     X=X+N
120     SUM=SUM-N*X^3
130   NEXT J
140   PRINT " ", "등분수 : "; I*10, "정적분"; SUM
150 NEXT I
160 END
```

범 위 : [-1, 1]
등분수 : 10 정적분 .1999999
등분수 : 20 정적분 .1000002
등분수 : 30 정적분 6.666662E-02
등분수 : 40 정적분 5.000038E-02
등분수 : 50 정적분 4.000038E-02
등분수 : 60 정적분 3.333414E-02
등분수 : 70 정적분 2.857145E-02
등분수 : 80 정적분 2.499918E-02
등분수 : 90 정적분 2.222225E-02
등분수 : 100 정적분 1.999934E-02

범위 : [0, 1]
등분수 : 10 정적분 .3025001
등분수 : 20 정적분 .2756251
등분수 : 30 정적분 .2669447
등분수 : 40 정적분 .262656
등분수 : 50 정적분 .2600398
등분수 : 60 정적분 .2584027
등분수 : 70 정적분 .2571944
등분수 : 80 정적분 .2562886
등분수 : 90 정적분 .2555871
등분수 : 100 정적분 .2550246

REFERENCES

Confreg, I. what constructivism lmplies for teaching. In NCTM monograph 4 (Ed.), Constructivist view on the teaching and learning of mathematics. Virginia : NCTM, inc,.

Cornu, B. limits. In Tall D. 0. (Ed.), Advanced mathematical thinking, Kluwer, Dordrecht.

Freudenthal, H. didactical phenomenology of mathematical structures, Reidel, 1983.

Ginsburg, H. P. Academic assessment, In J. Valsiner(Ed.), The role of the individual sudject subject in scientific psychology. New York : plenum, 1990. pp. 235-260

Hart, K. M. Children's Understanding of mathematics : 11 - 16 London : John Murray, 1981.

Howson A. G. Perspectives on Math Education. Cambridge : cambridge university press, 1981.

Lesh, R. Conceptual Models and applied mathematical problem solving, Acquisition of mathematics concepts and processes, Academic pres, Inc. 1981.

Piaget, J. The child's conception of geometry, New York : Basic, 1960.

Resmick, L. B. The psychology of mathematics for instruction, New Jersey : Lea. 1981.

Shavelson, E. J. Method for examining representations of a subject matter structure in student's memory. Journal of Research in science Teaching. Vol. 11, 1974.

APPLICATIONS OF SPREADSHEETS IN CALCULUS

Robert S. Smith
Department of Mathematics & Statistics
Miami University
Oxford Ohio 45056-1641
U. S. A.
rssmith@muohio.edu

The electronic spreadsheet has gained increasing popularity in secondary and tertiary mathematics pedagogy [4], [5], [7], [8], [9], [10]. A number of factors have contributed to this acceptance by the mathematical community. The spreadsheet is readily available since it is frequently bundled with other software packages. Spreadsheets are quickly learned and are easy to use. Further, learning to use the ubiquitous spreadsheet is good career training for our students.

From a mathematical point of view, the spreadsheet is a natural tool for many applications. It is ideal for implementing algorithms which rely on iterative procedures [2], [6]. Analyses which can be adapted to a tabular format are particularly well-suited to the spreadsheet since it can yield instant updates in response to parameter changes. For example, it is natural to apply the spreadsheet to topics such as approximating zeros of functions, numerical differentiation, and determining the convergence of sequences. Further, with the powerful graphics of a modern spreadsheet, it is equally natural to use this tool as an electronic chalkboard for visualization.

Approximating Zeros of Functions

Approximating the zeros of a function is a natural application of spreadsheets in mathematics. Finding zeros of functions (or solving equations) is a fundamental application of mathematics which is spread across many disciplines. Finding zeros of functions can be a nontrivial activity for students—even if the function is a simple polynomial of degree $n \geq 3$. A fortiori, solving an equation such as $x^2 = x \sin x + \cos x$ can be positively daunting [11]. Yet, there are methods which can be implemented on a spreadsheet which will effortlessly solve a wide variety of equations. The first such method we will discuss is based upon Bolzano's Theorem, a forerunner of the Intermediate Value Theorem. Bernard Bolzano (1781–1848), a Czechoslovakian Catholic priest, made many important contributions to mathematics. He observed the following result and first published it in 1817 [1].

Bolzano's Theorem: *Let f be continuous on the interval $[a, b]$, and let $f(a)$ and $f(b)$ have opposite signs. Then there exists a number c in (a, b) such that $f(c) = 0$.*

Bolzano's theorem can be used to good advantage in finding zeros of a function by successively applying it to the function over smaller and smaller intervals. The procedure is as follows:

1. Write the equation in the form $f(x) = 0$.
2. Determine a closed interval over which the function $f(x)$ is continuous and changes sign.
3. Divide the interval into 10 subintervals of equal length.
4. Evaluate the function at the endpoints of these subintervals.

5. Determine a subinterval over which the function changes sign.
6. Divide the subinterval into 10 subintervals of equal length.
7. Return to step (4) and repeat the process until the absolute value of the difference of successive x values is less than or equal to some previously specified error bound.
8. Then any x value which is between the aforementioned successive endpoints will be an approximate zero of the function or an approximate solution to $f(x) = 0$.

This method of successive approximation by tenths [12] is easily implemented on a spreadsheet. Once the cells are properly assigned, the algorithm produces accurate results with not much effort.

Let us approximate a solution to $x^5 + 4 = 2x^2$ which is accurate to 10^{-3}. We rewrite the equation in the form $f(x) = x^5 - 2x^2 + 4 = 0$ and begin the approximation by tabulating $f(x)$ over an interval, say $[-5, 5]$, and observe its values. Then, we set up the spreadsheet so that we can implement the algorithm by changing entries in only two cells, F2 and F3 (see Figure 1). After several iterations, one obtains a spreadsheet solution to this problem. One such solution is -1.0975 (see Figure 2).

While the method of successive approximation by tenths is reliable, it is not efficient. A more efficient but less predictable method of approximating the zeros of a function is the Newton-Raphson Method. The general idea behind this method is as follows:

1. Write the equation in the form $f(x) = 0$, where $f(x)$ is a differentiable function.
2. Make a table of values of the function or a graph of the function so that an interval over which $f(x)$ has a zero can be identified.
3. Select a value, x_1, which is close to a zero. This will be the first approximation of a zero of $f(x)$.
4. Compute the second approximation of a zero of $f(x)$ using

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

5. Use the general iterative scheme

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \ n \geq 1$$

to compute $x_3, x_4, \ldots$

	A	B	C	D	E	F	G	H
1	Solve f(x) = x ^{5} - 2x ^{2} + 4 = 0.							
2	x	f(x)			a =	-5	error bound =	0.001
3	-5	-3171			h =	1	approx. zero=	NA
4	-4	-1052						
5	-3	-257						
6	-2	-36						
7	-1	1						
8	0	4	Try Again					
9	1	3						
10	2	28						
11	3	229						
12	4	996						
13	5	3079						
14								
15								
16								
17								
18								
19								

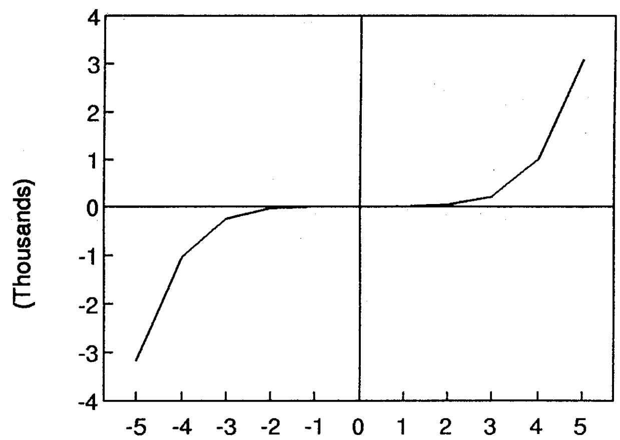


Figure 1.

	A	B	C	D	E	F	G	H
1	Solve f(x) = x ^{5} - 2x ^{2} + 4 = 0.							
2	x	f(x)			a =	-1.098	error bound =	0.001
3	-1.098	-0.00713			h =	0.0001	approx. zero=	-1.0975
4	-1.0979	-0.00596						
5	-1.0978	-0.00479						
6	-1.0977	-0.00363						
7	-1.0976	-0.00246						
8	-1.0975	-0.00130	OK					
9	-1.0974	-0.00013						
10	-1.0973	0.001023						
11	-1.0972	0.002187						
12	-1.0971	0.003350						
13	-1.097	0.004514						
14								
15								
16								
17								
18								
19								

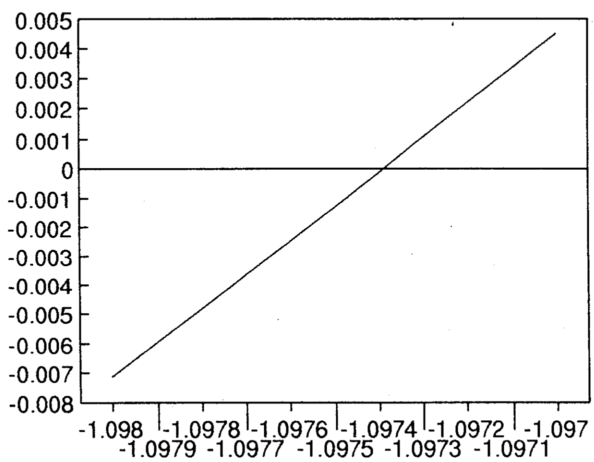


Figure 2.

Let us turn our attention to solving the equation $2\cos x = 2 - x$. We formulate the function $f(x) = 2\cos x + x - 2$ and construct a graph, or a table of values, over an interval, say $[-10, 10]$, to estimate a zero of the function. This activity will lead to the conclusion that $f(x)$ probably has three zeros in the interval $[-10, 10]$ and no others (see Figure 3).

Let us set up a spreadsheet with the iterate index in column A, the iterate x_n in column B, the function $f(x)$ in column C, and the function $f'(x)$ in column D. The initial approximation would be installed in cell B3. The remainder of row 3 will be linked to cell B3, and cell B4 will be linked to cells C3 and D3. This pattern is repeated throughout subsequent rows. The result would be a spreadsheet like Figure 4.

A point frequently made about the Newton-Raphson Method is that small changes in the initial estimate can make large changes in the approximation. This is easily and vividly demonstrated with this example. An initial estimate of 0.481 yields a zero approximation of 1.109144, an initial estimate of 0.482 yields a zero approximation of 0, an initial estimate of 0.483 yields a zero approximation of 3.698153, and an initial estimate of 0.485 yields chaos! These calculations are almost immediate with a spreadsheet.

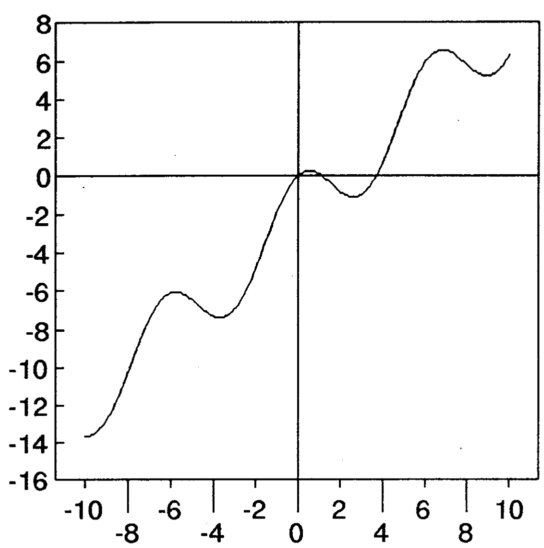


Figure 3.

	A	B	C	D
1	Solve f(x) = 2cos(x) + x - 2 = 0.			
2	n	x_n	$f(x_n)$	$f'(x_n)$
3	1	2	-0.83229	-0.81859
4	2	0.983265	0.091880	-0.66462
5	3	1.121510	-0.00984	-0.80151
6	4	1.109227	-0.00006	-0.79071
7	5	1.109144	-0.00000	-0.79063
8	6	1.109144	0	-0.79063
9	7	1.109144	0	-0.79063
10	8	1.109144	0	-0.79063
11	9	1.109144	0	-0.79063
12	10	1.109144	0	-0.79063
13	11	1.109144	0	-0.79063
14	12	1.109144	0	-0.79063
15	13	1.109144	0	-0.79063
16	14	1.109144	0	-0.79063
17	15	1.109144	0	-0.79063
18	16	1.109144	0	-0.79063
19	17	1.109144	0	-0.79063
20	18	1.109144	0	-0.79063

Figure 4.

Numerical Derivatives

While the spreadsheet is frequently thought of as a tool to analyze data which is in a tabular form, it is more versatile than these applications would suggest [10]. The spreadsheet can also be a very effective tool when used in class as an electronic chalk board, for example to illustrate the concept of the convergence of the difference quotient to the derivative. A discussion could begin by graphing the function $f(x)$ and its derivative $f'(x)$ on the same coordinate system. This would be followed by plotting the derivative function and the difference quotient function,

$$f_h(x) = \frac{f(x+h) - f(x)}{h},$$

for values of h approaching 0. Upon displaying the graph of $f'(x)$ and $f_h(x)$ students can immediately see the convergence of the difference quotient as $h \to 0$. Figure 5 shows a spreadsheet tabulation of the function

$$f(x) = -\frac{2x^3 + x^2 - 4x - 3}{10}$$

along with the functions f', f_2, $f_{1.3}$, and $f_{0.6}$.

	A	B	C	D	E	F
1	f(x) = -(2x³ + x² - 4x - 3)/10					
2			h =	2	1.3	0.6
3	x	f(x)	f'(x)	$f_2(x)$	$f_{1.3}(x)$	$f_{0.6}(x)$
4	-10	186.3	-57.6	-46.6	-50.268	-54.132
5	-9.9	180.5988	-56.426	-45.546	-49.172	-52.994
6	-9.8	175.0144	-55.264	-44.504	-48.088	-51.868
7	-9.7	169.5456	-54.114	-43.474	-47.016	-50.754
8	-9.6	164.1912	-52.976	-42.456	-45.956	-49.652
9	-9.5	158.95	-51.85	-41.45	-44.908	-48.562
10	-9.4	153.8208	-50.736	-40.456	-43.872	-47.484
11	-9.3	148.8024	-49.634	-39.474	-42.848	-46.418
12	-9.2	143.8936	-48.544	-38.504	-41.836	-45.364
13	-9.1	139.0932	-47.466	-37.546	-40.836	-44.322
14	-9	134.4	-46.4	-36.6	-39.848	-43.292
15	-8.9	129.8128	-45.346	-35.666	-38.872	-42.274
16	-8.8	125.3304	-44.304	-34.744	-37.908	-41.268
17	-8.7	120.9516	-43.274	-33.834	-36.956	-40.274
18	-8.6	116.6752	-42.256	-32.936	-36.016	-39.292
19	-8.5	112.5	-41.25	-32.05	-35.088	-38.322

Figure 5.

While this spreadsheet has no shortage of information, it may not be very comprehendible to the typical student. It is much more illuminating to study a graphical interpretation of these data, as given in Figure 6. In addition to analytical and graphical interpretations of these data, students can also experience an animated graphical interpretation. In this later representation, the function $f_h(x)$ can be made to visually converge f', as $h \to 0$, with the help of an easily written spreadsheet macro.

A similar approach can be taken to the chain rule, one of calculus's most enigmatic concepts. It is not uncommon for students to have a difficult time with this principle. Yet, a numerical approach can prove to be an effective tool in untangling the mysteries of this powerful and insidious rule [3]. The following example, using numerical derivatives and visualization, demonstrates this approach.

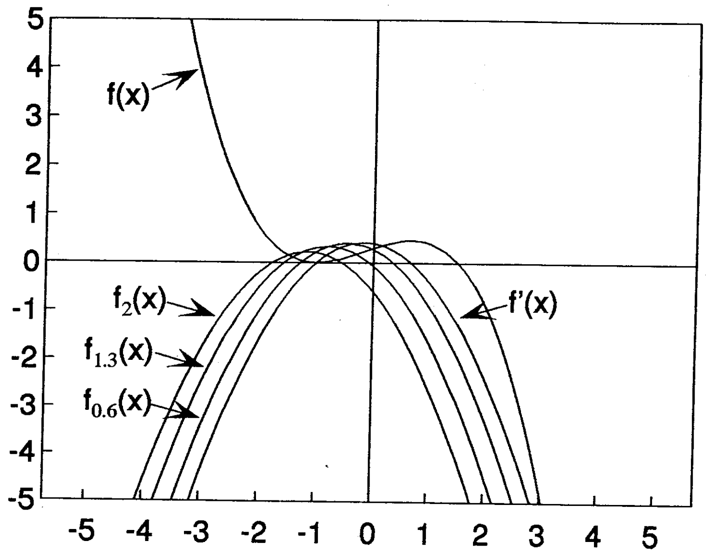


Figure 6.

Let us find the derivative of $y(u(x))$ where $u(x) = \sqrt{x^3}$ and $y(u) = \sin u$. We enumerate values of x in column A, $u(x)$ in column B, $y(u)$ in column C, the change in u with respect to x ($\triangle u/\triangle x$) in column D, the change in y with respect to u ($\triangle y/\triangle u$) in column E, the change in y with respect to x ($(\triangle u/\triangle x)(\triangle y/\triangle u) = (\triangle y/\triangle x)$) in column F, and $y'(u(x))u'(x)$ in column G. While the values in columns F and G agree quite favorably (see Figure 7) and help students to come to grips with the reality of the chain rule, a visualization of these data is even more impressive (see Figure 8). Note that $+$ represents $\triangle y/\triangle x$ and $\diamond$ represents the derivative $y'(u(x))u'(x)$.

	A	B	C	D	E	F	G
1	u(x)=x^{3/2}		y(u)=sin u	Change in	Change in	Change in	
2	x	u(x)	y(u)	u wrt x	y wrt u	y wrt x	y'(u(x))u'(x)
3	0	0	0	0.158113	0.999997	0.158113	0
4	0.025	0.003952	0.003952	0.289099	0.999969	0.289090	0.2371689
5	0.05	0.011180	0.011180	0.374370	0.999870	0.374321	0.3353892
6	0.075	0.020539	0.020538	0.443327	0.999654	0.443174	0.4107052
7	0.1	0.031622	0.031617	0.502855	0.999274	0.502491	0.4741044
8	0.125	0.044194	0.044179	0.556023	0.998684	0.555291	0.5298122
9	0.15	0.058094	0.058062	0.604520	0.997836	0.603212	0.5799674
10	0.175	0.073207	0.073142	0.649398	0.996683	0.647245	0.6258142
11	0.2	0.089442	0.089323	0.691366	0.995181	0.688034	0.6681389
12	0.225	0.106726	0.106524	0.730925	0.993281	0.726014	0.7074640
13	0.25	0.125	0.124674	0.768448	0.990939	0.761485	0.7441482
14	0.275	0.144211	0.143711	0.804222	0.988108	0.794658	0.7784413
15	0.3	0.164316	0.163578	0.838469	0.984743	0.825677	0.8105173
16	0.325	0.185278	0.184220	0.871371	0.980800	0.854641	0.8404959
17	0.35	0.207062	0.205586	0.903074	0.976235	0.881613	0.8684559
18	0.375	0.229639	0.227626	0.933701	0.971003	0.906627	0.8944450
19	0.4	0.252982	0.250292	0.963355	0.965062	0.929698	0.9184869

Figure 7.

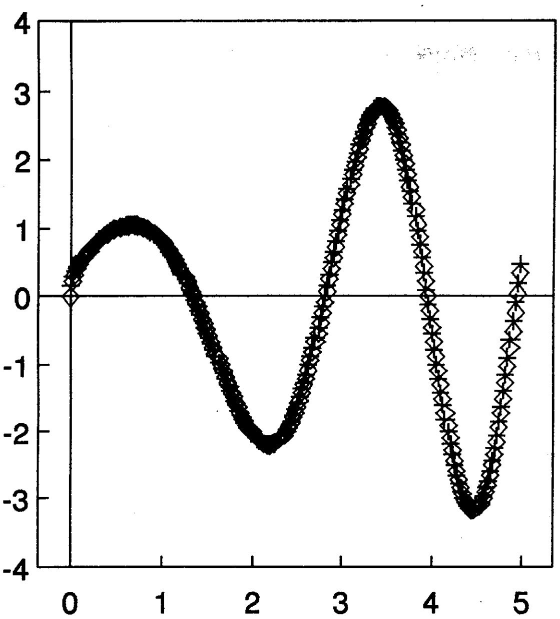


Figure 8.

Sequences and Series

Let us look at several applications of spreadsheets in a second calculus course. This first application is a very natural and predictable way of applying spreadsheets in calculus. Students seem to have little difficulty in solving problems of the form

$$\lim_{n \to \infty} a_n$$

where $\{a_n\}$ is a sequence given explicitly in a non recursive form. However, solving such problems when a_n is given recursively can cause considerable consternation. In addition to being interesting limit problems, such problems can lead to fascinating mathematical discoveries. The spreadsheet can be a great aid in both finding the solutions and making discoveries.

Consider the following problem. Compute the $\lim_{n \to \infty} a_n$ where

$$\begin{aligned} a_1 &= 2, \\ a_{n+1} &= \frac{1}{2}\left(a_n + \frac{2}{a_n}\right), \quad n \geq 1. \end{aligned}$$

What results if the value of a_1 is some other nonzero value?

A professorial view of this problem might be to prove the existence of the limit before trying to calculate it. On the other hand, a student using a spreadsheet would probably construct a spreadsheet giving the first 100 terms of the sequence and see if a limiting value is indicated (see Figure 9). In addition to the numerical analysis provided by the spreadsheet, the student could also make a graphical analysis by plotting data points of the form (n, a_n). It would not take long for the student to conclude that $\lim_{n \to \infty} a_n = \sqrt{2}$.

	A	B	C	D	E	F	G	H
1	n	a_n		$a_1 = 2$		$a_{100} = 1.414213$		
2	1	2		$a_{n+1} = (a_n + 2/a_n)/2$				
3	2	1.5						
4	3	1.416666						
5	4	1.414215						
6	5	1.414213						
7	6	1.414213						
8	7	1.414213						
9	8	1.414213						
10	9	1.414213						
11	10	1.414213						
12	11	1.414213						
13	12	1.414213						
14	13	1.414213						
15	14	1.414213						
16	15	1.414213						
17	16	1.414213						
18	17	1.414213						
19	18	1.414213						
20	19	1.414213						

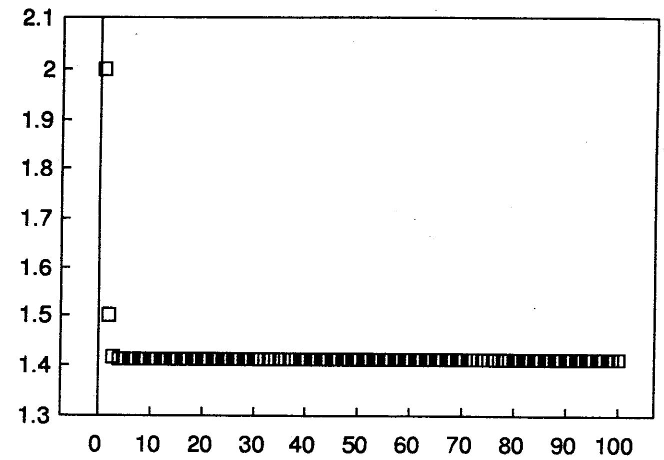


Figure 9.

The spreadsheet is an invaluable tool in addressing the more general question. Indeed, after trying a variety of positive and negative values for a_1, the student is apt to conjecture the following:

$$\lim_{n\to\infty} a_n = \begin{cases} \sqrt{2} & \text{for } a_1 > 0 \\ -\sqrt{2} & \text{for } a_1 < 0 \end{cases}$$

This conjecture is easily verified if one guides the student to consider the question, "For what real value x is the equation

$$x = \frac{1}{2}\left(x + \frac{2}{x}\right)$$

satisfied?" This would provide a nice stepping-off point to introduce the topic of fixed points.

Spreadsheets can be very useful in helping students understand convergence of series and estimating sums of series. It is easy to show that

$$\sum_{n=1}^{\infty} \frac{1}{n2^n} = \ln 2 = \sum_{n=1}^{\infty} \frac{(-1)^n}{n} .$$

However, the rates at which these series converge are quite different. The spreadsheet in Figure 10 provides analytical data which suggests that the left hand series converges more rapidly than the right hand series. Graphical representations of the left and right hand series are given in Figures 11 and 12, respectively. We use □ to represent the series terms a_n and b_n, and we use $+$ to represent partial sums A_n and B_n.

	A	B	C	D	E	F	G	H
1	n	a_n	A_n	b_n	B_n	a_n =	$1/(n2^n)$	
2	1	0.5	0.5	1	1	A_{1000} =	0.693147	
3	2	0.125	0.625	-0.5	0.5	b_n =	$(-1)^{(n+1)}/n$	
4	3	0.041666	0.666666	0.333333	0.833333	B_{1000} =	0.693646	
5	4	0.015625	0.682291	-0.25	0.583333			
6	5	0.00625	0.688541	0.2	0.783333			
7	6	0.002604	0.691145	-0.16666	0.616666			
8	7	0.001116	0.692261	0.142857	0.759523			
9	8	0.000488	0.692750	-0.125	0.634523			
10	9	0.000217	0.692967	0.111111	0.745634			
11	10	0.000097	0.693064	-0.1	0.645634			
12	11	0.000044	0.693109	0.090909	0.736544			
13	12	0.000020	0.693129	-0.08333	0.653210			

Figure 10.

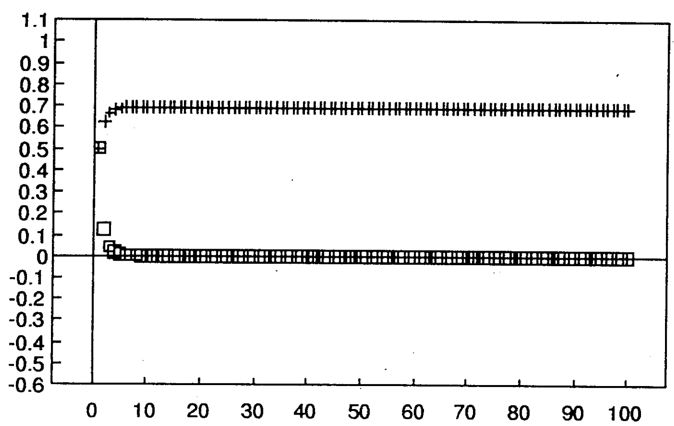


Figure 11.

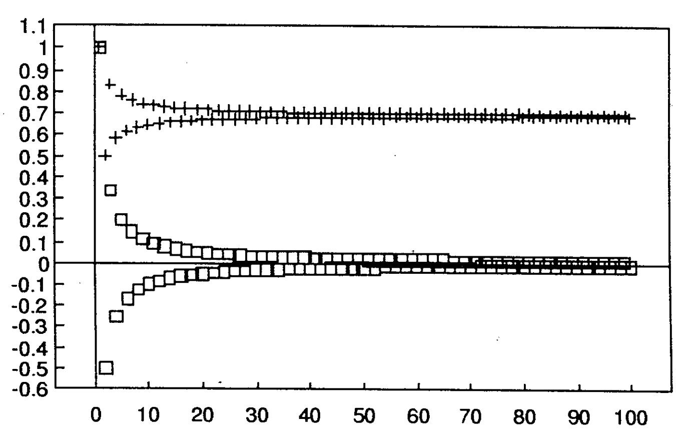


Figure 12.

Conclusion

The electronic spreadsheet is an easy-to-use and versatile pedagogical tool. Many students learn to use spreadsheets in courses outside of mathematics, and this is good career training. However, when students do mathematics with a spreadsheet they benefit in two ways—they enhance their mathematical experience and gain a dynamic new perspective on the uses and analytical powers of this software tool. With a variety of built-in mathematical and statistical functions, and excellent graphics, the spreadsheet is a powerful instrument for teaching and learning calculus.

References

[1] T. Apostle, Calculus, volume 1, 2nd ed., Blaisdell, Waltham, MA, 1967.

[2] D. Arganbright, Spreadsheet Curves and Geometric Constructions, CRC Press, Boca Ratan, FL, 1993.

[3] J. Kellett, Math 105 Class Notes, Gettysburg College, Gettysburg, PA, 1993.

[4] W. Masalski, How to use the spreadsheet as a tool in the secondary school mathematics classroom, National Council of Teachers Mathematics, Reston, VA, 1990.

[5] D. McLaren, Spreadsheets and Numerical Analysis, to appear.

[6] E. Neuwirth, Spreadsheet structures as a model for proving combinatorial identities, Journal of Technology in Mathematics and Science Education (to appear).

[7] J. Sandefur, Technology, linear equations, and buying a car, The Mathematics Teacher 85, 562–567, 1992.

[8] R. Smith, Teaching Precalculus with a Spreadsheet, Proceedings of the Fifth Annual International Conference on Technology in Collegiate Mathematics, Addison-Wesley, 262–269, 1992.

[9] R. Smith, Spreadsheets as a mathematical tool, Journal on Excellence in College Teaching 3, 131–148, 1992.

[10] R. Smith, Spreadsheets at Joint Meetings in Baltimore, UME Trends 4, 1–3, 1992.

[11] R. Smith, Rolle over Lagrange—Another Shot at the Mean Value Theorem, College Mathematics Journal 17, 403–406, 1986.

[12] E. Swokowski, Precalculus Functions and Graphs, 6th ed., PWS-Kent, Boston, MA, 1990.

The use of some of the Statistical Packages in fitting the Logistic Curve

by
Vasanth B. Solomon
Professor of Statistics
Radford University
Radford, Virginia

INTRODUCTION

Growth in many instances is rapid but only for a limited time. The demand for a consumer product is high but as time goes by the demand either drops or maintains a steady state. Human population or any other population for that matter could grow rapidly and it would seem that this growth is unbounded. If this was true, we would eventually run out of resources and space, which in turn could cause catastrophic illness or shortage of food or both. The natural consequence of this will be the destruction of a part of the population and thereby maintaining a delicate balance between our resources and the size of the population. Growth in most cases is not rapid. In fact, rapid growth cannot hold up indefinitely, eventually growth will slow down, fall or maintain a steady state. Almost in all cases, growth starts gradually, gathers momentum and then steadies itself. The shape associated with this kind of growth is the logistic curve (the S shaped curve).

The curve is of the form

$$y_t = \frac{k}{1+be^{-at}} \tag{1}$$

The curve is best suited in the study of Demography, Agriculture and Business and Technological Forecasting. Hence, there has been much literature into the study of the logistic curve and estimation of its parameters. Many of the methods are found in Nair (1954). Others are found in Nelder (1961), Oliver (1964) and Rasch (1988). The surprising fact is, in most papers is the absence of reference to Nair's paper in which he aludes to the methods of Hotelling, Fisher, Yule and his own which he calls as the 'New' method. Solomon (1992), in his paper refers to these methods and also to another method of estimation of the parameters of the logistic curve, which when fitted to historical data on the percentage of houses with electricity in the US, the R^2 is .9692, which seems to indicate a fairly good fit.

METHODOLOGY I

The method was based on the fact that the logistic curve is asymptotic at k as t tends to plus infinity. If k could be estimated, then estimates of a and b could be obtained by finding the regression of $\ln\left(\frac{k}{y} - 1\right)$ on t. The relationship between $\ln\left(\frac{k}{y} - 1\right)$ and t is of the form

$$\ln\left(\frac{k}{y} - 1)\right) = \ln b - at \qquad (2)$$

As an initial value of k, an educated guess is made knowing fully well that this value k_0 should not be less than the highest value of y. Using equation (3), we obtain the regression equation between $\ln\left(\frac{k_0}{y} - 1\right)$ and t. If the value of R^2 is high, we know that in the neighborhood of k_0 lies the value of k which will be a good estimate in that it gives a large R^2 value. To do so, we obtain two regression equations, one with $\ln\left(\frac{k_0+\delta}{y} - 1\right)$ and the other with $\ln\left(\frac{k_0-\delta}{y} - 1\right)$, $(\delta > 0)$ as the dependent variables and t as the the independent variable. If the R^2 value was to increase in one case, say for example with $k+\delta_0$ and decrease with $k-\delta_0$, then we do know that for increased values of k_0 we will get higher R^2 values. Obtaining many regression equations with increments to k_0, we will ultimately come to the value of k for which the R^2 value is a maximum. This value will be then used to obtain estimates of a and b. It is believed the equation obtained in this manner best fits the data.

METHODOLOGY II

This 'new' iterative procedure that is proposed today, is due to a relationship that is stated in Çambell (1993). "According to Verhulst the population growth depends on both the per capita resources needed to survive and the maximum population that can be sustained. This means that Malthus's equation must be adjusted to take into account the negative feed back that depends on the carrying capacity y_{max}, namely the maximum population that can be sustained ... for a fixed carrying capacity, the rate of growth will slow as the population increases. The following equation (3) results."

$$\frac{dy_t}{d_t} = ay_t\left(1 - \frac{y_t}{y_{max}}\right) \quad (3)$$

After integration, the population at any time t is

$$y_t = \frac{y_{max}}{1 + be^{-at}} \quad (4)$$

Thus as an initial value of k, we take the sum of $y_1 + y_n$ to be y_{max}. Regressing $\ln\left(\frac{y_{max}}{y} - 1\right)$ on t, we are then able to obtain initial estimates of a and b. With these values of k, a and b as starting values in Marquardt's* (1963) iteration procedure for non linear regression, we obtain the best estimates of a, b and k. (Marquardt's method is available in SAS as an option on Non Linear Regression.) The curve obtained in this manner is a better fit in that the sum of squares for error is least and consequently a larger R^2 value.

The above technique is demonstrated using Rasch's (1991) data on the growth of hemp plant. As the initial value of k we use 129.4 (8.3 + 121.1) and the values b and a to be 11.09039 and .419025 respectively obtained from regressing $\ln\left(\frac{129.4}{y} - 1\right)$ on t. At the end of five iterations, the convergence criterion is met and we obtain the final estimates of k, b and a to be 126.191, 12.449 and 0.4607 respectively. One would note that the method found in Rasch (1991) starts with initial values k= 125.21, b = 12.358 and a = .4657 and ends after five iterations with the same estimates k, b and a. The two methods are identical and Table I gives the observed, expected, error = observed − expected and the sum of squares for error. However, compared to many other methods, the proposed method is a better.

This technique was also applied to data on percentage of dwellings with electric power in the US. Since y_{max} cannot possibly exceed 100, we use 100 as the value for k and obtain estimates of a and b. Once again two methods are compared.

[*This is the iteration that is used by Rasch (1991) on the growth of hemp plants]

Method I was the method proposed in Solomon (1992) and Method II is the new iterative method. As before the observed, expected, error and sums of squares for error are presented. The new procedure is certainly the better method.

Table I

Age x	Plant height y (cm)	Method I and Method II * $y = \frac{126.19}{1+12.449\ e^{-.4607x}}$	
	Observed	Expected	Error
0	8.3	9.39	− 1.08
1	15.2	14.25	0.95
2	24.7	21.19	3.51
3	32.0	30.59	1.41
4	39.3	42.47	− 3.17
5	55.4	56.25	− .85
6	69.0	70.72	− 1.72
7	84.4	84.42	− .02
8	98.1	96.17	1.93
9	107.7	105.43	2.27
10	112.0	112.25	− .25
11	116.9	117.02	− .12
12	119.9	120.25	− .35
13	121.1	122.38	− 1.28
SUM OF SQUARES ERROR		40.7477	

*Please see Appendix I and III.

Table II

No.	Year	Percentage of of Dwellings with electricity	Method I $y = \frac{100}{1+31.5308\, e^{-.6382\, t}}$		Method II $y = \frac{100}{1+27.1361\, e^{-.6382\, t}}$	
			Expected	Error	Expected	Error
1	1905	3.7*	5.63	−1.93	6.52	− 2.82
2	1910	11.2*	10.09	1.11	11.67	− .47
3	1915	19.2*	17.43	1.77	20.00	− .80
4	1920	34.7	28.42	6.28	32.12	2.58
5	1925	53.2	42.75	10.45	47.26	5.94
6	1930	68.2	58.42	9.78	62.91	5.29
7	1935	68.0	72.55	− 4.55	76.25	− 8.25
8	1940	78.7	83.25	− 4.55	85.87	− 7.17
9	1945	85.0	90.34	− 5.34	92.00	− 7.00
10	1950	94.0	94.62	− .62	95.61	− 1.61
11	1955	98.4	97.07	1.33	97.63	0.77
SUM OF SQUARES ERROR			305.5308		250.5514	

1. Sources: Historical statistics of the United States, US Bureau of the Census, Washington, D.C.

2. *Estimated Values.

CONCLUSION:

In comparing the procedures that are available in fitting the logistic curve, the SAS program is easier than the SPSS program for the following reasons:

1) The mathematical model proposed in SAS is readily understood than the one proposed in SPSS.

2) To obtain initial values of the parameters of the logistic curve stated in SAS is much simpler than the one stated in SPSS.

As one should expect the two models are identical in the sense that the predicted values are the same and they are. However, my preference is to use the SAS program.

References

1. Balakrishnan, N. (1991). Handbook of the Logistic Distribution. New York: Marcel Dekker Inc.

2. Çambell, A. B. (1993), Applied Chaos Theory. New York: Academic Press Inc.

3. Marquardt, Donald W. (1963), "An Algorithm for Least Squares Estimation of Non Linear Parameters," Journal of SIAM, 11, 443−441.

4. Martino, Joseph P. (1975). Technological Forecasting for Decision Making. New York: American Elsevier Publishing Company Inc.

5. Nair, K. R. (1954). The Fitting of Growth Curves in Statistics and Mathematics in Biology (Eds) Kempthorne, Bancroft, Gowan and Lush, Ames, Iowa: Iowa State Press, 119-132

6. Nelder, J. A. (1961). The Fitting of a Generalization of the Logistic Curve, Biometrika, 17, 89

7. Oliver, F. R. (1964). Methods of Estimating the Logistic Growth Function, Journal of Applied Statistics, 13, 57−66

8. Rasch, D. (1988). Recent Results in Growth Analysis, Statistics, 19, 585−604

9. Rasch, Dieter (1991). Logistic Growth Models and Related Problems in Handbook of the Logistic Distribution. Balakrishuan, N. (ed.) New York: Marcel Decker Inc. (427−447).

10. Solomon, Vasanth B. (1992). An Exposure to Growth Curves − Especially the Logistic Curve. Paper presented at the joint meetings ORSA/TIMS in San Francisco.

```
options ls=80 ps=58;
data same;
infile 'hemp.dat';
input age 1-2 ht 3-6;
ht=ht/10;
proc nlin method=marquardt;
parms k=129.4 b=11.09039 a=.419025;
denom=(1+b*exp(-a*age));
model ht=k/denom;
der.k=1/denom;
der.b=-k*exp(-a*age)/denom**2;
der.a=k*b*age*exp(-a*age)/denom**2;
run;
```

Non-Linear Least Squares Iterative Phase

Dependent Variable HT   Method: Marquardt

Iter	K	B	A	Sum of Squares
0	129.400000	11.090390	0.419025	83.514032
1	125.401483	12.261983	0.460720	42.681568
2	126.211009	12.435871	0.460463	40.748782
3	126.189984	12.449078	0.460748	40.747676
4	126.191096	12.449019	0.460741	40.747674
5	126.191035	12.449039	0.460741	40.747674

NOTE: Convergence criterion met.

Non-Linear Least Squares Summary Statistics   Dependent Variable HT

Source	DF	Sum of Squares	Mean Square
Regression	3	94865.012326	31621.670775
Residual	11	40.747674	3.704334
Uncorrected Total	14	94905.760000	
(Corrected Total)	13	22904.617143	

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95 % Confidence Interval Lower	Upper
K	126.1910350	1.6630348663	122.53070592	129.85136400
B	12.4490390	0.8923553624	10.48497064	14.41310746
A	0.4607412	0.0163072714	0.42484905	0.49663345

Asymptotic Correlation Matrix

Corr	K	B	A
K	1	-0.405078847	-0.742954703
B	-0.405078847	1	0.8646614884
A	-0.742954703	0.8646614884	1

```
options ls=80 ps=58;
data same;
infile 'power.dat';
input yr 1-2 prct 3-5;
prct=prct/10;
proc nlin method=marquardt;
parms b=31.5308 a=.6382;
d=(1+b*exp(-a*yr));
model prct=100/d;
der.b=-100*exp(-a*yr)/d**2;
der.a=b*exp(-a*yr)/d**2;
run;
```

Non-Linear Least Squares Iterative Phase
Dependent Variable PRCT   Method: Marquardt

Iter	B	A	Sum of Squares
0	31.530800	0.638200	305.191174
1	27.118286	0.638200	250.552435
2	27.135024	0.638200	250.551387
3	27.136054	0.638200	250.551383
4	27.136117	0.638200	250.551383

NOTE: Convergence criterion met.

Non-Linear Least Squares Summary Statistics   Dependent Variable PRCT

Source	DF	Sum of Squares	Mean Square
Regression	1	45504.038617	45504.038617
Residual	10	250.551383	25.055138
Uncorrected Total	11	45754.590000	
(Corrected Total)	10	11448.727273	

NOTE: The Jacobian is singular.

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95 % Confidence Interval Lower	Upper
B	27.13611657	2.6676179359	21.192256017	33.079977115
A	0.63820000	0.0000000000	0.638200000	0.638200000

Asymptotic Correlation Matrix

Corr	B	A
B	1	.
A	.	.

APPENDIX III

```
Title Study of the use of SPSS  to fit a Logistic Curve
file handle input/name='hemp.dat'
data list file=input/age,ht(f2.0,f4.1)
model program A=3.6 B=-.419 C=465.84
compute pred=C/(A+Exp(A+B*age))
[c]NLR ht with age
```

There are 14 cases. There is enough memory for them all.

Iteration	Residual SS	A	B	C
1	77.83182687	3.60000000	-.41900000	465.840000
1.1	40.98175833	3.85854042	-.45919840	486.403465
2	40.98175833	3.85854042	-.45919840	486.403465
2.1	40.74782593	3.87599018	-.46064894	489.130237
3	40.74782593	3.87599018	-.46064894	489.130237
3.1	40.74767372	3.87659579	-.46074185	489.190835
4	40.74767372	3.87659579	-.46074185	489.190835
4.1	40.74767356	3.87660189	-.46074108	489.192463

Run stopped after 8 model evaluations and 4 derivative evaluations.
Iterations have been stopped because the relative reduction between successive residual sums of squares is at most SSCON = 1.000E-08

Nonlinear Regression Summary Statistics   Dependent Variable HT

Source	DF	Sum of Squares	Mean Square
Regression	3	94865.01233	31621.67078
Residual	11	40.74767	3.70433
Uncorrected Total	14	94905.76000	
(Corrected Total)	13	22904.61714	

R squared = 1 - Residual SS / Corrected SS =   .99822

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95 % Confidence Interval Lower	Upper
A	3.876601888	.096599264	3.663988342	4.089215433
B	-.460741083	.016307304	-.496633216	-.424848950
C	489.19246305	11.246739998	464.43855522	513.94637089

Asymptotic Correlation Matrix of the Parameter Estimates

	A	B	C
A	1.0000	-.8647	.8517
B	-.8647	1.0000	-.5113
C	.8517	-.5113	1.0000

Based on the SPSS output the fitted curve is of the form

Pred Ht = 489.1925/(3.8966+exp(3.8966-.4607*age))

WHAT LEIBNIZ MIGHT HAVE DONE ...
(to introduce calculus with a computer)

Walter G Spunde
Department of Mathematics and Computing
University of Southern Queensland
Toowoomba, Australia, 4350
E-mail: spunde@usq.edu.au

Abstract

Calculus means a *method of calculating*, but few students relate a formula such as $d(x^2)/dx = 2x$ to anything even remotely connected with the concept they have of calculation. Calculation, for them, deals with numbers; in calculus one manipulates symbols.

Formulae are often seen by students as character strings describing, or naming, a function, rather than as algorithms for computing function values. A formula such as $d(x^2)/dx = 2x$ implies that the *limit* of a difference quotient has been computed at any and every point, of an infinite number of points. If an approximate computation of these limits is done numerically, from first principles, at every point of any reasonable size sample of x values, it involves a massive amount of work and makes the simple formula look quite fantastic. The computing power available on a modern personal computer or programmable calculator, however, makes such calculations quite feasible.

In this paper we suggest that rather than deriving, from first principles, *formulae* for derivatives of the elementary functions (a task usually done only once or twice), students should instead compute approximate derivative values from first principles, and, indeed should do this as a matter of course, checking the results of symbolic manipulations by comparing function values obtained in this way with those obtained from first principles. The same direct approach can be applied to integration, where indefinite integrals in closed form cannot always be found. Finally we show how to demonstrate the Fundamental Theorem of the Calculus, differentiating integrals and integrating derivatives, for neither of which we have any formulae. The approach mirrors Leibniz's own early conception of the processes.

1 Introduction

Leibniz recalls that upon reading Pascal's brief paper, *Traité des sinus du quart de cercle*, a great light burst upon him. It occurred to him that the slope of the tangent line to a curve could be approximated by measuring the ratio of the change in the ordinates to the change in the abscissae at two neighbouring points of the curve. He saw that the area under the curve up to any given point was the cumulative sum of thin rectangles drawn to the height of the curve and it was obvious that the differences in a cumulative sum were just the terms of the sum, while the cumulative sum of pairwise differences would be the difference between the last and first terms. In short, the basics of differentiation and integration and the Fundamental Theorem of the Calculus were clear to him. In a letter to Johann Bernoulli of February 22, 1696, he wrote([3]),

> Accordingly I wish to get your opinion, whether you approve of marking by the sign $\int$ the sum, just as the sign d is displayed for differences; ...

In modern parlance, the $\int$ and the d are being used as symbols for functions: $\int y \cdot d\ x$ is the sum of the products of y and differences in x. Leibniz regarded differentials as exceedingly small, but finite, differences, and the d symbol was used for both finite and *infinitesimal* differences. He writes ([4]),

> Supposing that $1+1+1+1+1+$ *etc.* $= x$ or that x represents the natural numbers, for which $dx = 1$, then $1+3+6+10+$ *etc.* $= \int x, \ldots$

In describing x as the string of natural numbers, he clearly saw it as a vector, the cumulative sum of its differences, ie. $\int d\ x = x$, where $\int x$ (just as interpretable as $\int x \cdot dx$) is a vector of cumulative sums, and $d\ x$ is a vector of differences. It seems he has confused the sequences and the sums in the above extract, and surely meant:

$$1+1+1+1+1+etc = x \quad or \quad 1\ 2\ 3\ 4\ 5\ etc = x$$
$$1+2+3+4+5+etc = \int x \quad or \quad 1\ 3\ 6\ 10\ 15\ etc = \int x$$

from which we see not only that $dx = 1$, but also that $d \int x = x$. Both $\int$ and d apply to vectors and return vectors.

Leibniz did not have the conceptual structures available to describe his ideas adequately, but the computational aspects of his original vision were also beyond the capabilities of the age. Progress came from the development of symbolic manipulations, but it is the computational aspect of two vector operators that we will build on here. Like Leibniz ([7]), in whose day functional relationships were given by algebraic expressions and curves, *mechanical* curves, like the cycloid, and tables of logarithms and sines, we would like to stress that

> ...our method also covers transcendental curves –those that cannot be reduced by algebraic computation, or have no particular degree–and thus holds in the most general way ...

For today's students, it seems functional relationships are only expressed by algebraic formulae, involving perhaps transcendental functions, but not processes (mechanical or mathematical). We will follow Leibniz's advice ([7]),

> ...to keep in mind that to find a *tangent* means to draw a line that connects two points of the curve at an infinitely small distance, or the continued side of a polygon with an infinite number of angles, which for us takes the place of the *curve*.

for, in fact, that is how computer graphing programs represent curves (interpreting *infinite* as *sufficiently many*) and how we, using the slopes of *all* the polygon's sides will represent the derivative.

2 Samples, Derivatives and Integrals

Ironically the logical foundations of the calculus produce definitions of continuity and differentiability *at a point*, not because these properties are usually observed at a point, but because they may fail at a point. Standard texts invariably concentrate attention on the derivative at a single point (which admittedly could be any point) and the limit of the difference quotient at a single point. However, the functions dealt with in introductory calculus courses are all smooth at most points, and it is likely that there will be an increment that is sufficiently small for the difference quotient at any point of a given interval to provide an adequate approximation to the limiting value. In short, if we assume a domain, $[a, b]$, of uniform differentiability, given a sufficiently large number of equally spaced sub-division points, each interval width will be less than the required increment and it will be possible to compute approximate derivatives by dividing the increments in the function values by the x increment. The sections on numerical computation of the integral in standard texts deal exclusively with the definite integral, presumeably because there is sufficient work involved in doing just one integral. Maintaining a cumulative sum of the area calculation is a simple matter with a computer and restores the functional aspect to the results of the computation.

To perform our computations we assume that our computer software provides several vector functions, the most basic of which is a function, called say, *sample*, that delivers a vector of equally-spaced points for any given interval, as for example, in

$$x = sample(0,\ 1,\ 100)$$

0 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.1 0.11 ...

We will not, however, assume that our sample of x values is necessarily a uniform discretization of the interval. We may have finer (possibly uniform) samples in various sub-intervals of the domain. As indicated above we shall operate on this vector, x, as an entity and compute the corresponding function values, y, for any defined function, f, with $y = f(x)$ as for example in

$$y = \cos(x)$$

1 0.99995 0.9998 0.99955 0.9992001 0.9987503 0.9982005 0.997551 ...

Most particularly, we will need the two functions d and S to compute the differences between successive pairs in a string and to return the cumulative sum of entries in a vector, as in

$$dx = d(x)$$

0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 ...

$$S(dx)$$

0 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.1 0.11 ...

We have used the symbol S to indicate that this is an executable function— $\int$ has developed the connotations of a theoretical symbol. We also assume that the arithmetic functions will operate on vectors in a component-wise fashion and that we have available a function to plot sample points, (x, y), of any graph that might have been computed, with a statement like *plot y vs x*.

Chord slopes may be computed with

$$Dy = dy \div dx$$

$$-0.00499996 \quad -0.01499938 \quad -0.02499729 \quad -0.03499271 \quad -0.04498463 \ \ldots$$

where $dy = d(y)$. These slopes are better approximations to the derivative at the mid-points of successive pairs of the x sample than at the sample points themselves (the central difference approximation) and function, *mids*, delivers us these values, with $xm = mids(x)$ as for example in

$$-sin(mids(x))$$

$$-0.004999979 \quad -0.01499944 \quad -0.0249974 \quad -0.03499285 \quad -0.04498481 \ \ldots$$

Function *mids* also proves to be useful in computing the integral of a function (using the trapezoidal approximation) with

$$S\ (mids(y) \times dx)$$

$$0\ 0.00999975\ 0.0199985\ 0.02999525\ 0.0399890\ 0.04997875\ 0.05996351\ \ldots$$

cf. $sin(x)$

$$0\ 0.00999983\ 0.0199987\ 0.02999550\ 0.0399893\ 0.04997917\ 0.05996401\ \ldots$$

Thus we are in a position to graph both the derivative and integral of a function, without having any formula for these functions.

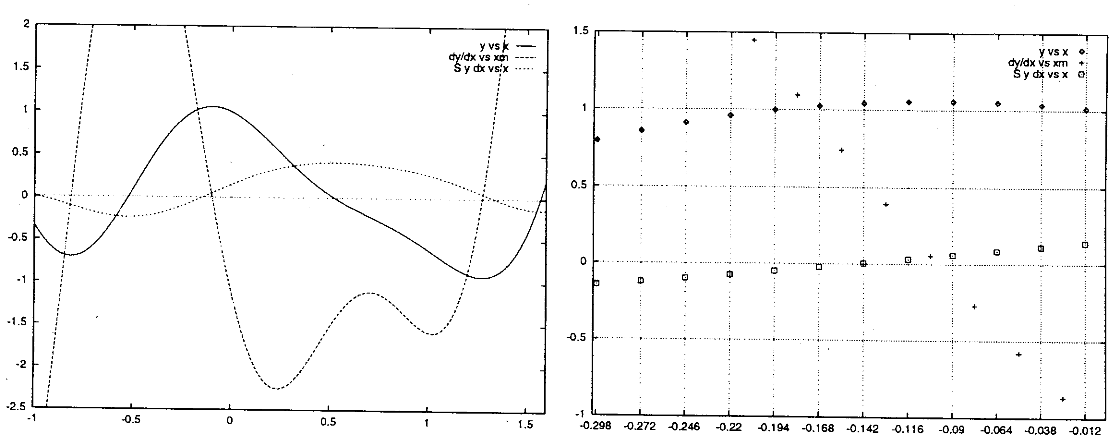


Figure 1: *Sample graph of function, derivative and integral with local detail*

Work with rapidly oscillating functions and functions undefined at odd points (like $\tan x$) can be used to convince students that one can never be sure that a sample is adequate unless there is some theoretical reason for believing this to be true. The requirement for students to produce a sample of x values on which to base the computations of derivative and integral, focusses attention on the domain of the function.

$$F(x) = \arcsin\left(\frac{1}{1+x^2} - \cos(x)\right)$$

for example, is defined nearly everywhere. Students may be surprised that some sample choices result in DOMAIN ERROR messages, while others do not. From a quick plot it appears to happen only in two regions near $-\pi$ and π:

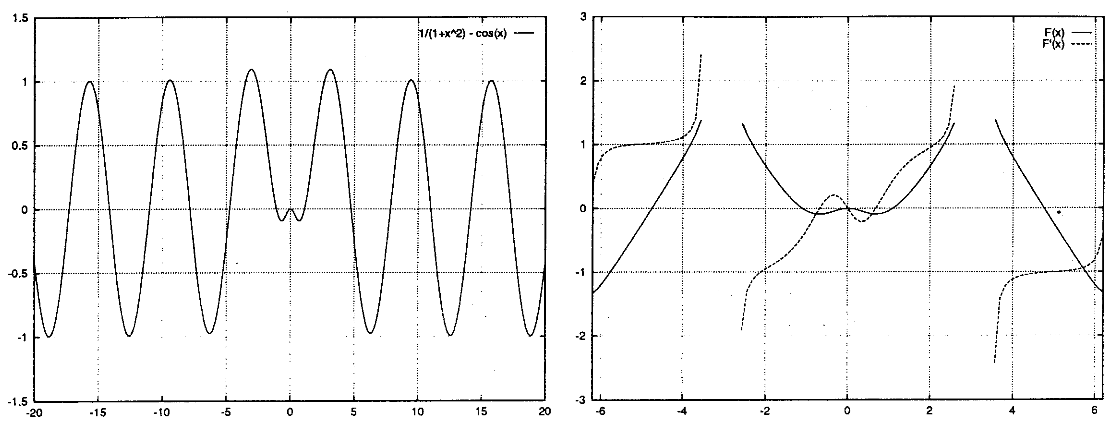


Figure 2: *Determining the domain of and plotting $F(x)$ and $F'(x)$*

However a little reflection shows that there must be a region, of increasingly small width, as x increases, around every value of x where $\cos(x) = -1$. Here samples must be drawn from disjoint intervals and dealt with separately, or nonsensical results will be obtained. Determining the domain of a function is seen to be of practical significance in a way seldom suggested by a symbolic formula.

3 The Rules of Calculus

With the rules of differentiation established theoretically, students, who have the tools of numerical calculus available to them, have several ways of computing derivatives of the usual arithmetic combinations of functions given in standard texts. The availability of computer algebra systems makes the requirement to do the symbolic manipulations required by hand rather artificial. Certainly some simple exercises will help in establishing an understanding of what is involved, but organizing a computation to apply the rules numerically to function samples is an alternative way of becoming familiar with the rules.

3.1 Arithmetic Combinations of Elementary Functions

We take as an example the problem of differentiating the function

$$f(x) = \left(2 - 3x^3 + x^4 - \frac{\sin(x)}{x + \cos(x)}\right) e^{-0.25x}$$

A computer algebra package may deliver the result

$$f'(x) = \left(-\frac{1}{2} - 9x^2 + \frac{19}{4}x^3 - \frac{1}{4}x^4 - \frac{\sin(x) - \cos(x)}{4(x + \cos(x))} + \frac{\sin(x)\,(1 - \sin(x))}{(x + \cos(x))^2}\right) e^{-0.25x}$$

There is little virtue in being able to produce this, but it is apparent that, unless we restrict oursleves to very simple functions, the algebraic manipulations can become so involved that they, rather than the rules of calculus take up the major part of the effort. Without a lot of practice on this sort of problem, however, the student will not appreciate what is involved in producing such a result.

With a numeric function sample approach, attention is shifted to organizing the computation. To compute the function values for a given sample of points, x, it may be convenient to compute the values of the component functions:

$x = sample(-1, 3.5, 100)$

$-1 \quad -0.955 \quad -0.91 \quad -0.865 \quad -0.82 \quad -0.775 \quad -0.73 \quad -0.685 \ \ldots$

$p = 2 - 3x^3 + x^4$

6 5.44474 4.94646 4.50148 4.10623 3.7572 3.45103 3.18443 ...

$c = \cos(x)$

0.5403023 0.5776088 0.6137457 0.6486401 0.6822212 0.714421 0.7451744 ...

$s = \sin(x)$

$-0.841471 \quad -0.8163137 \quad -0.7895037 \quad -0.7610953 \quad -0.7311458 \quad -0.6997161 \ \ldots$

$e = \exp(-0.25 \times x)$

1.28403 1.26966 1.25546 1.24141 1.22753 1.21379 1.20021 ...

The function values can be then be computed simply from

$y = (p - s \div (x + c)) \times e$

$5.35376 \quad 4.16664 \quad 2.86434 \quad 1.22125 \quad -1.47357 \quad -9.45942 \quad 56.8878 \ \ldots$

To differentiate the function, however, this reduction to components in terms of elementary functions is *essential*, because only for these functions is there a theoretically derived rule for differentiation. Applying the rules for differentiating the elementary functions, we may compute the values of the derivative functions at the sample points (suppressing the output):

$$\begin{aligned} Dp &= -18 \times x + 14.25 \times x^2 \\ Ds &= c \\ Dc &= -s \\ De &= -0.25 \times e \end{aligned}$$

To effect the differentiation, we analyse the function into its arithmetic components.

$$y = g \times e \quad ; \quad g = p - h \quad ; \quad h = s \div k \quad ; \quad k = x + c$$

These three statements must be executed in reverse order: k, then h then g (and then y if it has not already be done). Applying the rules for differentiation we have

$$Dy = Dg \times e + g \times De \quad ; \quad Dg = Dp - Dh \quad ; \quad Dh = (Ds \times k - s \times Dk) \div k^2 \quad ; \quad Dk = 1 + Dc$$

Again the statements are executed in reverse order, computing first Dk, then Dh, then Dg and finally Dy. As in the case of the y computation, if one is adept at analysing the structure one can by-pass the intermediate functions.

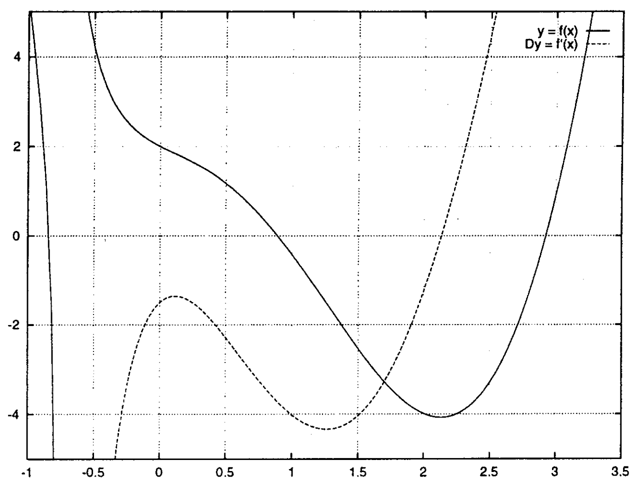


Figure 3: *Function $f(x)$ and its derivative (SIX graphs)*

Note that the analysis of the way function $f(x)$ is composed is exactly the task required of the student when (s)he is asked to do the differentiation *by hand*, but the computational sequence does not require that intermediate results be *substituted in* to create the monolythic result that the symbolic formula eventually produces. Thus it successfully separates the tasks of knowing and implementing the rules of calculus from those of knowing and implementing the rules of algebraic manipulation.

The values in the function table Dy *vs* x should be exactly the same as that obtained from the symbolic differentiation by tabulating $f'(x)$ *vs* x, and plots of each should be identical, since the only approximations made are those done internally by the machine. There is, of course, the ultimate check of computing the derivative from first principles and plotting

$$((d(y) \div d(x)) \; vs \; mids(x)$$

Where before students had no way of checking the results of their efforts in symbolic manipulation, they now have two: computation from first principles and, what, in another setting ([1, 5, 7]), might be called, *automatic differentiation* or, following Rall ([6]), *differentiation arithmetic.*

3.2 Composite Functions

Functions of functions expose students to the dreaded *Chain Rule* for differentiation. Seen from the perspective of differences, the rule appears trivial. For the composite function $f(g(x))$ we have, for example, writing f and g for the values of the functions f and g, to avoid additional symbols.

x	g	f	dx	dg	df	xm
1.000	12.376	-2.843				
			0.001	0.002	0.004	1.0005
1.001	12.378	-2.839				
			0.002	0.001	0.003	1.0015
...	...	...	...	...	...	...

Clearly

$$\frac{df}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx}$$

but the subtle point to note is that when we neglect the first column and regard f as a function of g, then the difference quotients df/dg approximate the derivatives at the points $mids(g)$. Expressing this symbolically:

$$\frac{df}{dx}(xm) = \frac{df}{dg}(gm) \cdot \frac{dg}{dx}(xm)$$

If function g is well enough behaved, gm may be regarded as an approximation to $g(xm)$ but a limit argument is required to deduce the chain rule.

$$\frac{df}{dx}(x) = \frac{df}{dg}(g(x)) \cdot \frac{dg}{dx}(x)$$

Using the table of values seems to be more effective in explaining to students the significance of the first term of the product on the right than the algebraic argument involved in the proof of the chain rule.

The chain rule can be implemented on sample values. Consider the differentiation of the function

$$y = \sin^3(x^2 - 3)$$

Taking a sample of values x from the domain, we must compute in order the values of

$$g = x^2 - 3 \quad ; \quad h = \sin(g) \quad ; \quad y = h^3$$

and then

$$dydh = 3 \times h^2 \quad ; \quad dhdg = \cos(g) \quad ; \quad dgdx = 2 \times x \quad ; \quad dydx = dydh \times dhdg \times dgdx$$

One can now plot y *vs* x and $dydx$ *vs* x if the sample is adequate.

Again, the analysis of the functional dependence is the task at hand and not the substitution of the various parts into some final formula in x: $y' = 6x\cos(x^2 - 3)\sin^2(x^2 - 3)$.

3.3 The Fundamental Theorem

Neither in computing the derivatives nor the integrals from first principles was a formula obtained for the resulting function. Being accustomed to dealing only with algebraic expressions in calculus, one might not envisage the prospect of applying the operations of the calculus to these tabulated functions, but both d and S are defined for vectors, and since we are not dealing with experimenatal data, but indeed functions for which it it is reasonable to assume that there is an *adequate* sample, we may compute

$$dx = d(x) \; ; \; xm = mids(x) \; ; \; dy = d(y) \; ; \; ym = mids(y)$$

$$Dy = dy \div dx \quad - \quad \text{derivative values at } xm$$

$$Sy = S(ym \times dx) \quad - \quad \text{integral values at } x$$

$$SDy = S(mids(Dy) \times d(xm) \quad - \quad \text{integral of derivative at } xm$$

$$DSy = d(Sy) \div dx \quad - \quad \text{derivative of integral at } xm$$

Note that both SDy and DSy should be plotted against xm and that they will not usually be the same. For example, for the function f above, the first few values of a 200 point sample of $[-1, 3.5]$ gave the values:

xm	-0.98880	-0.96625	-0.94375	-0.92125	-0.89875	-0.87625
DSy	5.06034	4.46678	3.85294	3.20179	2.48708	1.66553
SDy	0.00000	-0.59356	-1.20740	-1.85856	-2.57326	-3.39481
$SDy - DSy$	5.06034	5.06034	5.06034	5.06034	5.06034	5.06034
ym	5.06034	4.46678	3.85294	3.20179	2.48708	1.66553

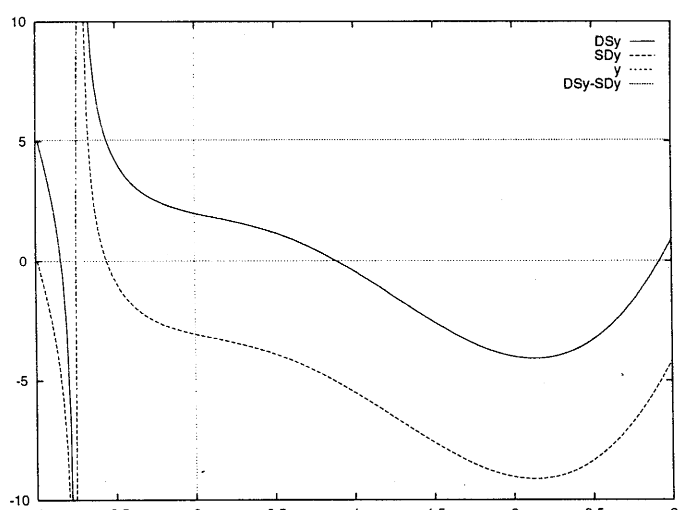


Figure 4: *The derivative of the integral & the integral of the derivative*

Students could be encouraged to speculate on the effect on the arbitrary constant of integration of a singularity of the function in the range of integration. The joy in seeing the graphs of DSy, SDy and y running parallel or coinciding lies in the knowledge of how each set of values was computed. It is certainly worthy of a theorem.

4 Conclusion

The use of function samples provides an alternative approach to the fundamental operations of the calculus, one that can be more easily understood by students who are insecure about their skills of algebraic manipulation and for whom the classical functional notation presents some difficulties. The technique allows the instructor to examine in meaningful detail the significance of the domain and range of a function and to provide an alternative way to implement the rules of the calculus. It provides a check on symbolic differentiation, and security for the student in the knowledge that (s)he can plot the integral of any continuous function whether (s)he is proficient in the techniques of finding anti-derivatives or not. If desired, the approach may be used to motivate, and illustrate $\epsilon - \delta$ computations for limits. It exposes students to simple vector operations and points the way to infinite dimensional function spaces.

Above all, it demonstrates the manipulation of functions whose values have been computed by a process which is not necessarily expressible in terms of an algebraic formula involving the elementary functions. Thus attention is focusses on the abstract definition of a function as a *rule whereby for any x in a certain domain, there is a unique y associated with it.*

References

[1] Griewank,A. & G. Corliss (eds.) *Automatic Differentiation of Algorithms: Theory, Implementation & Application.* SIAM, Philadelphia (1991)

[2] Leibniz,G.W. *Acta Eruditorum*,(1684) quoted in D.J.Struik, *A Source Book in Mathematics 1200-1800*, p.276, Harvard U.P., Harvard,(1969).

[3] *Leibnizens Math. Schriften*, Vol.III,(1855) p.262 quoted in Florian Cajori, *A History of Mathematical Notations*, Vol.II, p.182, Open Court Publishing, Chicago, (1929).

[4] *Leibnizens Math. Schriften*, Vol.V,(1858) p.397 quoted in Florian Cajori, *A History of Mathematical Notations*, Vol.II, p.264, Open Court Publishing, Chicago, (1929).

[5] Neidinger,R.D. Automatic Differentiation and APL *College Mathematics Journal* 20(3), 238-251, (1989)

[6] Rall,L.B. The Arithmetic of Differentiation *Mathematics Magazine* 59(5),275-282, (1986).

[7] Spunde,W.G. Point-wise Calculus, *APL Quote Quad* 24(1) 259-266,(1993).

ON THE EFFICIENT USE OF COMPUTER ALGEBRA IN TEACHING MATHEMATICS

Tadashi Takahashi*, Hirokazu Hashiba**

*Division of Mathematics and Informatics
Faculty of Human Development, Kobe University
3-11 Tsurukabuto, Nada-ku, Kobe 657, JAPAN
e-mail: takahasi@kobe-u.ac.jp
**Graduate School of Education, Kobe University
3-11 Tsurukabuto, Nada-ku, Kobe 657, JAPAN
e-mail: hashiba@kobe-u.ac.jp

Abstract

Progress in our highly information-oriented society is accelerating every day. Great importance is attached to education, particularly teaching mathematics, to meet future needs. It is 'efficient use of computer algebra' in education. It gives the traditional education the innovative and qualitative changes which include the revision of educational materials and enhancement of lecture style. Several approach by using cpmputer algebra is necessary to give the well-selected exercises which make the students confirm the materials in various ways. To make the use of computer algebra in this style enoguh efficient, the further research on teaching materials is expected. In this paper, we will consider the following three subjects for the efficient use of computer algebra in teaching mathematics:

- Getting into Workstation.
- Computer communication.

- Development of teaching materials by using anonymous ftp.

1 EFFICIENT USE OF COMPUTER IN EDUCATION

Progress in our highly information-oriented society is accelerating every day. Great importance is attached to education, particularly teaching mathematics, to meet future needs. It is 'efficient use of computer algebra' in education. In teaching mathematis, there are two important aims:

- Understanding the mathematical knowledges and concepts.
- Understanding the mathematical thoughts for problem solving process.

We must consider 'efficient use of computer algebra in teaching mathematics' on the above two aims. The most attention is paid to keep the students' attitude of positive participation, and to support the active study. Needless to say, the educational activity is not a routine.

Efficient use of computer algebra gives the traditional education the innovative and qualitative changes which include the revision of educational materials and/or enhancement of lecture style.

2 STYLES OF THE COMPUTER USES

The styles of computer users, a rough design of class is able to be drawn as follows.
(I) Situation leaded by teacher
(II) Situation leaded by student

In the near future, the computer softwares for the educational use can be classified into these categories. (I) is the situation in which computer is used as the presentation tool for teachers. For example, teahers will give graphs of functions or various curves as suitable materials in this situation. It is necessary to research how to give the educational materials.

In the situation(II), the students utilize the computers as the tools for their thinking. They are able to inquire mathematical facts and find them. The use of computer algebra systems is enough efficeent in these situations.

In general, certain application software packages are usable as the tools for the students' thinking. However, it must not be ignored that these programing languages or application software are not designed to aim at the use of mathematical education. They are designed for the education of computer science, software development, or use in an office.

To make the use of computer algebra systems in these situations enough efficient, the futher resarch on educational user-interface is expected.

3 EFFICIENT USE OF COMPUTER ALGEBRA SYSTEMS

The use of computer algebra systems is the test plans to design the class. It is based on the principle which says that the desirable change of mathematical education is from traditional one-side lecture with a blackboard to students' activity supported by computer algebra systems. Several approach by using cpmputer algebra systems is necessary to give the well-selected exercises which make the students confirm the materials in various ways. We will consider the following three subjects for the efficient use of computer algebra systems in teaching mathematics.
Getting into Workstation
Computer communication
Development of teaching materials by using anonymous ftp

3.1 GETTING INTO WORKSTATION

There are few computers in the classrooms of Japanese schools on which one can run can run computer algebra systems efficientry. So enabling students to access the workstations via a computer network would provide them with a better opotunity to use computer algebra systems efficientery.

we will show examples in which Risa/Asir was used.

Risa/Asir(Institute for Social Information Science, FUJITSU LABORATORIES LIMITED) is an experimental computer algebra system for use on personal computer/workstation. Risa/Asir is a computer algebra system, which provides a programming system Asir with several subroutine libraries that can also be used as parts of other programs.

Example.1

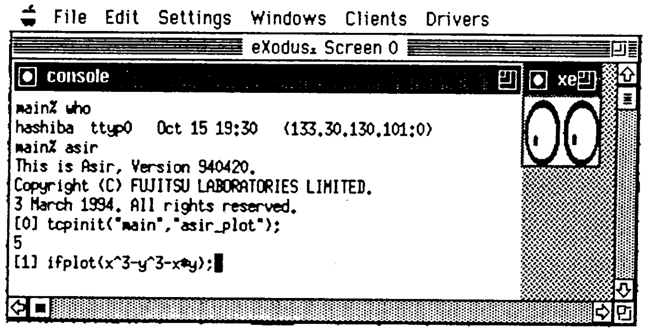


Figure 1: Input the defining equation of nordal curve

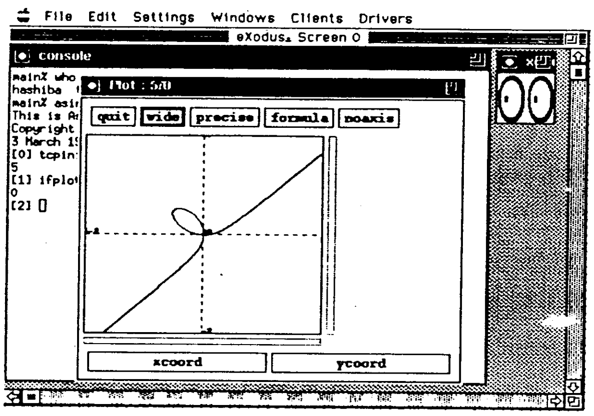


Figure 2: The graph of nordal curve

The first example is shown in Figs.1-4(see [1]).

The scenario below.

Get into Workstation(Sun SS10, machine name "main", IPR:133.30.130.1) by telnet from

Macintosh computer, start up Asir and input the defining equation of nordal curve((Fig.1)

Draw the graph of nordal curve(Fig.2)

Zoom up the graph near nordal point(Fig.3)

Place those graphs(Fig.4)

In this example, Risa/Asir was used a tool for presenting the configuration of plane curve. This example come under the above situatin (I). Risa/Asir

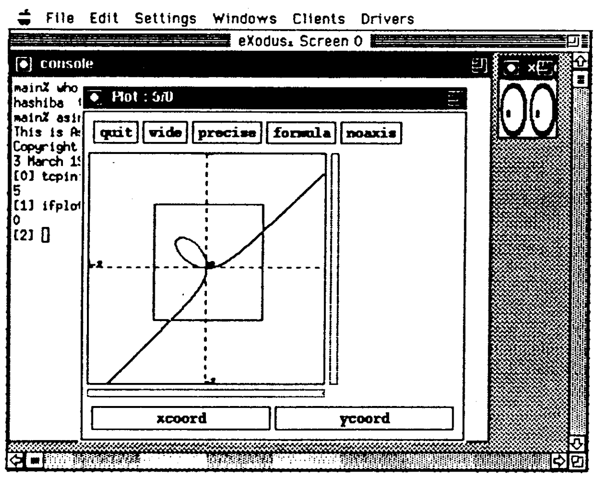


Figure 3: Zoom up the graph

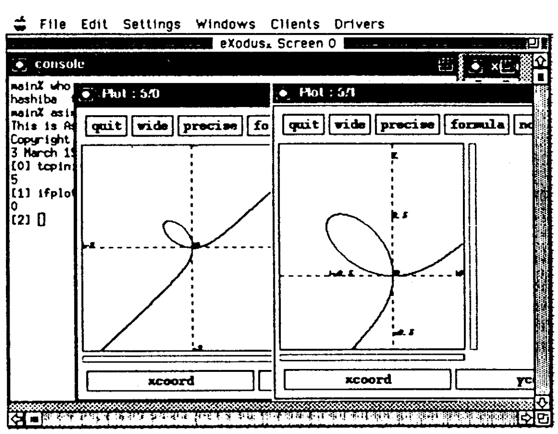


Figure 4: Place those graphs

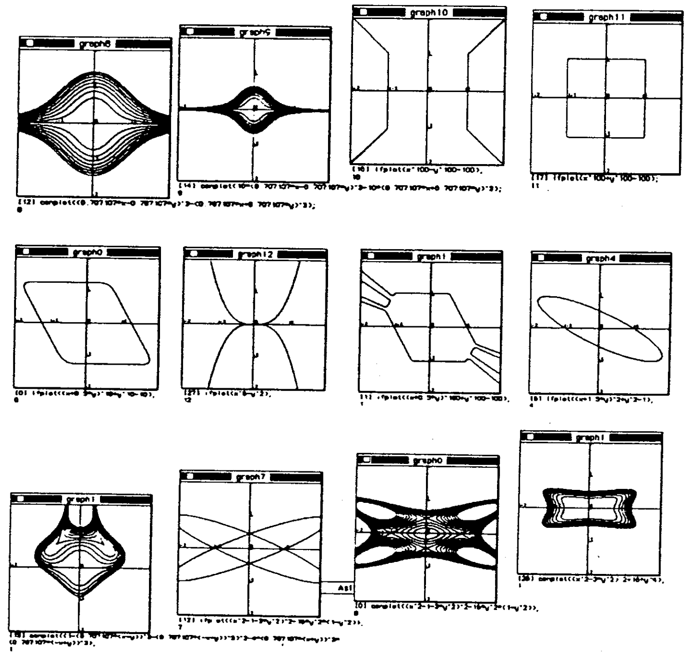


Figure 5: Various images drawn by the implicit functions

is very powerful as a drawing tool for the graphs of implicit functions.

Example.2

The second example is shown in Fig.5(see [2]). Using Risa/Asir, this example will show various images was drawn by the implicit functions.

In this example, Risa/Asir was used as a creative tool for conjecturing and finding. By using Risa/Asir, students can search geometry on the basis of direct and repearted experience, assimilate geometrical concepts, and acquire a positive attitude toward the geometrical world. Moreover, if it is used under proper guidance, it will be a good tool for encouraging students to think mathematically. This example come under the above situation (II).

The usage of the software is expected to be easier by the enhancement of user interface. In particular, input interface of computer algebra systems is not useful for educatinal activity.

3.2 COMPUTER COMMUNICATION

We will summarize the functions of computer communication which are supported by the existing systems, Fundamentally, most of systems has the following functions.

(a) Mailing

(b) File transmission

(c) Read-Time chatting

Communication capability of computer which is pointed as one of the functional chracteristics in 1. This capability is seemed to be a big potential for the educational use. The discussion will give some hints to utilize computer as a tool for educational communication. The concept of groupware has the possibility to provide an advanced view about the educational communication between the teacher and the students.

The function of the computer to support the cooperative activity will be implemented. This cooperative activity should be done in some small groups of students. The computer system will be able to support organizing the small groups. Consider that some of the students find no interesting things though the exploration. Suppose the computer system recognize the situation, it will be able to make suitable groups, while each group contained the students who missed the objective and the ones who got it. In each small group, the students communicate through the computer system discussing or reporting exploratory activity.

The communication by computer establish the connection among the teacher and the students even if they are not in the classroom. The students are able to continue their discussion while keeping their interest. The teacher is able to receive their questions and to repond them via the mailing system in anytime. Here is the educational communication system, it has to be capable not only to exchanging the normal letter but also to deal the images of the mathematical expressions, figures, and/or graphs(see [3]).

It will be realized by the use of computer algebra systems with the educational user-interface supporting.

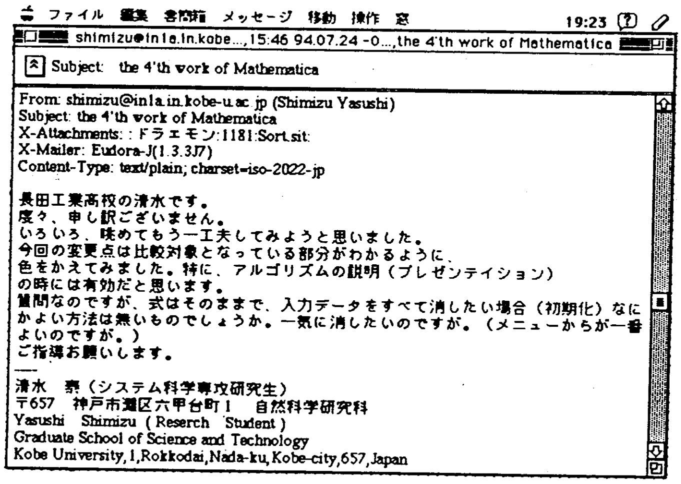


Figure 6: Cooperative work by usig e-mail

3.3 DEVELOPMENT OF TEACHING MATERIALS BY USING ANONYMOUS FTP

We should be construct the ftp site of educational programs for the computer algebra systems which created by teachers and students. This site can be used for development of teaching materials. As experimentally trial, we developed the educational matrials(mathematica programs). We use the mailing function for the cooperative work and sent the programs as attachment documents. This cooperative work was held at Nagata Technical High School in Hyogo Prefecture by Mr. Y. Shimizu, and followed one of them below.

Now, as experimentally trial, Mathematica programs which developed by students(Fig.7) were put in anonymous ftp site(main[IPR:133.30.130.1]).

Mathematica is a sophisticated software pakage designed for use in mathematical education. Students manipulate figures and mathematical information related to them interactively. With these good user-interfaces designed for educational use, students can enjoy manipulating constrained mathematical concept, and can clarify or correct their own ideals through practical activities.

The efficient use of computer algebra systems expects the students to discover, to think and to confirm the contents of lesson through their ex-

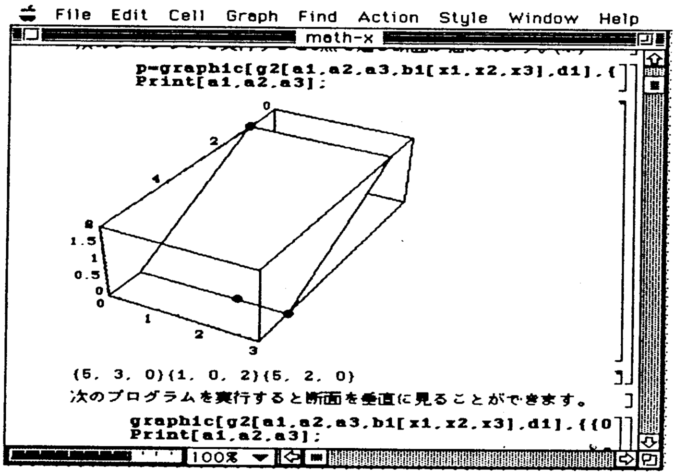


Figure 7: Mathematica programs which developed by student

priemental activity.

After the individual exploration, it is effective to give the students an opportunity to report or to discuss about their discoveries each other.

4 CONCLUSION

The use of computer algebra systems in education will be a very effective way of changing current mathematical classes into ones that allow students to assimilate mathematical concepts through their own directed and repeated exprience. The learning activity of the students supported by computer algebra systems should supersede the traditional one-side lecture. To approach the ideal use of computer algebra systems, it is necessary to study the way to integrate the various uncoordinated attempts into an integrated learning system. However, as the efficient use of computer algebra systems has already shown, the focus of the lesson should move to the students' activity supported by computer.

There is a lack of good computer algebra systems that provide easy-to-use educational (input) interface.

As an ideal in the future,it is excepted to create an enhanced lerning environment software which is designed from the educational viewpoint of the well experience teachers. So it is neccsary to correct such examples. We will correct them on our anonymous ftp site, and open them.

It is expected to make futher study about this potential of the efficient use even though it will not be implemented in the near future.

References

[1] H.Hashiba and Y.Shimizu, Mathematics Education by using Computer Network(in Japanese), Journal of Kinki Mathematics Society Vol.8, Kinki Society for Mathematics Education, pp.30-38, 1995.

[2] T.Kumada, Development of Mathematical Materials by using personal computers(in Japanese), Master Thesis(Kobe University), 1994.

[3] F.Terada and et al., Mathematical Education towards the Next Century, Jornal of Science Education in Japan, Vol.15, No.3, pp.101-109, Japan Society for Science Education, 1991.

THE EDITING SYSTEM FOR HAND-WRITE MATHEMATICAL FORMULA ON THE NEMANTIC DISPLAY

Katsuhide Tsushima , Masayuki Ueno , Sinjiro Wada and Katsunori Nakamura
(Osaka Electro-Communication University)
18-8 , Hatsucho , Neyagawa , JAPAN
(tsushima@tslab.osakac.ac.jp)

1. Introduction

Paper and pencil is DMI for mathematical problem solving

In the problem solving using mathematics, a man thinks about the subject with help of formula manipulation with pencil on paper.
The man write two-dimensional symbol and formula on paper with pencil. He manipulates the term and formula by himself on paper with help of mathematical knowledge.
The written series of formula can be looked as a faithful memo of his mathematical thinking concerning the present subject.
In this case , the flow of formula manipulation on paper coincidents with that of his thinking in his brain. The difference between manipulation and thinking does not arise using paper and pencil in his problem solving. In this sense , paper and pencil can be considered as DMI(Direct Manipulation Interface) for him about mathematical thinking.

The usual computer algebra system is not DMI

As is well known , the interface of the computer algebra system was developed on the TTY environment. The essential improvement concerning the input style of computer algebra system have not done.
So , using the usual computer algebra system , the user is forced to use one-dimensional command to manipulate formula on the computer .
So, the user must convert the two-dimensional mathematical expression that is used by him on paper to one-dimensional command in his problem solving using computer algebra system. This indirect input method prevents the user from thinking about mathematical subject. The mental model of mathematical thinking using the ordinary symbols does not coincident with that of manipulation of mathematical formula on the computer algebra system. So, the computer algebra system is not DMI for mathematical problem solving. In spite of this inadequate human interface , the number of

user of computer algebra system is gradually increasing.

The educational objective to use computer algebra system

In general ,the computer algebra system is used in next three purposes.

1) To avoid troublesome formula manipulation on the paper.
2) To get the mathematical knowledge from the computer.
3) To understand the effect of manipulation commands on the terms and to realize the mathematical meaning of this manipulation.

2) and 3) are important in education. On the other hands , 1) is an important facility for the researcher of physical and mathematical science. Most of all the user begin to use computer algebra system to get this benefit from computer algebra system.

Hand-write input method for computer algebra system

We have engaged for a long time in researching the human interface of computer algebra system.[1),2),3)] But, appropriate direct hardware interface to receive hand-written two dimensional formula does not exist before.

Hand-write input method for formula of computer algebra system have been expected for a long time in the domain of both research and education. Especially , this input method is necessary for the primary and secondary education.

Recently , large size nemantic display can be used as peripheral device of computer. The user can write down the two-dimensional symbol and formula using electric pen on the nemantic display , then these information can be detected by computer to construct appropriate pursing software. The computer can draw this written symbol faithfully on the nemantic display . So, the user feels as if he wrote the formula on the nemantic display in the same way on the paper. We have developed a completely new method that enables the user to input hand-written symbol and formula in two-dimensional figure to computer algebra system using this nematic display.[4),5)]

The details of developed hand-write interface is shown in another paper proposed by us in this conference. (We call this paper 1.)

We focus our attention on editing of handwritten formula and the symbolism which will be used on the developed hand-write system in this paper.

2. Editor for mathematical formula on the nemantic display

A man manipulates the formulae and symbols on paper successively in his problem solving. So, continuous thinking about the problem solving is kept using the help of paper. On the other hand , a man think about idea concerning the next manipulation modifying the formulae with pencil in two-dimensional style.(Fig.1)

$$A = x + y \qquad 1)$$

$$B = x - y \qquad 2)$$

$$\int (A^2 - B^2)\, dx \qquad 3)$$

$$= \int \{(x+y)^2 - (x-y)^2\}\, dx \qquad 4)$$

$$= \int \{x^2 + 2xy + y^2 - (x^2 - 2xy + y^2)\}\, dx \qquad 5)$$

$$= \int 4xy\, dx \qquad 6)$$

$$= 2x^2 y \qquad 7)$$

Fig.1 The Symbolism on paper

In the preceding section , we have accomplished to finish the research of hand-write input method for computer algebra system.

If the processed two-dimensional output on the display can be used for the next input , then the successive manipulation can be possible on the display in the same way as on paper.

A user uses part of formulae in some case as is shown in Fig.2 .

$$\frac{d}{dx}(x^3 + 2x)$$

$$\text{-->} \int 3x^2 + 2\, dx$$

$\int$, dx are added by hand to the output $3x^2+2$

$$\text{-->} \quad x^3 + 2x$$

Fig. 2 Partial use of outputted formula

We have developed graphical manipulation symbols which operate on the two-dimensional formula appearing on the display to ensure the user successive manipulation of formula on the computer.

The following three types of operators are developed in our graphical editor.(Fig. 3)

1) Referencer
2) Modifier
3) Executer

Referencer is used to take out the value of formula and variables. There are two types of value , expression and symbol itself. Each value is transferred to another symbol by drawing line with arrow shown in Fig.4

	Name	Operational Symbol	Example
Referencer	Expressional Referencer	Round Enclosure	$\int_0^{\sqrt{2}} \frac{\sqrt{x+y}}{\textcircled{x}} dx$
	Symbolic Referencer	Boxed Enclosure	$\int_0^{\sqrt{2}} \frac{\sqrt{x+y}}{\boxed{x}} dx$
Modifier	Forbidden Modifier	Upper Waveline	$\frac{d}{dx}F(\widetilde{x})\widetilde{G(x)}$
	Exclusive Modifier	Referencer with a diagonal	$x^3+3x^2+3x+1+\textcircled{c}$
	Central Modifier	Underline	$\underline{\frac{d^2}{dx^2}f(x)+(x+1)\frac{d}{dx}f(x)+f^2(x)+1+\sin x}$
Executer	Move	Arrow (with Referencer)	$\int_0^{\sqrt{2}} \boxed{\frac{\sqrt{x+y}}{x}} dx \searrow$
	Delete	Cross (with Referencer)	$x^3+\cancel{3x^2}+3x+1+c$
	Execute	Double line and Arrow	Expand[(a+b)^2] $\Rightarrow$ $a^2+2ab+b^2$

Fig.3 Operational Symbols in our Graphical Editor

Almost all the graphical manipulation of the formula on our editing system generate the command for computer algebra system.
The temporary variables are generated to take in the effect of graphical manipulation in some case.
And the additional information structure is necessary to take into account the various effect caused by graphical manipulation.
For example , some of them such as movement , do not

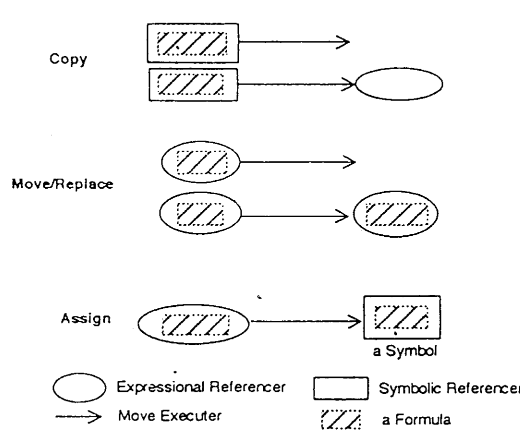


Fig.4 The Move Executer and Referencer

generate the command which can be read by computer algebra system, but varies geometrical information of the corresponding formula.. The data of formula position table is rewritten by this type of graphical manipulation.

3. The symbolism used in ICAS

As is well known , there are the following four types of mathematical entities.

1) Operator or Functional
2) Function
3) Terms and Variable
4) Number

But, the user cannot manipulate operator, functional directly on the computer using the ordinary computer algebra system. In the usual computer algebra system , there are only two information structures , atom and list. So, the system has only two informational entities , function and variable.

In other words , the system does not have the information structure corresponding to the structure of mathematical entities. The mathematical operator and mathematical function is described as function using LISP language in the ordinary computer algebra system. So, the computer algebra system cannot have the mechanism distinguish mathematical operator from mathematical function .

We have already developed unique computer algebra system called ICAS (Intelligent Computer Algebra System). 6),7) ICAS has four informational structures corresponding to the mathematical entities. Operator algebra and functional treatment is possible directly on the computer using ICAS.

Continuous direct manipulation of mathematical entities on the computer was accomplished at the first time. The educational applications of ICAS is researched carefully in reference 1 and 2.

And the application of ICAS to differential equation and difference equation are researched by us .

Symbolic iteration method can be used to solve differential equation and difference equation on ICAS.

The user must control the evaluation of hierarchical mathematical entities by using command which controls evaluation flags in ICAS.

The differential operator Dx and function F(x) is defined in ICAS in the following.

```
Define F(x) , x^4, enddef;          1)
Defop Dx ,Dif(#,x),endop;           2)
```

By controlling the evaluation flags , the user can obtain the various results which he want .

```
Disable(Dx,F);                              3)

(Dx+3).F(x)--->(Dx+3).F(x)                  4)
           ---->Dx.F(x)+3F(x)               5)
 Enable(F);                                 6)
(Dx+3).F(x)--->Dx.x^4+3x^4                  7)

Enable(Dx);                                 8)
(Dx+3).F(x)--->4x^3+3  x^4                  9)
```

The obtained results can be used the next manipulation in the following. Each results corresponds to 9),7) and 5).

```
Dx.@     ---->12  x^2+12  x^3               10)
Dx.@     ---->Dx^2.x^4+Dx.3  x^4            1,1)
Dx.@    ----> Dx^2.F(x)+Dx.3 F(x)           12)

disable(Dx,F);                              13)
(Dx+2)^2 ----->Dx^2+4  Dx+4                 14)
@.F(x)   ----->(Dx^2+4  Dx+4).F(x)          15)
enable(Dx,F);                               16)
 (Dx+2)^2.F(x) ----->12  x^2+8 x^2+4 x^4
                                            17)
```

4.New symbolism which control the evaluation of entity

The developed hand-written formula editor and unique computer algebra system ICAS.
The computer algebra system is developed considering command paradigm. But, the symbolism is used not only to get manipulating of the formula ,but to obtain the overall description of mathematical thinking.8)
The latter is not possible on the ordinary computer algebra system in which one-dimensional expression is inevitable. There is a possibility to realize these two things at the same time on the hand-written formula editing system developed by us.
There are the following three types of manipulation to solve problem using paper and computer algebra system.

7 .K.Tsushima,H.Kaga , K.Nakamura and T.Satou,"Intelligent Computer Algebra System",Rep. of OECU,25,131,(1990).
8. C.S.Peirce,"Collected Papers of Charles Sanders Peirće",Harvard University Press,(1960).

TRIGONOMETRIC FUNCTIONS WITH MOVIL: THE ROLE OF MOVING FIGURE THAT REPRESENTS MATHEMATICAL CONCEPTS.

Shigeru Tsuyuki
Kawaijuku Educational Institution
SF building, 4-22-3 Sendagaya, Shibuya-ku,
Tokyo 151, Japan

INTRODUCTION

We have developed a series of software MoViL at Kawaijuku Educational Institution since 1987. MoViL is an acronym for Moving Visual Language. It reflects our belief that interactive moving figures are new types of 'language' through which teachers can communicate with students. Moreover, moving figures promote students' ability to find their own image of various mathematical concepts. In this paper, we will focus on the applications of MoViL to trigonometric functions, and will discuss the role of moving figures.

HALF-ABSTRACT AND HALF-CONCRETE REPRESENTATIONS

One of characteristics of MoViL is moving figures as medium between the abstract and the concrete. Half-abstract and half-concrete representations are neither real things nor abstract concepts themselves, but they have both characteristics. Any manipulatives, models, stories, and games that represent some mathematical concepts are classified in this category.

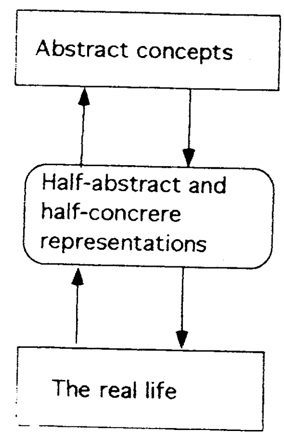


Figure 1. Half-abstract and half-concrete representations.

The first example represents the sine function. This moving picture links students' experiences with the idea of the concept, and have students obtain the necessity of this function. In other words, students recognize the meaning of the function.

The radian measure is sometimes confusing, so we made a program that gives the image of this measure. By winding a thread around a circle, we can measure the angle by the radius.

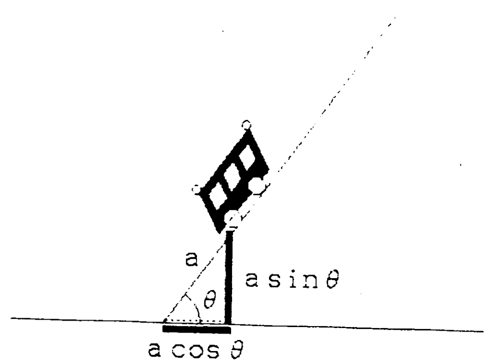


Figure 2. A cable car represents the sine function.

Some teachers use black boxes to introduce the idea of functions. The next example is a combination of a black box and the unit circle. Students understand that the sine function is really a function. This black box is a realization of the image of trigonometric functions.

Those three examples are all located somewhere between the abstract and the concrete. However, while the first example can not be applied to other concepts, black box can be used to show the idea of any functions. Black box has more generality and looks like a figure rather than a picture. Black box is considered to be a 'scheme'. Scheme is a Japanese-English word used in Japanese math education society, and the meaning is different from that used in psychology. Although the definition of the word is different among people, we defined a scheme as "a simple figure that represents mathematical concepts, and is between the abstract and the concrete." A scheme is a realization of our image of concepts in our mind, and has generality to some extent. A rectangle that represents the system of multiplication and division is a common scheme.

THE EFFECTS OF MOTION

Another characteristic of MoViL is that every picture has interactive moving capability. Every motion shows some image or mathematical concepts. In the example of black box, The motion represents the basic idea of the functions and the process of finding the values of the functions. It implies manipulation or somebody's hands.

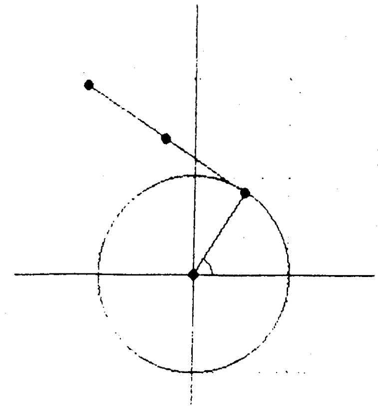


Figure 3. The radian measure.

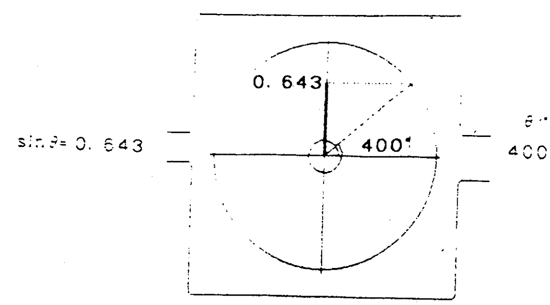


Figure 4. A black box for the sine function.

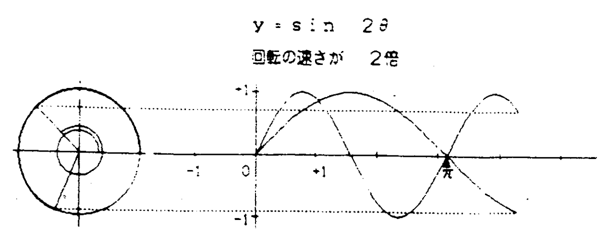


Figure 5. The graph of y=sin ax.

Figure 5 shows the generating process of the graph of y = sin ax from the unit circle. Students understand why the period of this function is $2\pi / a$ by observing that moving radius of y=sin ax is 'a' times faster then that of y=sin x. This relation is unchanged during the demonstration. It is another example of showing manipulation, but also it implies alternative way of understanding in motion.

Another effect of moving figure is to have students obtain transformative perspective. The graph of y=asin(bx+c) is transformed by changing the parameters. They can recognize a graph of y=3sin (2x) as one stage of transformation, and they regard a particular graph as a part of the whole set of graphs. Although y=asin(bx+c) is a simple case, there are more than enough applications of this type. The next program shows that the area of a sector is equal to (arc length) * (radius). This program uses the transformation of an area. The preservation of area during motion is used to show the result. In these cases, students recognize some characteristics are preserved. Moreover, in any continuous transformations, the preservation of some quality or quantity can be seen easily through moving figures.

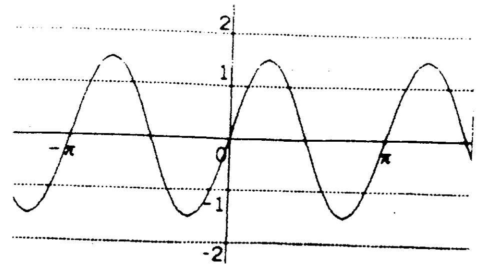


Figure 6. The graph of y=a sin (bx+c).

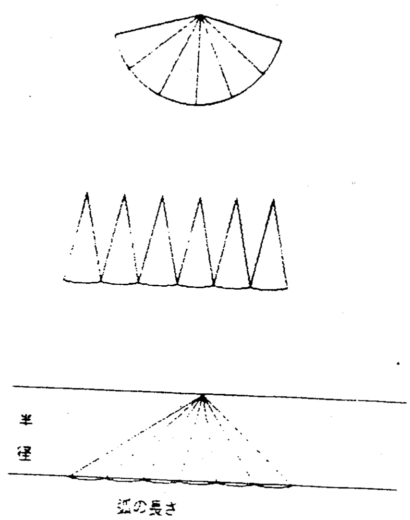


Figure 7. The area of a sector.

AN EXAMPLE OF MOVING SCHEME

The two ideas; half-abstract and half-concrete representations and the effect of motion are the most fundamental for MoViL. Since a scheme has more generality and is applicable to many cases, we are trying to find or create a new 'moving scheme'. In trigonometry, rotating-folded-line that represents the sum of sine functions is an effective moving scheme in our view. In the figure 8, each line segment has the length a_k, and angle x_k. Then the y-coordinate of the end point P is $y=\Sigma\ a_K \sin x_k$. This figure is static, but computer graphics show moving image of this idea. We can observe how the graph is generated by the folded line. If $x_k=bx+c_k$ for all k, the folded line keeps the shape. In other words, the shape is preserved in motion. Therefore it generates a simple sine curve. In this case, students' understanding is qualitative; they understand $\Sigma\ a_k \sin(b_k x + c_k) = A\sin(Bx + C)$, but they don't know the precise value of A and C. The followings are the applications of this scheme.

1)The sum Asin x+ Bcos x

Since $\cos x=\sin(x + \pi/2)$, this sum is equal to $A\sin x + B\sin(x+\pi/2)$. Since the rotating speed of each segment is the same, this function is represented by L-shaped folded line. The same graph is generated by the rotation of the hypotenuse of the right triangle. (Sasaki) It shows the formula;
$A\sin x+B\cos x= A+B \sin(x+a)$ where $\tan a = B/A$.

2)The formula for $\sin(\alpha+\beta)$

The formula $\sin(\alpha+\beta)=\sin\alpha\cos\beta + \cos\alpha\sin\beta$ is derived by similar method. Observing a right triangle whose hypotenuse is 1 and one of whose angle is α, we know the base is $\cos\alpha$ and the height is $\sin\alpha$. By rotating

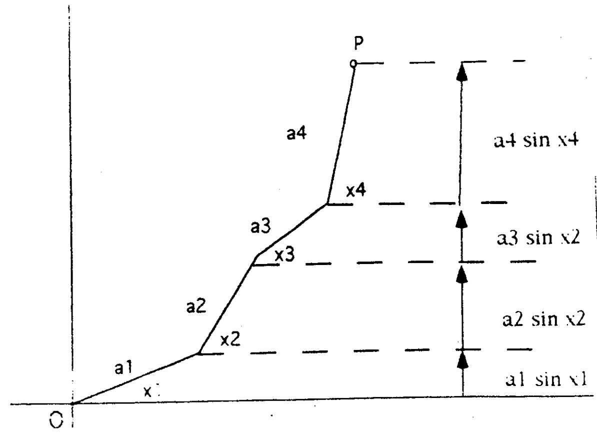


Figure 8. The basic concept of folded lines.

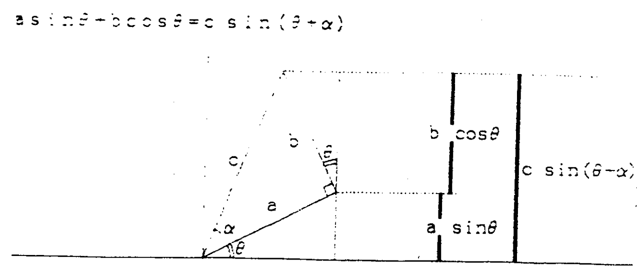


Figure 9. Asin θ + B cos θ

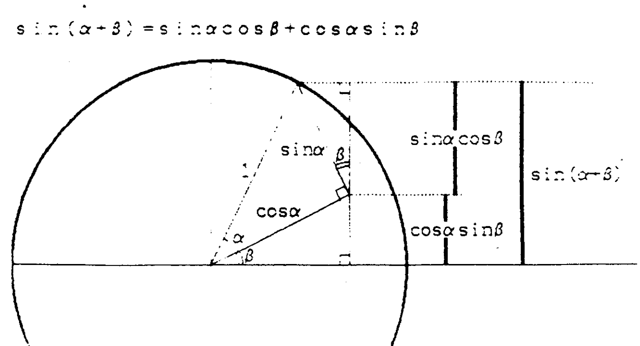


Figure 10. The formula for sin(α+β)

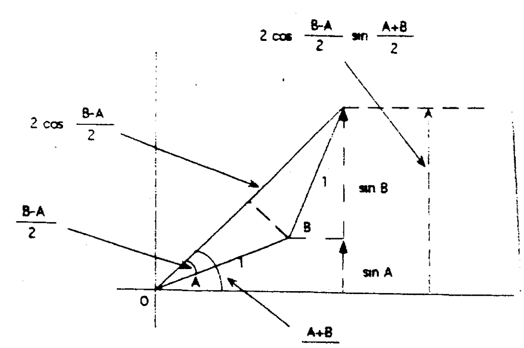


Figure 11. The formula for sin A + sin B.

this triangle by the angle β, we obtain the formula.

3)The formula for sinA+sinB

Figure 11 depicts the formula for $\sin A + \sin B = 2\sin((A+B)/2)\cos((A-B)/2)$. The left side of the formula corresponds to the two line segments whose lengths are 1, and whose angles are A and B respectively. The right side of the formula corresponds to the other side of the triangle.

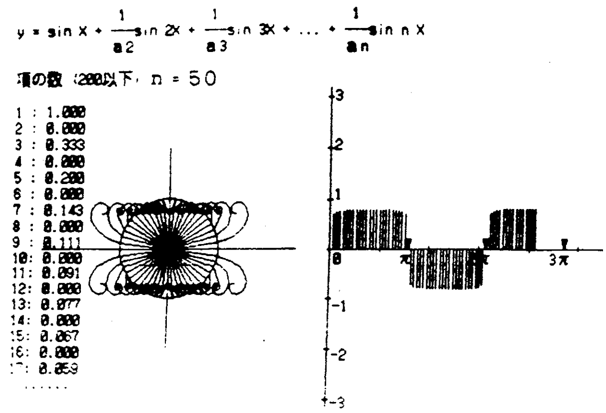


Figure 12. Fourier series.

4)Fourier Series

Fourier series are also shown by connecting an infinite number of line segments. Figure 12 shows the screen of $y=a_k \sin kx$ where $a_k=1/k$.

CONCLUSION

We started from half-concrete and half-abstract representations. Scheme is contained in the category of half-abstract and half-concrete representations, but it is a more general and simpler figure. One schema can be applied to more than several cases, so students use the same scheme to understand other concepts or solving other problems. Using computer graphics, we can add motion to those representations. The capability of moving figures enabled us to develop more effective representations. The following list is considered to be the outcome of adding motion.

1) The motion of figure helps students to obtain qualitative understanding before obtaining quantitative understanding.
2) Students can obtain transformative perspective through viewing continuous deformation.
3) Moving figures promote students' structural knowledge, and have students grasp the relation between the whole and the parts.
3) Preservation is easily observed in gradual transformations.
4) Process shown in the motion can imply manipulation.
5) We can create a new scheme by applying the capability of computer graphics.

REFERENCES

Sasaki.S, Compositions of trigonometric functions, Suugaku Kyoushitsu(Japanese), 5, 1988

HANDWRITING INTERFACE FOR COMPUTER ALGEBRA SYSTEM

UENO Masayuki, WADA Shinjirou, TSUSHIMA Katsuhide
Osaka Electro-Communication Univ.
18-8 Hatsu-chou, Neyagawa city, Osaka 572, Japan
ueno@tslab.osakac.ac.jp

Abstract:

Handwriting interface for computer algebra system has developed. In this system, the users are not only able to use two dimensional symbolism in mathematics using on paper but also operate computer algebra system with pen tablet. In this paper, we described the human interface and the formula recognition system.

0. Preface:

We expected computer algebra system to accelerate human thoughts in mathematics. Howerver, using ordinary computer algebra system instead of pen and paper is difficult for the user to think with formula. One reason for this is that ordinary computer algebra system compel the user to use special symbolism. The symbolism is called TTY expression that is designed one dimensional (like a+b/c^3) for keyboard and character display. This symbolism is always more difficult for the user to know the structure and write it than two dimensional formula expression using on paper.

Although, *display tablet device* (also called *nemantic display*) that can be used input and output in same place like pen with paper should solve the problem. Display tablet device display character and graphics, and sense the pen position and either of pen is touched or not touched on the tablet, and send the data to host computer. Thus, display tablet can handle two dimensional formula. However, two dimensional formula expression causes another problem. Editing interface for two dimensional formula expressions had not been developed. We have also developed *sign language* to edit for two dimensional formula expressions. We can summarize which we have developed seamless pen interface for computer algebra system.

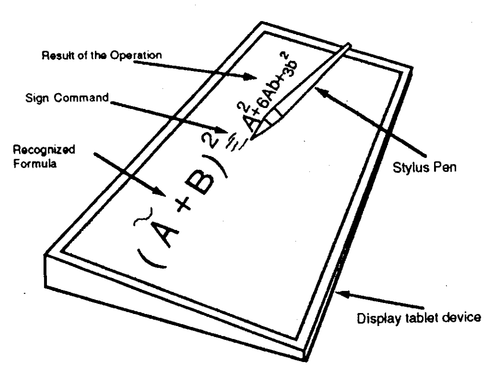


Fig.1 display tablet device and stylus pen

1. Handwriting Interface Overview:

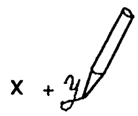


Fig.2 Writing a character

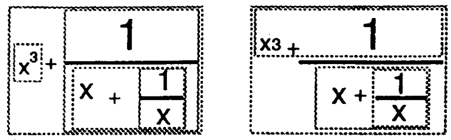


Fig.3 Structure display (good case and ill case)

Hardware of handwriting interface for computer algebra system is consisted of a Display tablet device, a stylus pen and a computer that is running computer algebra system. The user who interact with handwriting interface can write in two dimensional expression on the tablet with the stylus pen. Handwriting interaction between the user and the tablet is like Fig.1.

Strokes that have been written on the tablet are recognized to a symbol and typed out on the tablet quickly for making out what symbol is recognized to the user(Fig.2). If the user find out that the recognized symbol is not intended, the user can delete the symbol with the pen sign(In chapter 2, we describe pen sign language). After writing formula, when the user ordered mathematical operations with sign language, formula level recognition start and system display hierarchical structure of the formula (Fig.3).

Group	Name	Sign	Example
Designator	Designator (Designate a part of formula)	Enclosure	$\frac{\sqrt{x+y}}{(x)}$
Modifier	Forbidden Modifier (Forbid formula evaluation)	Upper Bar	$\frac{d}{dx}\widetilde{F(x)}G(x)$
	Focus Modifier (Evaluate only this)	(designator)	$(x^2+2x)+3(c+1)$
Verb	Move (Editing)	Simple Arrow with designator	$\int_0^1 \frac{\sqrt{x+y}}{x} \leftarrow (dx)$
	Copy(Editing)	Triangle Arrow with designator	$\int_0^1 \frac{\sqrt{x+y}}{x}\, dx$
	Delete (Editing)	Ex (with designator)	$x^2+2x+3c+1$
	Localize (Algebraic) (Localize definition environment)	Double Enclosure	$c=4$, $x^2+2x+3c+1=0$
	Simplify (Algebraic)	Double line and Arrow	$x^2+2x+3c+1 \Rightarrow$
	Factorization (Algebraic)	Double Line and ">"	$x^2+ax+bx+ab \geq$
	Expand (Algebraic)	Double Line and "<"	$(x+a)(x+b) \leq$

If the user finds out that the recognized formula is not intended, the user can stop the process of the operation and correct as mentioned earlier. Consequently, this interactivity such as correcting the mistaken recognition should permit low recognition rate in practical use.

2. Sign Language

Sign language which we have designed, is consisted of three elements; *Designator*, *Modifier* and *Verb*(Fig.4). Designator sign designates a part of formula to order editing or algebraic operation with an enclosure.

Modifier sign modifies algebraic meaning of a part of formula. *Forbidden modifier* forbid a part of formula evaluation. This modifier came out of I.C.A.S (Refer the reference-1). *Focus modifier* handle to evaluate a part of the formula. Focused formula is only evaluated in algebraic verb process if this modifier designated. Verbs are the commands that execute the operation when the system just recognized. *Move verb* which is written as an arrow indicates moving a formula or a part of formula with designator from the source to the destination. *Copy verb* which is written as another type arrow likewise copy from the source to the destination. *Delete verb* delete under the marked a part of formula with a designator, a modifier or a character. *Localize verb* which is written double enclosure is a definition of the separation from global definition environment. This function hide definitions before. *Simplify, Factorize* and *Expand* verbs are corresponding to the commands in normal Computer Algebra System.At last, their signs are functionally complete in editing on tablet device.

3. Handwriting Computer Algebra System:

The Handwriting Computer Algebra System - block diagram is expressed as Fig.5. Drawn Strokes on display tablet is read by INPUT/OUTPUT manager. If the stroke is formula stroke†, INPUT/OUTPUT manager sends the each stroke coordination sequence to the formula recognition unit. If the stroke is sign command stroke, INPUT/OUTPUT manager sends the stroke sequence to the sign recognition unit.

Formula recognition unit has three stages to recognize formula from stroke sequence(Fig.6). First stage is stroke level recognition. Each character is consisted of some strokes. Quantamized angle recognition(Q.A.R.) module and characteristic points recognition(C.P.R.) module recognizes a part of character from the character pattern templates(we describe each module details chapter4). Second stage, character recognition module constructs each strokes and makes a character. This two stages are executed when the user written the

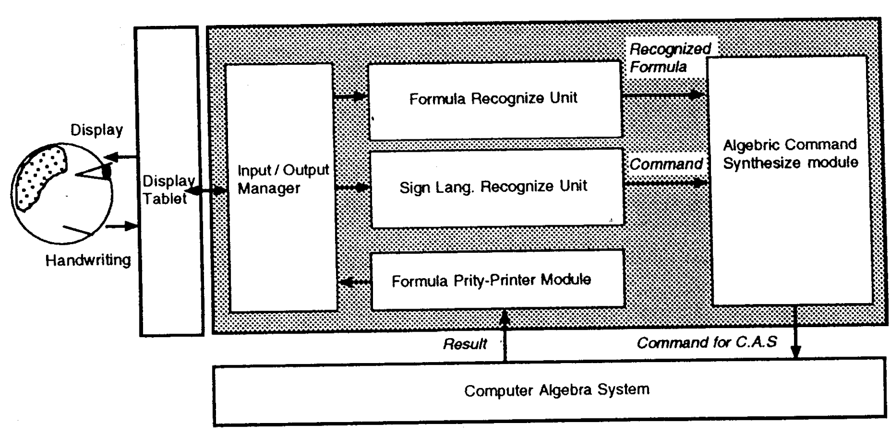


Fig.5 Block diagram

strokes. Particularly, Character recognition module uses a time between a stroke and next stroke, is executed if this time above 100ms. The character recognition module returns the result as an object in CLOS††. Third stage,If the command of sign required, formula recognition module construct an internal expression of the formula structure from arrangement of the recognized character objects.

Sign recognition unit which recognizes the signs is similar to formula recognition unit. Different points from the sign recognition unit to the formula recognition unit are which signs are revolvable and signs construct only simple structure. As a result, each stages are simpler than formula recognition unit and we omit the detail.

After the stroke recognition, Suppose that the commands of the sign require to execute computer algebra system. From the recognition, algebric command synthesize module convert expression, synthesize a command for the computer algebra system(this module have a table to convert for the target system), and send the command to the computer algebra system. The computer algebra system execute the command, and return the result to formula prity-printer module. Formula prity-printer module convert the result to the internal expression, and arrange the elements for ordinary formula expression in mathematics(each elements are expressed objects and manage the position themselves). INPUT/OUTPUT module accept the arrangement of the formula. At last the module display the result of the formula on the tablet.

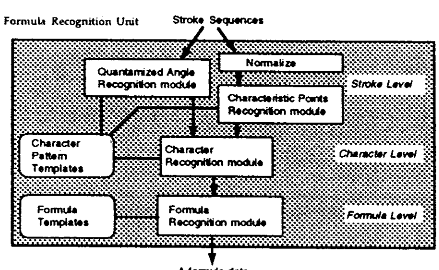


Fig.6 Formula recognize unit

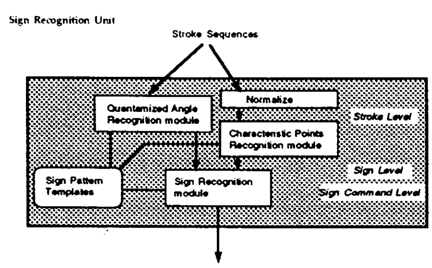


Fig.7 Sign recognize unit

† It is difficult to distinguish only stroke. Because, it is difficult to design a sign language that has no conflict with formula symbol set. This system is using a button on stylus pen to distinguish either formula stroke or sign command stroke.

†† This system programmed on Object Oriented Language - CLOS (Common Lisp Object System). And, each recognized symbols and formula structure elements are expressed in a class. The class of the instances contains information of the position and the symbol.

4. Recognition of character:

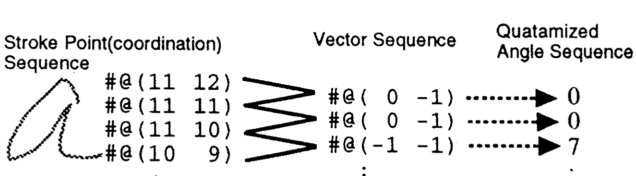


Fig.8 generating the angle sequence

Q.A.R. pick each two points in the coordination sequence, make the difference of the pair and the vector, calculate the angle, and quantamize the angle to eight

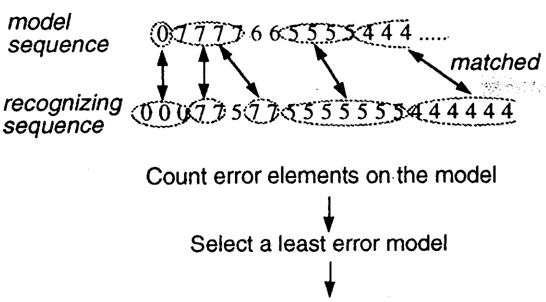


Fig.9 matching (Q.A.R)

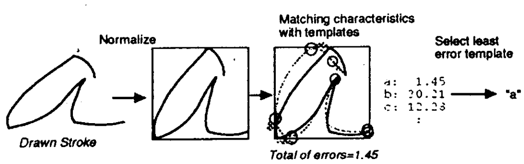


Fig.10 matchig (C.P.R.)

directions(Fig.8). Character model for Q.A.R is quantamized angle sequence which is made similar process. Matching with the model is looking same pattern in the model pattern, counting the matched element sequence. All models are checked the count, and selected a least error model(Fig.9).

C.P.R look and compare characteristic points on the model and the object. Characteristic points which C.P.R looks is start and end point of the stroke, kinked points (sudden angle changing point - In our case: above 90°) and the point of touching with the bounding box of the stroke. Recognition by characteristics points progress like Fig.10. First, a drawn stroke is normalized. Second, match characteristic points of the stroke with a model in the templates. Third, calculate all point errors and summed. Fourth, The models are competing in the error value and return a model as the solution of recognition.

If their two modules' answers are not conflicting, the system calculates total error value(summation of two error value) and select the least error value.

5. Recognition of formula:

Formula recognition process is expressed Fig.11. Before explaination of the figure, We must describe the template to recognize the formula. The template which is expressed formula structure in mathematics is called a *formula template*. Formula template is divided into *hierarchical* template and *relational* template according to matching style. Hierarchical template has a characteristic symbol which definite a template and some indefinite elements(Fig.12). Relational template has no characteristic symbol which definite a template, and is definite by arrangement of the elements(Fig.13).

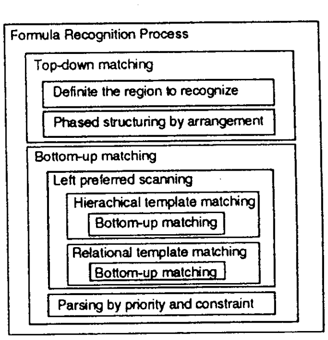


Fig.11 Formula recognition algorithm

Formula template description is like Fig.14. All elements have optionally name, relative position which is expressed a pair of a direction symbol and another element and match-able symbol type which is expressed a symbol that is defined by another. In addition, the template is also able to have necessary or sufficient conditions of place and size.

Formula recognition process is divided into two

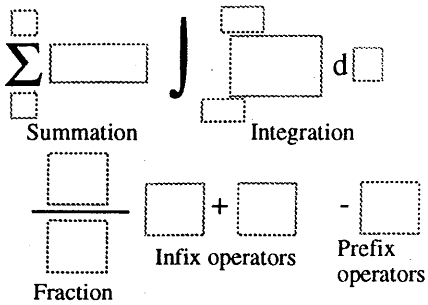


Fig.12 Hierarchical forms

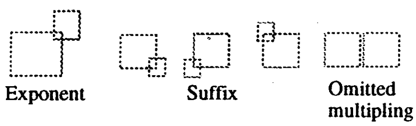


Fig.13 Relational forms

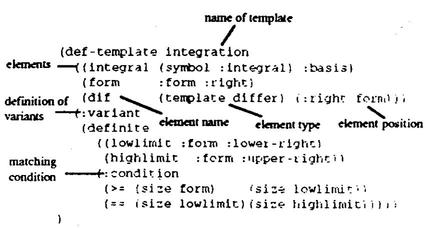


Fig.14 a formula template definition

stages. First stage is top-down matching. In this stage, the system definite the region to recognize by newly written, and look possibility to segment the region by vertical and horizontal space, and make the structure stepwise from the segmentation. Next stage is bottom-up matching. In this stage, the system make matching with templates that have the pattern of the formula notation. The system scans the segment horizontal from left to right. If the segment has a characteristic symbol, the system call bottom-up matching process recursive with all indefinite elements. And, all elements are checked to match with relational templates. But, the system cannot definite operator priority in the scanning because of not scanned operator. After the scanning, the system definite the priority based matching. At last, the formula structure is definited.

6. Conclusion:

Performances in recognition decide which is made practicable or not. In general, recognition rate of our character recognition module is 80%, recognition rate of our formula recognition module is 90%. However, their numbers are not very high, the system has interactive correction mechanism. As a result, we thought that the system is made practicable.

Another criterion to be made practicable is sign language. However, our sign language has sufficiently functions, we have not estimated the sign language cognitively. We should estimate the sign language practical, cognitive experiments.

Consequently, We have developed the Pen based Computer Algebra System. This interface will make of not only a tool for everyone but also an education system for students in algebra from computer algebra system.

References:

[1] TSUSHIMA K. and SATOH T., "OAP — Operator Algebra Package", SIGSAM Bull.,20,39(1986)

[2]WADA S., UENO M., MATSUMOTO M., TSUSHIMA K., "The Development of Handwrite Interface for Computer Algebra System", journal of Japan society for CAI, Vol.11, No.3, pp.136(1994 Oct.)

SOFTWARE DEVELOPMENT IN MMRC †

Dingkang WANG & Lihong ZHI
Institute of Systems Science, Academia Sinica
Beijing 100080, P.R. China
(dwang@mmrc.iss.ac.cn, lzhi@mmrc.iss.ac.cn)

Abstract

In this paper, we report three packages CSET, SACCS, WSOLVE and one system GPROVE. CSET and SACCS are packages implementing characteristic set method. WSOLVE is a package for solving system of polynomial equations. GPROVE is a mechanical proving system for elementary geometric theorems.

1. Introduction

The characteristic set method which considers the zero structure of polynomial sets was first introduced by J.F.Ritt (see [RITT]) and revived by Wu Wentsun (see [WU1]). Many improvements and modifications on this method have been made since 1978. This method is successfully used in many areas such as mechanical theorem proving in geometries (see [WU2] or [CHOU]), computer aided geometry design(CAGD) (see [WU3]), nonlinear programming (see [WU4]), Stewart platform equation solving (see [W-H]), Yang-Baxter equation solving (see [SHI]), etc.

Developing more efficient and practical algorithms for characteristic set method and its application is very important. CSET and SACCS are provided for this purpose. CSET was written in Maple, SACCS was written in C language by means of the basic subroutines in SACLIB (a library of C programs for computer algebra). These two packages are made of algorithms for computing characteristic set of any polynomial set, decomposing polynomial set into ascending sets and irreducible ascending sets, decomposing algebraic varieties into irreducible components, factorizing polynomial over algebraic extension fields and solving system of polynomial equations.

Geometric theorem proving and polynomial equation solving are two main applications of the characteristic set method. WSOLVE is designed specially for solving zero dimensional system of nonlinear algebraic equations, written in Maple.

†Mathematics Mechanization Research Center

GPROVE is a theorem prover, which can be used for proving theorems in Euclidean geometry , non-Euclidean geometry and projective geometry. For a geometric statement of constructive type(see [GLLL]), the prover can decide whether the statement is true. At the same time, it will generate the figure and the non-degenerate conditions in geometric form of the statement automatically. It is programmed in C and implemented on Sun SPARC station.

2. Preliminaries

Let $K[x_1, x_2, \cdots, x_n]$ be the ring of polynomials in indeterminates $x_1, x_2, \cdots, x_n$ with coefficients in a field K of character 0. Consider a fixed ordering on the set of indeterminates: $x_1 \prec x_2 \prec \cdots \prec x_n$. Let $f \in K[x_1, x_2, \cdots, x_n]$, if f is not a constant, $class(f)$ is the greatest i such that x_i occurs in f, otherwise $class(f) = 0$.

Definition 1. Given two polynomials $f_1, f_2 \in K[x_1, x_2, \cdots, x_n]$, we say f_1 is reduced with respect to f_2. if $deg_{x_c}(f_1) < deg_{x_c}(f_2)$, $c = class(f_2)$.

Definition 2. A sequence of polynomials $AS = \{f_1, f_2, \cdots, f_r\}$ is called an *ascending set* if $r = 1$ and f_1 is not identically zero, or $r > 1$ and $0 < class(f_1) < class(f_2) < \cdots < class(f_r) \leq n$ and each f_i is reduced with respect to the preceding $f_j (1 \leq j < i)$.

Definition 3. Consider an ascending set $AS = \{f_1, \cdots, f_r\}, f_i \in K[x_1, \cdots, x_n]$, and a polynomial $g \in K[x_1, \cdots, x_n]$. Let's pseudo-divide g by $f_r, \cdots, f_1$ consider successively as polynomials in $x_{c_r}, \cdots, x_{c_1}$, $c_i = class(f_i)$, and denote the final remainder by R. Then we shall get an expression of the form:

$$I_1^{s_1} \cdots I_r^{s_r} g = \sum_{i=1}^{r} Q_i f_i + R$$

where I_i is the initial of f_i and s_i the smallest possible power achievable for the expression above. R is called the *pseudo-remainder* of g with respect to AS, denoted as $R = Prem(g, AS)$.

Definition 4 An ascending set CS is called a *characteristic set* of polynomial set PS if CS is contained in the ideal generated by the polynomials in PS and $Prem(f, CS) = 0$ for all f in PS .

Well-Ordering Principle: There is an algorithm which permits to determine characteristic set CS for any given polynomial set PS in a finite number of steps that satisfies

$$Zero(CS/J) \subset Zero(PS) \subset Zero(CS)$$

J is the product of initials of polynomial in CS.(see [WU5,page103])

3. CSET and SACCS

CSET and SACCS are two packages implementing the characteristic set method. The basic structure of the algorithms in CSET and SACCS are almost the same. For explicity and convenience, we discuss briefly the algorithms in SACCS and some techniques related to them.

3.1 Description of algorithms

SACCS consists of six main algorithms. Each algorithm realizes a function for characteristic method.

3.1.1 Charset

This algorithm computes a characteristic set CS of any polynomial set PS . CS satisfies:

$$Zero(CS/J) \subset Zero(PS) \subset Zero(CS)$$

J is the product of initials of polynomial in CS.

3.1.2 Charser

Programmed here is a function for decomposing any polynomial set PS into a finite sequence of ascending sets CS_i which satisfies:

$$Zero(PS) = \bigcup_{i=1}^{n} Zero(CS_i/J_i)$$

in which each J_i is the product of all initials of polynomials in CS_i.(see [WU5,page 103])

3.1.2 IRRcharser

This routine computes the decomposition of a polynomial set PS into a sequence of irreducible ascending sets $IRRCS_i$ which satisfies:

$$Zero(PS) = \bigcup_{i=1}^{n} Zero(IRRCS_i/J_i)$$

where J_i is the product of all initials of polynomials in $IRRCS_i$. (see [WU6,page 11])

3.1.4 IRRvariety

This algorithm decomposes the algebraic variety defined by a polynomial set PS into a sequence of irredundant and irreducible algebraic varieties defined by polynomial set VS_i.

$$Zero(PS) = \bigcup_{i=1}^{n} IRRvar(VSi)$$

(see [WU6,page 11])

3.1.5 Pfactor

Pfactor computes the irreducible factorization of a polynomial F over algebraic extension fields defined by an irreducible ascending set or a polynomial set. It combines modular technique with characteristic method so that improves dramatically the efficiency of the algorithm for factorization. (see [ZHI]).

3.1.6 Psolve

For a given polynomial system PS, we have

$$Zero(PS) = \bigcup_{i=1}^{n} Zero(CS_i/J_i)$$

where CS_i is in triangular form.

The zero set of CS_i can be obtained by solving successively for the leading variables in turn. Psolve use this technique to give all the solutions of the original polynomial equations $PS = 0$.

3.2 Some techniques for improving speed

We adopted several strategies in the design and implementation of CSET and SACCS.

3.2.1 Optimization of variable ordering

The efficiency of all the algorithms in SACCS and CSET depends heavily upon the choice of variable ordering. For example

$$PS = \left\{ \begin{array}{l} (4-5x_1+4x_4)^2+4(-5x_1-2x_3)^2-8x_1^2, \\ (2+x_1+4x_4)^2+(4-8x_1-6x_3)^2-16x_1^2, \\ 4x_2^2-4x_1x_2-11x_1^2, \\ 4x_3^2+4x_2^2+4x_1x_2-15x_1^2-4x_1-8x_2+4 \end{array} \right\}.$$

$Charset(PS, [x_1, x_2, x_3, x_4])$ Time: 0.21 seconds.

$Charset(PS, [x_2, x_3, x_4, x_1])$ Time: 245.73 seconds.

So it is important to select an optimal variable ordering. In our implementation, if the variable ordering is not given, we choose a heuristic variable ordering according to some laws which comes from experience. We will pay more attention to this problem in the future.

3.2.2 Controlling the expansion of polynomials

The size of polynomials including the terms and coefficients may grow very quickly while using the characteristic set method (see [WU1]). Sometimes the number of terms increases to thousands. Several strategies are discussed by Wu Wentsun and others.

Wu introduced the concepts of quasi and weak ascending set. (see [WU7,page6]) $CS = \{f_1, f_2, ..., f_n\}$ is a *quasi-ascending set* if either $n = 1$ and $f_i \neq 0$, or $r > 1$

and $class(f_1) < class(f_2) < \cdots < class(f_n)$. A quasi-ascending set is called a *weak-ascending set* if the initial of f_j is reduced with respect to f_i for all $j \geq i$. S.C.Chou and X.S.Gao also define the weak ascending set and weak prem (see [C-G1,page 6]). We provide four options for user to choose.

Polynomial factorization is also an effective tool for controlling the expansion of polynomial.

During the computation of pseudo-remainder, some redundant factors will appear according to the theory of Collins (see [COL] and [Li]) We adopted Linear Equations Method (see [WU5]) to avoid partially the redundant factors.

3.2.3 Removing redundant branches

Because of the recursive property of characteristic set method, the number of branches may be hundreds or even thousands. Some of these branches are redundant and should be cut off.

The following theorem provides a trick to remove the redundant branches (see [C-G1])

Theorem : Let AS_1 and AS_2 be two ascending set. AS_1 is irreducible, if the non-zero remainder set of AS_2 with respect to AS_1 is empty and the remainder of J_2 is non-zero, then $Zero(AS_1/J_1) \subset Zero(AS_2/J_2)$, where J_1, J_2 are respectively the product of initials of the polynomials in AS_1 and AS_2, so that AS_1 can be removed.

There are many other techniques , we refer to (see [DMW] and [C-G2]).

4. WSOLVE

WSOLVE is a Maple package for solving zero dimensional system of polynomial equations. For a given system of polynomial equations $PS = 0$, WSOLVE will decompose PS into a series of ascending set, which is in a "triangular" form and can be solved easily.

For a polynomial set PS, if PS has only finitely many solutions, we have

$$Zero(PS) = \bigcup_{i=1}^{n} Zero(AS_i)$$

where each AS_i is an ascending set.

The specification of WSOLVE is as follows:

$LIST \leftarrow$ *wsolve*(PS,X)

Input:

PS is a list of polynomials;
X is a list of ordered indeterminates.

Output:

$LIST$ is either NIL, which implies $Zero(PS) = \emptyset$, or
a list, which consists of a finite number of ascending set AS_i,
$Zero(PS) = \bigcup_{i=1}^{n} Zero(AS_i)$

For a polynomial system PS, we can get a series of subsystems AS_i, each AS_i is a ascending set so that the zeros of the polynomial system PS are the union of the zeros of the subsystems AS_i. Each subsystem is in a "triangular" form and the leading coefficients of the polynomials in each AS_i are constant so that the numerical solutions of the system can be obtained easily.

There are many techniques and tricks used in WSOLVE to improve the efficiency of the algorithm. Every polynomial produced by the pseudo-remainder algorithm will be factored. This factorization will reduce the amount of the computation greatly. The redundant factors and branches will be cut off in the computing process.

5. GPROVE

GPROVE is a geometric theorems prover which was designed by Dr. Xiaoshan Gao and Dr. Dongdai Lin. It is based on Xlib and written in C language and consists of a main window and two sub-windows. One sub-window is the proving sub-window, the other one is the graphics sub-window. The geometric theorem will be given in the main window. When you press the prove item in the main menu, the proving detail will be shown in the proving sub-window.The users also can use mouse to draw the diagram according to the geometric statements in the graphics sub-window. The user can also move certain points in a diagram smoothly to change the shape and position of diagram to see if it is true when the free points are moved.

There are three main steps to prove the geometric statement in GPROVE. First, the geometric statement will be translated into a normal form so that it can be represented by polynomial equations. The hypotheses will be represented by polynomial equations $PS = 0$ while the conclusion statement will be described by $G = 0$. Second, the characteristic set CS of PS can be got in finite number of steps according to the Well Ordering Principle. Finally, the pseudo-remainder of polynomial G with respect to the characteristic set CS will be computed, if the pseudo-remainder is zero, then the geometric theorem is generic true.

An example is given to show the proving details.

1. The theorem can be loaded from a file or typed directly in the main window.

```
The input to the prover:

Example Pappus. Let A, B, and C be three points on a line; and A1, B1,
and C1, be three points on another line. P is the intersection of
line AB1 and line A1B. Q is the intersection of lines AC1 and CA1. R
is the intersection of lines BC1 and CB1. Show that P, Q, and R are
collinear.
```

2.The constructive description of the geometry statement will be given.

```
EXAMPLE Pappus
```

```
HYPOTHESES:                          HYPOTHESES:
POINT A B A1 B1;                         ;A,B,A1,B1 are free points
ON-LINE C A B;                           ;C,A,B are on the same line
ON-LINE C1 A1 B1;                        ;C1,A1,B1 are on the same line
INTERSECTION-LL P A B1 A1 B;             ;P is the intersection of line AC1 and CA1
INTERSECTION-LL Q A C1 C A1;             ;Q is the intersection of line BC1 and CB1
INTERSECTION-LL R B C1 C B1;             ;R is the in
CONCLUSION:                          CONCLUSION:
COLLINEAR P Q R.                         ;P,R,Q are collinear
```

3. The predicate form of the geometry statement and the non-degenerate conditions will be given.

```
The  HYPOTHESES:                     HYPOTHESES
COLL C A B.                              ;C, A and B are collinear
COLL C1 A1 B1.                           ;C1, A1 and B1 are collinear
COLL P A B1.                             ;P, A and B1 are collinear
COLL P A1 B.                             ;P, A1 and B are collinear
COLL Q A C1.                             ;Q, A and C1 are collinear
COLL Q C A1.                             ;Q, C and A1 are collinear
COLL R B C1.                             ;R, B and C1 are collinear
COLL R C B1.                             ;R, C and B1 are collinear
THE CONCLUSION:                      CONCLUSION:
COLL P Q R.                              ;P, Q and R are collinear
```

The non-degenerate conditions:

```
LINE AB IS NON-ISOTROPIC.
LINE A1B1 IS NON-ISOTROPIC.
AB1 DOES NOT PARALLEL TO A1B.
AC1 DOES NOT PARALLEL TO CA1.
BC1 DOES NOT PARALLEL TO CB1.
```

4. The coordinates of the points will be given and the geometry statement will be represented by polynomial equations.

```
A: (0 0)   B: (x1 0)  A1: (x2 x3) B1: (x4 x5)
C: (x6 x7) C1: (x8 x9) P: (x10 x11) Q: (x12 x13)
R: (x14 x15)

THE HYPOTHESES:
x1x7 = 0                                 ;C, A and B are collinear
(x4-x2)x9+(-x5+x3)x8+x2x5-x3x4 = 0       ;C1, A1 and B1 are collinear
```

```
x4x11-x5x10 = 0                          ;P, A and B1 are collinear
(-x2+x1)x11+x3x10-x1x3 = 0               ;P, A1 and B are collinear
x8x13-x9x12 = 0                          ;Q, A and C1 are collinear
(-x6+x2)x13+(x7-x3)x12-x2x7+x3x6 = 0     ;Q, C and A1 are collinear
(x8-x1)x15-x9x14+x1x9 = 0                ;R, B and C1 are collinear
(-x6+x4)x15+(x7-x5)x14-x4x7+x5x6 = 0     ;R, C and B1 are collinear

THE CONCLUSION:
 CONC =                                  ;P, Q and R are collinear
(x12-x10)x15+(-x13+x11)x14+x10x13-x11x12 = 0
```

5. According to the Well-Ordering Principle, the characteristic set TS of the hypotheses polynomial system will be computed.

```
 TS =
(-x6+x4)x15+(x7-x5)x14-x4x7+x5x6 = 0
((x6-x4)x9+(-x7+x5)x8+x1x7-x1x5)x14+(-x1x6+x1x4)x9+(x4x7-x5x6)x8
          -x1x4x7+x1x5x6 = 0
x8x13-x9x12 = 0
((-x6+x2)x9+(x7-x3)x8)x12+(-x2x7+x3x6)x8 = 0
x4x11-x5x10 = 0
((-x2+x1)x5+x3x4)x10-x1x3x4 = 0
(x4-x2)x9+(-x5+x3)x8+x2x5-x3x4 = 0
x7 = 0
```

6. The successive pseudo remainder of CONC w.r.t to TS will be computed. The class, leading degree and terms of the pseudo remainder are given. If the pseudo remainder is 0, It means the theorem is generic true.

```
Index: (15,1,6)
Index: (14,1,10)
Index: (12,1,27)
Index: (10,1,18)
Index: (9,2,12)
Index: (0,0,0)
The statement is true.
```

6. Remarks The packages and system together with user's guide are available via anonymous ftp at

mmrc.iss.ac.cn:/pub/software

A new system STAR - A Small Tool for Algebraic Research is being developed in MMRC. It will be written in C++ and provide a symbolic environment for polynomial operations and the computing of the characteristic set. At last, we must point out that much of the material summarized here was done by the other members of MMRC software group.

Acknowledgement The authors would like to thank Prof. Wu Wenda, Prof. Shi He and Dr. Liu Zhuojun for their valuable suggestions.

References

[CHOU] S.C.Chou ,"Mechanical Geometry Theorem Proving" , D.Reidel Publishing Company, Dordrecht, Netherlands (1988).

[C-G1] S.C.Chou & X.S.Gao, "Ritt-Wu's Decomposition Algorithm and Geometry Theorem Proving" Proceedings of CADE-10, Lecture Notes in AI, Vol. 449, 1990.

[C-G2] S.C.Chou & X.S.Gao, "An Algebraic System Based on the Characteristic Method ", IWMM'92 , (1992)1-19.

[COL] G.E.Collins, "Subresultants and Reduced Polynomial Sequences", J.ACM, 14 (1967)128-142.

[DMW] D.M.Wang, "An Implementation of the Characteristic Set Method in Maple" RISC-LINZ series no.91-25.0 Johannes Kepler University,Austria (1991).

[GLLL] X.S. Gao,Y.L. Li, D.D. Lin, X.S.Lu, "A Geometric Theorem Prover based on Wu's Method", IWMM'92 , (1992).201-205.

[LI] Li Ziming, "A new proof of Collins's theorem", MM-Res. Preprints, No.1 (1987) 33-37.

[RITT] J.F.Ritt, "Differential Algebra " , AMS Colloquium Publications, New York,1950.

[SHI] Shi He"On Solving the Yang-Baxter Equations", IWMM'92, (1992)79-89.

[WU1] Wu Wen-tsun, "On the Decision Problem and the Mechanical Theorem Proving and the Mechanization of Theorem in Elementary Geometry", Scientia Sinica 21(1978)159-172.

[WU2] Wu Wen-tsun, "Basic Principles of Mechanical Theorem Proving in Geometries" J.of Sys.Sci.and Math.Sci, (1984)207-235 .

[WU3] Wu Wen-tsun, "On Surface-Fitting Problem in CAGD", MM-Res. Preprints, No.10, (1993)1-11

[WU4] Wu Wen-tsun, "On a Finiteness Theorem about Optimization Problems" MM-Res. Preprints No.8, (1992)1-19

[WU5] Wu Wen-tsun, "On the Char-Set Method and the Linear Equations Method on Non-linear Polynomial Equations-Solving", IWMM'92 , (1992) .

[WU6] Wu Wen-tsun, "On the Generic Zero and Chow Basis of an Irreducible Ascending Set" MM-Res. Preprints No.4, (1989)

[WU7] Wu Wen-tsun, "A Zero Structure Theorem for Polynomial Equations Solving and Its Applications" , MM-Res. Preprints No.1 , (1987).

[W-H] Wu Wen-da & Yu-zhen Huang , "Kinematic Solution of A Stewart Platform", IWMM'92 , (1992)181-189 .

[ZHI] L.H.Zhi, "Factorizing Multivariate Polynomials over Algebraic Extension Fields", MM-Res. Preprints No.12, (1994)77-84.

GEOMETRY THEOREM PROVING WITH EXISTING TECHNOLOGY*

Dongming Wang

LIFIA–IMAG, 46, avenue Félix Viallet, 38031 Grenoble Cedex, France

Abstract. The first part of this paper examines the existing techniques and software tools for mechanical geometry theorem proving (GTP) and the second part describes the functionality of a GTP package GEOTHER implemented by the author. We give an account to early developments on GTP, followed by a review on algebraic techniques developed later. Some recent advances on the subject are reported and a list of existing geometry theorem provers is provided. A few functions and features of GEOTHER are explained in detail with illustrations.

This paper is a shortened and updated version of a project report [23]. The original report contains a long list of references, which are referred still but not reproduced here. Some examples and discussions specific to the project are eliminated.

Part I. A Brief Review

1. Early Developments

Although the idea of mechanizing proofs of geometric theorems may be traced back to R. Descartes and G. W. Leibniz, we begin our account to early developments of GTP around the turn to this century. According to W. T. Wu, one of the earliest theoretical contributions to the mechanization of geometry is due to D. Hilbert, who gave a method for proving a certain class of pure point of intersection theorems in his book "Grundlagen der Geometrie" published in 1899. However, Hilbert's method had never been recognized before Wu pointed out its merit (Wu, 1982a, 1983). This method was extended later by the author to a wider application domain (Wang, 1987, 1989). On the other hand, in 1930 A. Tarski discovered a general decision method for the theory of elementary algebra and geometry. Tarski's method, published later in 1948, provides a remarkable positive result on the mechaniziablity problem for a large domain. Subsequently, A. Seidenberg (1954) and P. J. Cohen (1969) proposed other methods for the same decision problem of elementary algebra and geometry. These methods are considered of high complexity and have not yet been implemented on any computer up to the present. It was G. E. Collins who discovered in 1975 another method based on cylindrical algebraic decomposition (CAD) which deals with the same decision problem as Tarski's but is practically more efficient.

H. L. Gelernter (1959a) was the first who realized a program for proving geometric theorems on a modern computer. The program developed by Gelernter, called Geometry Machine, ran on an IBM 704 and proved geometric theorems mechanically for the first time in early spring 1959. This program was extended and improved to prove a good number of theorems; further work in this direction was reported in Gelernter et al. (1960), Gelernter (1963), Gilmore (1970),

*Supported by FWF and CEC under ESPRIT BRP 6471 (MEDLAR II).

Goldstein (1973), Nevins (1975), Elcock (1977) and other relevant papers of these authors. Nevertheless, the axiomatic approaches along with the line of Gelernter are hardly possible to reach the desired altitude of proving and discovering non-trivial theorems. Although various strategies were subsequently adopted (see the comprehensive study of Coelho and Pereira, 1986), the problem of searching space amongst the others still remains and makes the method highly inefficient.

In 1969, E. Cerutti and Ph. Davis applied a symbolic manipulation system, called FORMAC, to prove theorems in elementary geometry, following the analytic method of R. Descartes. They proved the well-known theorem of Pappus, pointed out the way of attacking Pascal's theorem and obtained a couple of new theorems by inspecting the machine printout. Also contained in Cerutti & Davis (1969) are interesting discussions on the aspects of machine proofs. This work is very much in the spirit of the algebraic approaches to GTP to be discussed in the following section. Unfortunately, the techniques of Cerutti and Davis were not further developed to constitute a general GTP method.

2. The Success of Algebraic Approaches

A breakthrough of theorem proving in geometry is marked by the work of Wu (1978a). Wu observed that most of the fundamental geometric relations can be expressed by means of polynomial equations via algebraization and coordinatization. Thus, proving most of the theorems in geometries can be reduced to manipulating the corresponding algebraic relations, for which he developed a powerful method based on J. F. Ritt's concept of characteristic sets (CS). Wu's method is capable of proving hundreds of non-trivial theorems in various elementary geometries and differential geometries, leading to researcher's renewed interest in the subject. Wu's work has become widely known outside China since 1984 when a book edited by Bledsoe and Loveland (1984) was published. This book reprints Wu's 1978 paper and contains a notable paper by Chou (1984) in which he described a prover based on Wu's method that proved 130 theorems already at that time. Awareness of Wu's method in the West motivated researchers to apply the well-known method of Gröbner bases (GB; Buchberger, 1985) to prove the same class of theorems that Wu's method addresses. In 1986 several papers (Chou & Schelter, 1986; Kapur, 1986a-b; Kutzler & Stifter, 1986a-c) on the application of GB to GTP were published. See also Carrà Ferro & Gallo (1987), Hussain, Drew & Noble (1986), Ko (1985a, 1988), Ko & Hussain (1985a) and Winkler (1988, 1990, 1992). Researchers made extensive experiments with comparisons on Wu's method and the GB method for GTP around that time. A large number of non-trivial theorems were proved by different provers on different machines (see Chou, 1988a; Ko & Hussain, 1985b; Kutzler & Stifter, 1986f). Among them the most gigantic scale of experiments was carried out by Chou (1988a).

Meanwhile, Wu as well as his students in Beijing continued improving his method and applying it to prove and discover non-trivial theorems (Gao, 1987; Wang & Hu, 1987; Wu, 1987d). Relying upon the limited computing facilities and access to high level software systems, each member of Wu's group wrote his own Fortran code for polynomial manipulation and GTP. They proved more than hundred theorems (Wang & Gao, 1987) and also discovered several non-trivial ones (Wu & Lu, 1985; Wu, 1986b; Wang, 1989a).

Wu's method also stimulated the discovery of another type of methods — proving by examples — for geometric theorems (Hong, 1986a-b; Wang, 1988), and found applications in formula derivation (Chou, 1987; Wu, 1986b, 1987b), geometric computation (Brüderlin, 1988; Mundy, 1986; Swain & Mundy, 1986) and on many other subjects beyond geometry. The method was also applied to prove theorems in geometries besides plane Euclidean geometry (Wang, 1987; Chou & Ko, 1986; Gao, 1989) and extended to differential geometry (Wu, 1978b, 1979, 1982b, 1987a). Procedures based on GB were developed to prove theorems in differential geometry, too

(Carrà Ferro & Gallo, 1990). Various applications of GB to geometric problems were reported in Buchberger (1987).

Wu's method and the GB method mainly deal with geometric theorems whose algebraic formulations only involve polynomial equations and inequations. When inequalities are involved, these two methods may be inapplicable. A general quantifier elimination method and thus a decision method for the theory of elementary algebra and geometry based on CAD was discovered by G. E. Collins (1975). Collins' method has been subsequently improved and can be applied to a wide class of problems in geometry and algebra with reasonable efficiency (cf. Arnon & Buchberger, 1988). The method has been applied to GTP by D. S. Arnon (1986).

3. Recent Advances

We consider the time from later 1950's through middle 1970's, and from later 1970's through 1988 as the first two exciting periods for GTP. There have been many advances on the subject after 1988. Some of them are reported in this section.

3.1. Extensions, Improvements and Applications of the Existing Methods

The algebraic methods of CS, GB and CAD are being improved constantly. A large number of theoretical and practical improvements of these methods are made through the research activities on the methods themselves, for which we are not able to mention one by one. See Wang (1991a) and references therein for more information. For examples of special improvements directed to GTP, we may mention Chou & Gao (1989b, 1990c-e), Gao (1990), Wang (1990) and Wu (1992a).

Wu's method has been further applied and extended to prove theorems in space geometry (Wang, 1990, 1992) and finite geometry (Lin & Liu, 1992, 1993), and to derive formulas (Chou & Gao, 1990a; Gao & Wang, 1992). It has been combined with Collins' method for proving theorems involving inequalities (Chou et al., 1989, 1992 and [19, 20]). Wu's method of proving theorems in differential geometry has also been further studied and improved (Chou & Gao, 1989a,c, 1990e, 1991, 1993; Li, 1991; Wu, 1991, 199? and [6, 7]). The dimension method of Carrà Ferro and Gallo for proving theorems in differential geometry has been extended in [5]. A formalism based on the complex numbers with conjugation and GB computation for GTP is proposed in [22]. See [16] for a set of mechanically proved theorems in real geometry. The reasoning about a number of selected geometric problems using Wu's method, Buchberger's GB method and Collins' CAD method are explained in Wang (1991a). Some general logical deduction methods have been applied to prove geometric theorems in various cases (Quaife, 1989; Balbiani & Fariñas del Cerro, 1992; Balbiani et al., 1993 and [1, 2, 3, 4]).

3.2. Development of New Methods

Methods Using Coordinates. Some new methods via coordinatization have been proposed for GTP. One of them, developed by L. Yang, J. Z. Zhang and their co-workers (Deng, 1989; Yang, 1989, 1991; Yang et al., 1992; Zhang & Yang, 1989; Zhang et al, 1990), follows the idea underlying the single-instance method of J. W. Hong, in which numerical verification is incorporated. The new method uses multiple-instances (m examples) with parallel numerical verification, where m is calculated from the degrees of the hypothesis polynomials.

Examples of methods merely using algebraic manipulation are those proposed in Kalkbrener (1992), Wang (1992d) and [26, 18, 28, 29, 30, 31]. Kalkbrener's method is based on a generalized Euclidean algorithm which computes unmixed-dimensional decompositions of algebraic varieties with no need of polynomial factorization. The underlying idea of this method also appears in the method presented in [28, 30, 31]. Resultant computation is used for the methods suggested in [29, 18]. The reader may consult the above-mentioned references for details of these method.

The author's method, based on some elimination procedures (Wang, 1992c, 1993), is described in detail in Wang (1992d) (for elementary geometry) and [26] (for differential geometry). In this method as well as our implementation of Wu's method, we make use of polynomial factorization over algebraic extension fields. This is appropriate in particular when the reducibility problem is involved [24]. The application of our method to reason about selected geometric problems is explained in [21, pp. 147–185].

Coordinate-Free Methods. A drawback of the algebraic approaches via coordinatization is that the constructed proofs of geometric theorems are neither readable nor interpretable. In order that the attraction of geometric theorems is not drowned by complicated algebraic expressions, one way is to create enjoyable environments for manipulating and proving theorems, either leaving out the algebraic proofs as black-box from the user or presenting them with good structure at different levels. The other is to express geometric statements in terms of other geometric entities such as distances, vectors, angles, areas and volumes, rather than point coordinates. It is feasible for some classes of geometric theorems, and in such cases the algebraic manipulations performed among the geometric entities are usually simpler. This is the so-called coordinate-free approaches. Work in this direction started in Crapo (1987), Fearnley-Sander (1989), Havel (1991), Sturmfels & White (1991) and has been highlighted recently (Richter-Gebert, 1992; Chou, Gao & Zhang, 1992, 1993; Stifter, 1993 and [8, 9, 10, 11, 12, 13, 14]). The suggested methods employ different notations and different techniques for algebraic manipulation, but some of them have obvious relevance. They are very efficient and the proofs produced by them are short and readable, in which quantities like areas and brackets have clear geometric meanings.

3.3. Cooperation between Algebraic and Logical Provers

It was illustrated in Wang (1991a) how algebraic methods can be used to support logical approaches in proving geometric theorems. Continuing in this direction, we are investigating the cooperation between algebraic provers and logical deduction systems, with the integration of GEOTHER (see Part II) and the inference laboratory ATINF (Caferra et al., 1991) as a case study. In GEOTHER, there is no capability of editing, manipulating and presenting the produced proofs. We are now trying to use the ATINF proof editor to handle the proofs generated by GEOTHER with structural and hierarchical presentation at different levels. Imprecise statements of geometric theorems due to the missing of non-degeneracy conditions are being used as doubtful assertions to test the capability of model construction in searching for counterexamples and for making the statements precise by generating the missed non-degeneracy conditions.

3.4. Construction of an Integrated System for Computer Geometry

Our examination on the state-of-the-art of GTP, the implementation of GEOTHER and the study of cooperation between algebraic and logical approaches to GTP are made towards a long-term project on the design and implementation of an integrated system for computer geometry. The system is planned to be one that integrates logical, combinatorical, algebraical, numerical and graphical algorithms with knowledge base and powerful reasoning and computing capabilities for geometry research, education and application. Such a system could be very sophisticated and will involve many fundamental algorithms and have to make use of various software tools. As the first step, we are now engaged in integrating theorem provers and computer algebra systems (cf. Hadzikadic et al., 1986; Beeson, 1989; Clarke & Zhao, 1993, [17] and other references in [23]). The design issues are being focused on the internal data structures, strong data types of geometric objects and the input/output communication with other systems. The object-oriented paradigm with abstract specification of geometric structures and objects (cf. Miola, 1990 and Limongelli & Temperini, 1992) has been taken into consideration for our system.

4. Existing Provers

In this section is provided a list of exiting geometry theorem provers with references. We have tried to include most of the important and significant provers. Some of their features are reviewed in Hong et al. (1993) and [23].

- **Geometry Machine** (Gelernter, 1959a; Gelernter et al., 1960)
- **BTP/PTP** (Goldstein, 1973)
- **GEOM** (Coelho & Pereira, 1979, 1986)
- **China Prover** (Wu, 1984)
- **GEO-Prover** (Chou, 1984, 1985, 1988a, 1992)
- **ALGE-Prover/ALGE-Prover II/GEOMETER** (Ko & Hussain, 1985b; Ko, 1986; Cyrluk et al., 1988)
- **CPS/PS/DPMS1/DPMS2** (Gao & Wang, 1987)
- **GPP** (Mayr, 1988; Kusche, Kutzler & Stifter, 1989)
- **GEOTHER** (see Part II of this paper)
- **EUCLID** (Chou, Gao & Zhang, 1992, 1993)
- **GEO** (Alberti, Lammoglia & Torelli, 1992)
- **GEOPROVER** (Kanchanasut & Nam: see Hong et al., 1993)
- **CINDERELLA** (Crapo & Richter-Gebert: see Hong et al., 1993 and Richter-Gebert, 1992)
- **G-Prover** (Gao et al., 1992)
- **Other Provers** (Hadzikadic et al., 1986; Smietanski, 1986/87; Yang et al., 1992; K. Sonmez: see Proc. 1992 Mathematica Conf.).

Part II. GEOTHER

GEOTHER (for GEOmetry THeorem provER) is a software package implemented in Maple for manipulating and proving geometric theorems. The core of GEOTHER consists of five algebraic provers based on the methods of CS and GB and an elimination method proposed by this author. Along with the implementation of this package, we keep in mind that the provers should be practically efficient and the environment be flexible and easy to be acquainted and in which the geometric fascination and attraction are retained as much as possible. In what follows we present some functions and features of the package. Implementational issues and discussions about the underlying algorithms are left for a forthcoming paper.

1. Specification via Predicates

A geometric theorem in GEOTHER is specified by means of predicates. We have reserved `Theorem` as a special predicate with `Theorem(H, C, X)` asserting that `H` implies `C`. Among the three arguments, the first is a list or a set of other predicates that correspond to the hypothesis of the theorem, the second is either a single predicate, or a list or a set of predicates that correspond to the conclusion of the theorem, and the third is optional, and if presents, is a list of variables used for the internal algebraic computation. For example, the so-called Simson theorem can be specified as

```
Simson := Theorem( [arbitrary(A,B,C), oncircle(A,B,C,D), perpfoot(D,P,A,B,P),
                    perpfoot(D,Q,A,C,Q), perpfoot(D,R,B,C,R)], collinear(P,Q,R) ).
```

In this representation, **Simson** is the name of the theorem, **arbitrary(A,B,C)** means that A, B and C are three arbitrary points, **oncircle(A,B,C,D)** the point D is on the circumcircle of the triangle ABC, **collinear(P,Q,R)** the three points P, Q and R are collinear, and so forth. The predicates like **arbitrary**, **oncircle** and **collinear** take different number of arguments and declare the corresponding geometric relations among their arguments — the points. Internally, associated with each predicate a Maple procedure that contains necessary information about the predicate is defined. The logical **not** may prefix to any predicate. At the present stage, there are about 30 predicates available in GEOTHER.

2. Assignment of Coordinates

The assignment of coordinates of points in GEOTHER can be done either artificially along with the specification by the user using the function **Let**, or automatically by the prover using the function **Coordinate**. The function **Let** may take an arbitrary number of arguments, each one of which is an equation of the form A=[x,y], where A is the name of the point and [x, y] are the coordinates to be assigned to A. For example, the coordinates to the points in the Simson theorem can be assigned by

```
Let(A=[-x1,0], B=[x1,0], C=[x2,x3], D=[x4,x5], P=[x4,0], Q=[x6,x7], R=[x8,x9]).
```

When the coordinates are assigned by **Let**, a hash table is created by the prover to store them. The coordinates of every point can be inquired by using the function **Algebraic**. The user may present a list of (preferably only the geometrically dependent) variables as the third argument to **Theorem** for the internal computation (of CS and GB). If the list does not present, a "heuristically" optimized order is chosen internally by the prover.

The function **Coordinate** allows to assign coordinates to points automatically in some optimal manner. After calling **Coordinate**, the natural order of variables presents in default as the third argument to **Theorem**. For the above specification of Simson's theorem, application of **Coordinate** to **Simson** yields

```
Theorem( [[arbitrary(A,B,C), oncircle(A,B,C,D), perpfoot(D,P,A,B,P),
          perpfoot(D,Q,A,C,Q), perpfoot(D,R,B,C,R)], collinear(P,Q,R),
         [x1, y1, y2, x2, y3, x3, y4] )
```

where the specification remains unchanged excepting that the coordinates of points are reassigned and a new list (order) of variables presents as the third argument to **Theorem**. Now the coordinates of the points A, B, C, D, P, Q and R are assigned internally as

```
[u1, 0], [u2, 0], [0, v1], [x1, y1], [x1, y2], [x2, y3], [x3, y4]
```

respectively.

3. Translation and Drawing

Having information contained in the procedure associated with each predicate, the translation of a geometric theorem in predicate representation into English and algebraic expressions is straightforward. For the former there is a function called **English** which simply puts the interpretation of each predicate together with some English conjunctives as to form a complete statement of the theorem. Application of **English** to **Simson** yields

```
Theorem: If the points A, B, and C are arbitrary, the point D is on the circum-
circle of the triangle ABC,  P is the perpendicular foot of the line DP to the
line AB,  Q is the perpendicular foot of the line DQ to the line AC,  and R is
the perpendicular foot of the line DR to the line BC,  then the three points P,
Q and R are collinear.
```

The meaning of each predicate as well as its negation can also be inquired by using `English`. For instance, `English(not parallel(A,B,P,Q))` yields

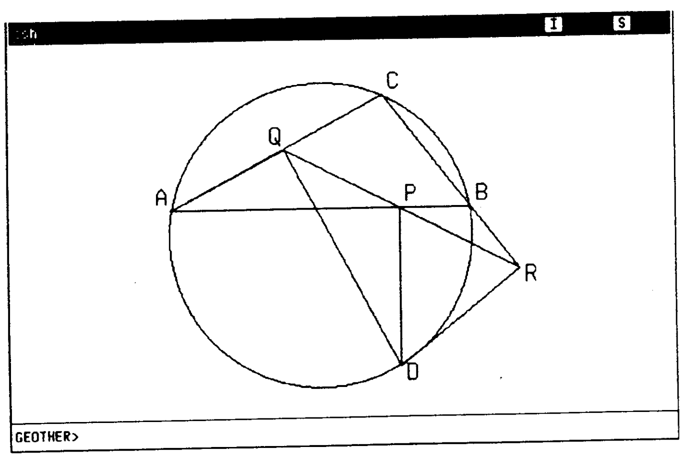
```
the line AB is not parallel to the line PQ.
```

A function called `Logic` is implemented in GEOTHER which translates any predicate representation of a theorem into a logical formula. For `Simson` specified above, `Logic(Simson)` yields

```
(Any D A B C P Q R) [ oncircle(A,B,C,D) /\ perpfoot(D,P,A,B,P) /\
perpfoot(D,Q,A,C,Q) /\ perpfoot(D,R,B,C,R) ==> collinear(P,Q,R) ].
```

The translation of theorems in predicate form to theorems represented by means of polynomials is done by a function called `Algebraic`. This function is also used to translate the individual predicates and to print the coordinates of points stored in the hash table as mentioned before.

The function `Geometric` is implemented for drawing figures from the predicate specification of a theorem. This function is composed of routines for solving the hypothesis equations with parameters specialized and drawing routines written in C, making use of Domain Graphics Primitives on Apollo workstations and Xlib on Sun SPARCstations. Special care has been taken of nice-look of the drawn figures and of labeling the letters in appropriate positions. For the above example, `Geometric(Simson)` yields



4. Algebraic Provers

The algebraic provers, which are based on the algebraic methods of CS and GB as well as a method of the author, are the kernel of our package. They contains a complete implementation of Wu's method with algebraic factorization and irreducible zero decomposition. In case of using Wu's and our elimination method, a theorem can be proved either in coarse form by providing subsidiary conditions, or in fine form by examining whether each subsidiary condition is a non-degeneracy condition indeed. A coarse form proof (sketch) of the Simson theorem by using `Wprover` (based on Wu's method) reads as follows

```
Proof:

char set:            [6 x1 2], [10 x2 1], [6 x3 1], [10 x4 1], [6 x5 1]
pseudo-remainder:    [4 x9 1] = -x4*x9   + ... (3 terms)
pseudo-remainder:    [8 x8 1] = -x2*x7*x8   + ... (7 terms)
pseudo-remainder:    [23 x7 1] = x2**2*x3*x5*x7   + ... (22 terms)
pseudo-remainder:    [17 x6 1] = x2**2*x6*x1   + ... (16 terms)
pseudo-remainder:    [12 x5 2] = 2*x2*x3**2*x5**2*x1   + ... (11 terms)
```

```
        The theorem is true under the following subsidiary condition(s):

.    the line BC is non-isotropic
.    the line AC is non-isotropic
.    the three points A, B and C are not collinear
.    the line AC is not perpendicular to the line AB
.    the line AB is not perpendicular to the line BC

QED.
```

The prover based on Wu's method has been parallelized and can run with parallel processors via workstation networks (cf. Wang, 1991b).

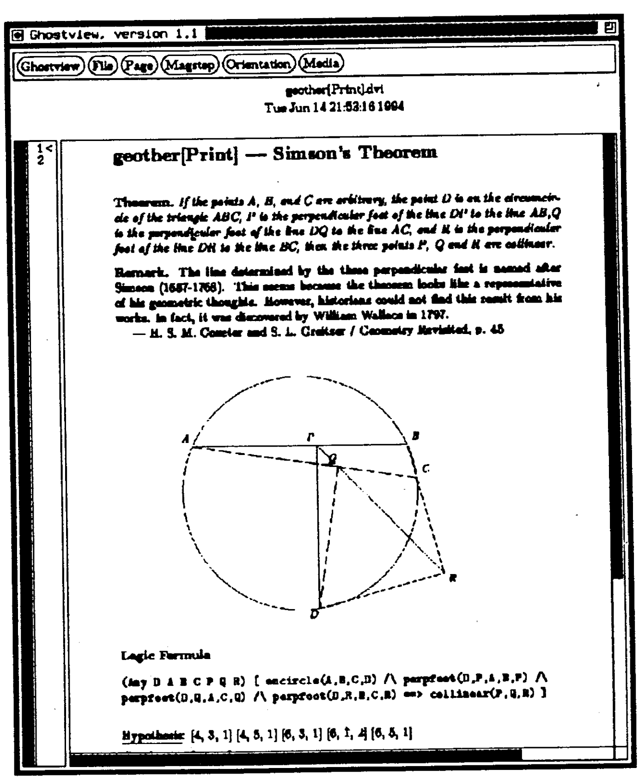
5. Predicate Form of Non-Degeneracy Conditions

From the preceding `Proof` one sees that the subsidiary conditions are provided in geometric/predicate form (English). They are generated from algebraic expressions by a function called `Generic`. `Generic` makes use of the information contained in the predicate definitions and a translator that translates algebraic expressions into geometric statements by heuristic search. The translator works for the algebraic expressions which correspond to some fundamental geometric relations and is rather fast. Considering for example the above Simson theorem, `Generic(Simson,x2-x1)` yields

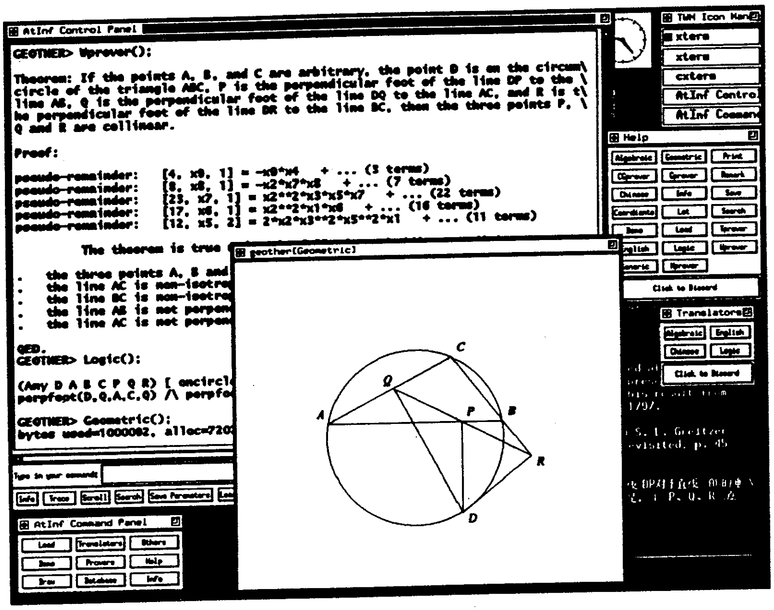
```
.    the line AB is not perpendicular to the line BC
                          not perpendicular(A, B, B, C)
```

6. Interface

GEOTHER has a function called `Print` which interfaces with LaTeX and uses Ghostview to preview a PostScript file documenting the theorem and its proof. The file is generated from a hash table which maintains useful data associated with the theorem produced in the previous manipulation and computation. The following is an example of the document viewed on the screen.



GEOTHER may run with the ATINF graphic interface implemented by M. Desbrun. Using this interface, all the GEOTHER commands can be entered with menu-driver. Only occasionally, one needs to use keyboard such as entering the name of a theorem to be loaded. A GEOTHER session with graphic interface is shown below.



7. Library and Help Facility

A library of geometric theorems with standardized specifications is being created. A couple of functions are implemented for handling the library, for example, searching for theorems in the library, loading theorems from the library into a GEOTHER session and adding theorems to the library. The package contains on-line help for all functions, explaining how to use them with examples.

References

The references cited in the text with authors' names plus years are listed in [23]. The following is a supplement to that list.

[1] Balbiani, Ph. Equation solving in projective planes and planar ternary rings. *Rapport Irit/93-31-R*, IRIT, Université Paul Sabatier (1993).

[2] Balbiani, Ph., Dugat, V., Fariãs del Cerro, L. and Lopez, A. *Eléments de géométrie mécanique.* Hermès, Paris (1994).

[3] Balbiani, Ph., Dugat, V. and Lopez, A. Le raisonnement sur les figures de la géométrie projective. *Rapport Irit/94-5-R*, IRIT, Université Paul Sabatier (1994).

[4] Balbiani, Ph. and Lopez, A. Simplification des figures de la géométrie affine plane d'incidence. In: *Neuvième congrès reconnaissance des formes et intelligence artificielle* (Paris, January 1994), pp. 341–351.

[5] Carrà Ferro, G. An extension of a procedure to prove statements in differential geometry. *J. Automated Reasoning* **12**, 351–358 (1994).

[6] Chou, S. C. and Gao, X. S. Automated reasoning in differential geometry and mechanics using characteristic method – III. In: *Automated Reasoning* (Z. Shi, ed.), pp. 1–12. North-Holland: Elsevier Science Publ. (1992).

[7] Chou, S. C. and Gao, X. S. Automated reasoning in differential geometry and mechanics using the characteristic method – IV. *Syst. Sci. Math. Sci.* **6**, 186–192 (1993).

[8] Chou, S. C., Gao, X. S. and Zhang, J. Z. Automated production of traditional proofs for constructive geometry theorems. In: *Proc. 8th IEEE Symp. Logic Comput. Sci.*, pp. 48–56 (1993).

[9] Chou, S. C., Gao, X. S. and Zhang, J. Z. The area method and affine geometries over any fields. *MM Research Preprints* **10**, 92–100 (1993).

[10] Chou, S. C., Gao, X. S. and Zhang, J. Z. *Machines Proofs in Geometry.* Singapore: World Scientific (1994).

[11] Chou, S. C., Gao, X. S. and Zhang, J. Z. A rule based geometry theorem prover using full angles. *MM Research Preprints* **11**, 72–103 (1994).

[12] Chou, S. C., Gao, X. S. and Zhang, J. Z. Vectors and automated geometry reasoning. *MM Research Preprints* **11**, 104–127 (1994).

[13] Chou, S. C., Gao, X. S., Yang, L. and Zhang, J. Z. Automated production of readable proofs for theorems in non-Euclidean geometries. *Tech. Rep. TR-WSU-94-9*, Wichita State Univ. (1994).

[14] Chou, S. C., Gao, X. S., Yang, L. and Zhang, J. Z. A collection of 90 automatically proved geometry theorems in non-Euclidean geometries. *Tech. Rep. TR-WSU-94-10*, Wichita State Univ. (1994).

[15] Grünbaum, B. and Shephard, G. C. From Menelaus to computer assisted proofs in geometry. *Preprint* (1993).

[16] Guergueb, A., Mainguené, J. and Roy, M.-F. Examples of automatic theorem proving in real geometry. In: *Proc. ISSAC '94* (Oxford, June 1994).

[17] Homann, K. and Calmet, J. Combining theorem proving and symbolic mathematical computing. In: *Artificial Intelligence and Symbolic Mathematical Computing* (J. Calmet and J. A. Campbell, eds.). LNCS (to appear 1994).

[18] Kapur, D., Saxena, T. and Yang, L. Algebraic and geometric reasoning using Dixon resultants. In: *Proc. ISSAC '94* (Oxford, June 1994).

[19] McPhee, N. F. *Mechanically Proving Geometry Theorems Using Wu's Method and Collins' Method.* Ph.D thesis, Univ. of Texas at Austin (1993).

[20] McPhee, N. F., Chou, S. C. and Gao, X. S. Mechanically proving geometry theorems using a combination of Wu's method and Collins' method. In: *Proc. CADE-12*, LNAI **814**, 401–415 (1994).

[21] Pfalzgraf, J. and Wang, D. M. (eds.). *Automated Practical Reasoning: Algebraic Approaches.* Springer-Verlag, Wien-New York (1995).

[22] Stokes, T. Conjugate polynomials and automated geometry theorem proving. *Preprint* (1993).

[23] Wang, D. M. Geometry theorem proving with existing technology. *MEDLAR II 12-Month Deliverables*, Dept. of Computing, Imperial College, Univ. of London (1993). Also available as RISC-Linz Report Series No. 93-40, Johannes Kepler Univ. (1993).

[24] Wang, D. M. Algebraic factoring and geometry theorem proving. In: *Proc. CADE-12*, LNAI **814**, pp. 386–400 (1994).

[25] Wang, D. M. Some advances in geometric and algebraic reasoning. *MEDLAR II 24-Month Deliverables*, Dept. of Computing, Imperial College, Univ. of London (1994).

[26] Wang, D. M. A method for proving theorems in differential geometry and mechanics. *Preprint*, LIFIA - Institut IMAG (1995).

[27] Wu, W. T. *Mechanical Theorem Proving in Geometries: Basic Principles* (translated from Chinese by X. Jin and D. Wang). Springer-Verlag, Wien-New York (1994).

[28] Yang, L. and Zhang, J. Z. Searching dependency between algebraic equations: an algorithm applied to automated reasoning. In: *Artificial Intelligence in Mathematics* (J. Johnson, ed.). Oxford: Oxford Univ. Press (1994).

[29] Yang, L., Zhang, J. Z. and Hou, X. R. A note on Wu Wen-tsün's nondegenerate condition. *Chinese Sci. Bull.* **38**, 86–87 (1993).

[30] Yang, L., Zhang, J. Z. and Hou, X. R. An efficient decomposition algorithm for geometry theorem proving without factorization. *MM Research Preprints* **9**, 115–131 (1993).

[31] Yang, L., Zhang, J. Z. and Hou, X. R. A criterion for dependency of algebraic equations, with applications to automated theorem proving. *Sci. China* Ser.**A** **37**, 547–554 (1994).

A NEW APPROACH OF TEACHING CALCULUS USING MAPLE

Xiaohui Zhong
Alberta Knowlton
Department of Mathematics and Computer Science
University of Detroit Mercy
Detroit, Mi 48221, USA
zhongk@udmercy.edu

Introduction

Calculus is an old and well-developed discipline that is required for all students majoring in engineering and science. Traditional instruction of Calculus consisted of the introduction of concepts followed by examples involving complicated algebraic manipulation. As a result, the reasoning skills necessary in problem solving were often overshadowed by this complex computational work. This lengthy training usually became an obstacle for many students to continue studying mathematics and engineering. Modern technology such as the computer algebraic system (CAS) Maple gives instructors and students a new tool to change this traditional teaching pedagogy. The use of Maple in classrooms will stimulate students' interest through a more exciting presentation by reducing the time spent on long and tedious drills and practices. Exercises in Maple will provide students an opportunity to supplement and reinforce the course content outside classrooms. In addition, it will allow the course emphasis to shift from a focus on laborious hand calculation to the development of students' analytical skills.

This paper describes how the University of Detroit Mercy, an urban university in the USA, addressed and solved the problems associated with conventional mathematical teaching methods. UDM has introduced Maple to its Calculus curriculum since Fall of 1993. This well received dynamical change is illustrated by four typical lectures from the calculus sequence. We then discuss both the advantages and problems associated with the implementation, as well as the anticipated solutions.

1. Intuitive concept of vertical asymptotes

Students start using Maple in the first course of the calculus sequence, Functions and Graphs. Topics in this course include coordinates and graphs, the distance formula, circles, formal definition of functions, function composition and inverses, as well as linear, quadratic, exponential, logarithmic and trigonometric functions. It intends to give a clear concept of functions through graphing and to prepare students for the intensive work on calculus. Maple's powerful graphical capabilities enable instructors and students avoid involving themselves from the lengthy, and often inaccurate, process of hand graphing each function. This Maple session introduces the concept of vertical asymptotes to students by intuition and pattern recognition. By answering questions prepared by the instructor, students are led to explore the behavior of rational functions near the singular points. Students will develop a more comprehensive understanding of the concepts and properties of functions through lectures and exercises similar to this one.

```
> n:=x->1;
```

$$n := 1$$

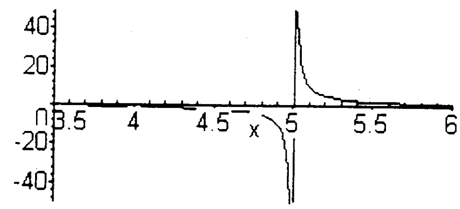
```
> d:=x->x-5;
```

$$d := x \rightarrow x - 5$$

```
> g:=x->n(x)/d(x);
```

$$g := x \rightarrow \frac{\mathrm{n}(x)}{\mathrm{d}(x)}$$

```
> plot(g(x), x=3.5..6, y=-50..50);
```




```
> g(4.9),g(4.99),g(4.999),g(4.99999), g(4.999999);
```

$$-10.00000000, -100.0000000, -1000.000000, -100000.0000, -.1000000000\ 10^7$$

$$-2.000000000, -10.00000000, -100.0000000, -1000.000000, -100000.0000, -.1000000000\ 10^7$$

```
> #Q: What are the function values approching as x gets close to 5 from the left?
> g(5.01), g(5.001), g(5.0001), g(5.00001),g(5.000001),g(5.0000001);
```

$$100.0000000, 1000.000000, 10000.00000, 100000.0000, .1000000000\ 10^7, .1000000000\ 10^8$$

```
> #Q: What are the function values approching as x gets close to 5 from the right?
> #Q:  How will these values affect the shape of the graph?
> #Q:  Is the line x=5 a part of the graph? Explain.
> solve(d(x)=0, x);
```

$$5$$

```
> # 5 is the location of the vertical asymptote, the equation of this asymptote is  x=5.  By solving d(x)=0, you
> # will find the location of the asymptote.
```

2. Geometric and mathematical definition of definite integral.

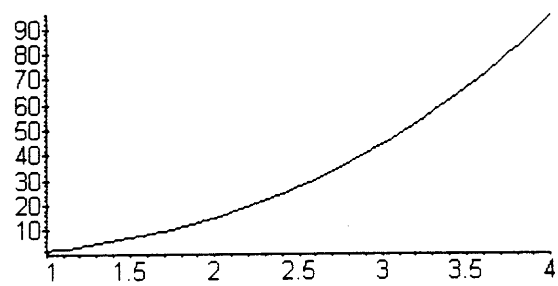
Calculus I is the first course where students are introduced the concepts of limits, derivatives and integrals, as well as the computational techniques. However, the emphasis on mechanical skills leaves the student lacking sufficient understanding of these important concepts. Maple's powerful graphical and computational abilities will present these concepts to students in a straightforward, step-by-step manner, without getting bogged down in drawn-out calculations.

This particular Maple session shows an intuitive approach to introduce the concept of the definite integral. Maple relieves the student of the necessity of graphing the function himself, yet visually reinforces the relationship between the function and its graph. Computer graphics also allows the student to see more clearly how the height and width of the rectangles drawn on the subintervals are measured. Although Maple contains a function which calculates the sum of the areas of the rectangles, the instructor can show students how to calculate the height, width and area of each rectangle using Maple's basic function definition and thereby derive the formula on their own. Students can then see how the results from the derived formula compare with those obtained through Maple's built-in function. The student may also dynamically change the number of rectangles under the curve and compare each successive sum of areas. This process leads naturally from the idea of the limit of the sum of the areas of retangles to the notion of an integral.

```
> f:=x->x^3+2*x^2-1;
```

$$f := x \rightarrow x^3 + 2x^2 - 1$$

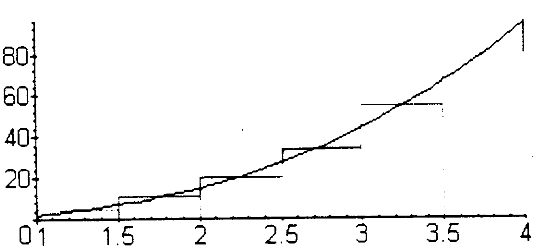
```
> plot(f(x),x=1..4);
```




```
> with(student), with(plots):
> # We want to find the area between this curve and the x-axis for the given interval. The following command
> # will draw a series of rectangles under the  curve. Each rectangle has as a height the function f(x)
> # evaluated at the midpoint of each sub-interval, while the width of each rectangle is given by the width of
>  # the entire interval divided by the number of sub-intervals, n .
> a:=n->middlebox(f(x), x=1..4, n);
```

$$a := n \rightarrow \text{middlebox}(\text{f}(x), x = 1 .. 4, n)$$

```
> a(6);
```




```
> # The width of each interval is
> w:=n->(4-1)/n;
```

$$w := n \rightarrow 3 \frac{1}{n}$$

```
> # The height of  the ith rectangle is
> h:=(`n`,`i`)->f(1+`i`*w(`n`)-1/2 *w(`n`));
```

$$h := (n, i) \rightarrow \text{f}\,1 + i\,\text{w}(n) - \frac{1}{2}\,\text{w}(n)$$

```
> # The area of the ith rectangle is
> b:=(`n`,`i`)->w(`n`)*h(`n`,`i`);
```

$$b := (n, i) \rightarrow \text{w}(n)\,\text{h}(n, i)$$

> # The area of all rectangles is
> area:=`n`->sum(b(`n`,`i`), `i`=1..`n`);

$$area := n \to \sum_{i=1}^{n} b(n, i)$$

> # Maple also has a fomula to define this area of the sum of all rectangles
> A:=n->middlesum(f(x),x=1..4,n);

$$A := n \to \text{middlesum}(\, f(x), x = 1 \,..\, 4, n\,)$$

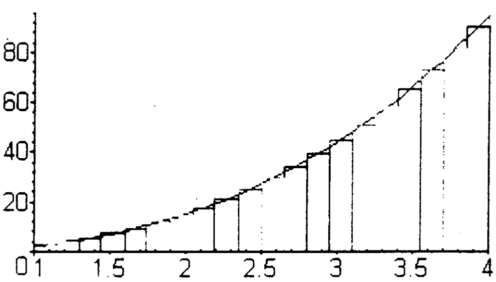
> evalf(area(6));

102.1562500

> evalf(A(6));

102.1562500

> a(20);



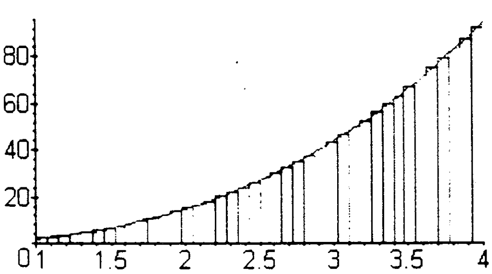
> evalf(A(20));

102.6965625

> evalf(area(20));

102.6965625

> a(40);



> evalf(A(40));

102.7366406

> evalf(area(40));

102.7366406

> #Q: Do you think A(n) and area(n) are the same? Why? What number are A(n) and area(n) approaching?
> #What does this number represent? What kind of process is this? Let's take the limit of area(n) and A(n)

```
> evalf(limit(area(n), n=infinity));
```

$$102.7500000$$

```
> evalf(limit(A(n), n=infinity));
```

$$102.7500000$$

```
> # This number is the area between the curve f and  the x-axis for x between  1 and 4,  it is called the definite
> # integral of  f(x) for x between 1 and 4.
> # Maple's command to find the definite integral is
>  int(f(x), x=1..4);
```

$$\frac{411}{4}$$

3. Precise mathematical concept of convergence.

Calculus II continues with more topics such as sequence, series, etc. Students already have some intuitive ideas about limits and convergence. However, the precise mathematical concept of covergence is still vague for most of students. A good place to introduce this concept is in studying series. The amount of calculation necessary before the pattern of convergence becomes clear can be prodigious, and opportunities for error numerous. Fortunately, Maple provides students a way to see the process of convergence without first engaging in a protracted computational exercise. Maple can quickly generate a sequence of values of partial sums from which the student may deduce the converging value. The instructor can then lead the student to make the connection between a series converging value and its limit by having Maple calculate the latter. The student may also have Maple compute the distance between partial sum of the series and the limit. By observing the behavior of this distance, students could provide the definition of convergence in thier own words.

```
> nsum:=n->sum(x^i, i=1..n);
```

$$nsum := n \rightarrow \sum_{i=1}^{n} x^i$$

```
> nsum1:=n->subs(x=1/2, nsum(n));
```

$$nsum1 := n \rightarrow \mathrm{subs}\, x = \frac{1}{2}, \mathrm{nsum}(n)$$

```
> seq(evalf(nsum1(j)), j=1..20);
```

.5000000000, .7500000000, .8750000000, .9375000000, .9687500000, .9843750000, .9921875000, .9960937500, .9980468750, .9990234375, .9995117188, .9997558594, .9998779297, .9999389648, .9999694824, .9999847412, .9999923706, .9999961853, .9999980927, .9999990463

```
> seq(evalf(nsum1(j)), j=21..50);
```

.9999995232, .9999997616, .9999998808, .9999999404, .9999999702, .9999999851, .9999999925, .9999999963, .9999999981, .9999999991, .9999999995, .9999999998, .9999999999, .9999999999, 1.0000000000, 1.0000000000, 1.0000000000, 1.0000000000, 1.0000000000, 1.0000000000, 1.0000000000, 1.0000000000, 1.0000000000, 1.0000000000, 1.0000000000, 1.0000000000, 1.0000000000, 1.0000000000, 1.0000000000, 1.0000000000

```
> #Q: What is this sequence of partial sums approaching?
> limit(nsum1(n), n=infinity);
```

$$1$$

```
> # Let's consider the "distance" between each partial sum and 1
> d:=n->abs(nsum1(n)-1);
```

$$d := n \rightarrow \text{nsum1}(n) - 1$$

```
> seq(d(n), n=4..15);
```

$$\frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \frac{1}{256}, \frac{1}{512}, \frac{1}{1024}, \frac{1}{2048}, \frac{1}{4096}, \frac{1}{8192}, \frac{1}{16384}, \frac{1}{32768}$$

```
> # Q: What is the distance approaching? Can the distance be made as small as possible? How? Can d be    #
>  made smaller than 0.05, 0.005, 0.0005?
> # We can try to solve the inequality
> solve(d(n)<0.05, n);
  Error, (in unknown) unable to order boundaries of equation

> # Maple does not have the ability to solve this inequality. However, we can solve this inequality recursively
> num:=0;
```

$$num := 0$$

```
> for j from 1 while d(j)>0.05 do num:=num+1 od;
```

$$num := 1$$

$$num := 2$$

$$num := 3$$

$$num := 4$$

```
> # Use your own words to develop the definition of a partial sum converging  to a certain number.
```

4. Level curves of 3-D surfaces.

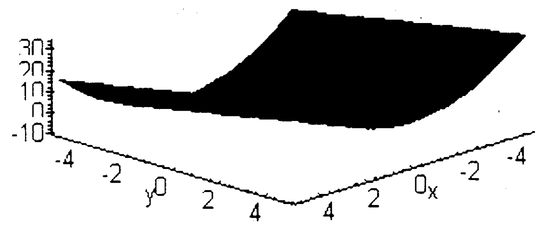
Calculus III is the last course in the Calculus sequence. Two or more variable calculus are taught here. One of the important concepts studied is the level curves of surfaces, which is extremly useful for engineering students. This Maple session demonstrates the procedure of drawing a 3-D graph. By invoking the plots package, special plot commands such as 3-D surface plot, 3-D curve plot and countour plot can be utilized. The results of these commands produced clear, accurate three-dimensional graphs of a surface, a surface cut by a plane, and the trace of that cut. This application demonstrates one of Maple's most significant strengths: the ability to quickly produce a three-dimensional drawing which would be difficult or impossible to accurately draw on a chalk board. From these pictures, the student is easily lead to develop the concept of level curves.

```
> f:=(x,y)->x^2+2*y;
```

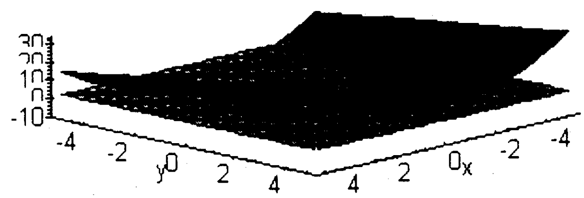
$$f := (x, y) \rightarrow x^2 + 2\,y$$

```
> with(plots):
> A:=plot3d(f(x,y), x=-5..5, y=-5..5, color=red):
> B:=plot3d(2, x=-5..5, y=-5..5, color=green):
> y2:=solve(f(x,y)=2,y);
> C:=spacecurve([x, y2, 2], x=-5..5, color=yellow):
```

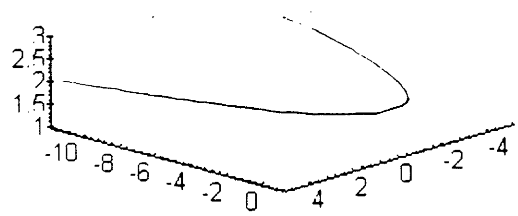
> display(A);



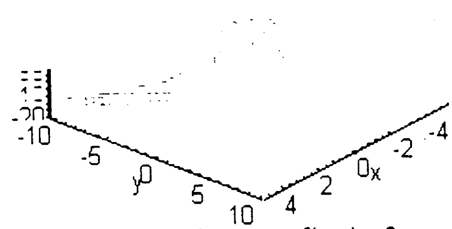
> display({A,B});



> display(C);



> #Q: What is the yellow curve? This curve is the trace where the plane z=2 cuts the surface f(x,y). What is
> #the equation of this curve when it is projected to the xy plane? This curve is called the level curve of f(x,y)
> #at level z=2. The following command displays a family of level curves of f(x,y).
> contourplot(f(x,y),x=-5..5,y=-10..10);



> #In general, a level curve of f(x,y) at level z=z0 is the curve f(x,y)=z0.

5. Conclusion

These examples have demonstrated how, by using Maple, the whole curriculum can be dynamically restructured to fit the needs of conventional education. Using a CAS is a first step in calculus reform at University of Detroit Mercy. Looking for or writing a more technologically oriented textbook is the next phase of our implementation. Finally, building an interactive computer classroom utilizing Maple and other advanced technologies is our ultimate goal.

As usual as in the development of software, it is inevitable that there is some room for Maple to perfect to a more user-friendly, easily mastery CAS. The following is a list of suggested enhancements which we will communicate to the Maple software developers. Incorporating these features into the existing software would let students spend less time learning the software and more time using it.

- Command structure could be simplified so that it associates more with the spoken language and mathematics language.
- Error messages could be more specific and clearly stated
- An enhanced editing or word processing feature could be incorporated so that a user can edit the worksheet easily.
- A user should be able to retrieve a worksheet and its calculations results without executing the commands again.

Overall, the power and versatility of Maple outweigh any inconveniences. Through using Maple in and out of the classroom, calculus can be taught in a more dynamic, interactive manner than by traditional methods. In addition, it provides the students with a tool for building analytical, rather than just computational, skills.

ON THE DERIVATIVES OF A FAMILY OF ANALYTIC FUNCTIONS

O. P. AHUJA

Division of Mathematics

School of Science

Nanyang Technological University

Singapore 1025

AND

M. JAHANGIRI

Department of Mathematics

Kent State University

Burton, Ohio 44021

ABSTRACT. For $\alpha < 1$ and $n = 0, 1, 2, \ldots$, we let $M_n(\alpha)$ denote the family of functions $f(z) = z + \ldots$ that are analytic in the unit disk and satisfy there the condition $Re(D^n f)' > \alpha$, where $D^n f(z)$ is the Hadamard product or convolution of f with $z/(1-z)^{n+1}$. We find extreme points and then determine sharp lower bounds on $Re\{z^{-1}(z^n f'(z))^{(n-1)}\}$. We also show that the partial sums of the functions in $M_n(\alpha)$, $1 \le n \le 4$, are contained in $M_{n-1}(\beta)$ for $\beta = \beta(n, \alpha) \le \alpha$. Finally we find $\beta = \beta(n, \alpha) \ge \alpha$ so that the convolution of any two functions in $M_n(\alpha)$ belongs to $M_n(\beta)$.

KEYWORDS AND PHRASES. Analytic functions, convex hull, extreme points, Hadamard product, partial sums.

1991 AMS SUBJECT CLASSIFICATION CODES. Primary 30C45; Secondary 30C50

1. INTRODUCTION.

Let $\mathcal{A}$ denote the family of functions f of the form

$$f(z) = z + \sum_{n=2}^{\infty} a_k z^k \tag{1.1}$$

that are analytic in the unit disk $U = \{z : |z| < 1\}$. Denote by $M_n(\alpha)$, $\alpha < 1$, $n = 0, 1, 2, \ldots$, the subfamily of $\mathcal{A}$ consisting of functions f of the form (1.1) for which $Re(D^n f)' > \alpha$ in U. Here $D^n f(z) = z(z^{n-1}f(z))^{(n)}/n! = f(z) * (z/(1-z)^{n+1})$, see Ruscheweyh [9]. The operator $*$ stands for the Hadamard product or convolution of two power series $f(z) = \Sigma_{k=0}^{\infty} a_k z^k$ and $g(z) = \Sigma_{k=0}^{\infty} b_k z^k$, that is, $(f * g)(z) = f(z) * g(z) = \Sigma_{k=0}^{\infty} a_k b_k z^k$. It is obvious that $M_n(\beta) \subset M_n(\alpha)$ if $\beta > \alpha$. We also know that $M_{n+1}(\alpha) \subset M_n(\alpha)$, c.f., e.g. [7]. Alexander [2] showed that $M_o(0)$ is a subfamily of analytic univalent functions. Singh and Singh [11] proved that the functions in $M_1(0)$ are starlike in U. Note that the functions in $M_n(\alpha)$ for $\alpha < 0$ need not be univalent in U. Ahuja [1] found coefficient bounds, convolution invariance, and distortion theorems for $M_o(\alpha)$, $0 \leq \alpha \leq 1$. Silverman [10] found extreme points of $M_1(\alpha)$ and showed that for $f \in M_1(\alpha)$, the partial sums $S_m(z, f)$ satisfy $Re(S_m(z, f))' > \alpha$. Ali [3] and Silverman [10] independently showed that for $f \in M_1(\alpha)$, $Ref'(z) > 2\alpha - 1 + 2(1-\alpha)log\, 2$. In this note we extend most of their results to more general case $M_n(\alpha)$. Furtheremore, we find $\beta = \beta(n, \alpha) \geq \alpha$ so that for f and g in $M_n(\alpha)$, their convolution is in $M_n(\beta)$. This is an improvement to the result obtained by Singh and Singh [12] for $M_1(0)$.

2. MAIN RESULTS.

THEOREM 1. The extreme points of $M_n(\alpha)$ are

$$f_x(z) = z + 2(1-\alpha) \sum_{k=2}^{\infty} \frac{n!(k-1)!}{(k+n-1)!k} x^{k-1} z^k \ , \quad |x| = 1 \ , \ z \in U. \tag{2.1}$$

PROOF. From the definition of $M_n(\alpha)$ it follows that $f \in M_n(\alpha)$ if and only if $D^n f \in M_o(\alpha)$. Therefore the operator D^n is a linear homeomorphism from $M_o(\alpha)$ to $M_n(\alpha)$ and thus preserves extreme points. Hallenbeck [6] (page 288) showed that the extreme points of $clcoM_o(\alpha)$ are given by

$$(2\alpha - 1)z - 2(1-\alpha)\bar{x} log(1 - xz) \ , \quad |x| = 1 \ , \ z \in U,$$

which can be written as

$$z + 2(1-\alpha) \sum_{k=2}^{\infty} \frac{1}{k} x^{k-1} z^k \ , \quad |x| = 1 \ , \ z \in U.$$

Also note that

$$D^n f(z) = \left(z + \sum_{k=2}^{\infty} a_k z^k \right) * \frac{z}{(1-z)^{n+1}} = z + \sum_{k=2}^{\infty} \binom{k+n-1}{n} a_k z^k.$$

Thus the extreme points of $clcoM_n(\alpha)$ are given by

$$z+2(1-\alpha)\sum_{k=2}^{\infty}\frac{1}{k}\binom{k+n-1}{n}^{-1}x^{k-1}z^k\ ,\quad |x|=1,\ z\in U,$$

which simplifies to (2.1). Since the family $M_n(\alpha)$ is convex and therefore is equal to its convex hull, (2.1) gives the extreme points of $M_n(\alpha)$.

For the cases $n=0$ and $n=1$, Theorem 1 gives the extreme points of $M_o(\alpha)$ and $M_1(\alpha)$ obtained in [6] and [10], respectively.
Since the coefficient bounds are maximized at an extreme point, as an application of Theorem 1, we have

COROLLARY. If $f\in M_n(\alpha)$, then

$$|a_k|\le\frac{n!(k-1)!2(1-\alpha)}{(k+n-1)!k}\ ,\quad k\ge 2.$$

Equality occurs for $f_x(z)$ defined by (2.1).
From (2.1) we see for $f\in M_n(\alpha)$ and $|z|=r<1$ that

$$|f(z)|\le r+2(1-\alpha)\sum_{k=2}^{\infty}\frac{n!(k-1)!}{(k+n-1)!k}r^k.$$

By letting $r\to 1$ we obtain

$$|f(z)|\le 1+2(1-\alpha)\sum_{k=2}^{\infty}\frac{n!(k-1)!}{(k+n-1)!k}. \tag{2.2}$$

Since for $n\ge 1$, $M_n(\alpha)\subset M_1(\alpha)$, we let $n=1$ in (2.2) to get

$$|f(z)|\le 1+2(1-\alpha)(\frac{\pi^2}{6}-1).$$

This shows that the family $M_n(\alpha)$, $n\ge 1$ is bounded in U for all real α, $\alpha<1$, even tough its functions may not be univalent. For $n=0$, (2.2) becomes $|f(z)|\le 1+2(1-\alpha)(\Sigma_{k=2}^{\infty}1/k)$ which diverges for all α, $\alpha<1$. So the functions in $M_o(\alpha)$ are not necessarily bounded in U. The above corollary for $n=0$ yields Corollary 4.1. by Hallenbeck [6] (page 289), and a special case of Theorem 3 by Ahuja [1] (page 178). When $n=1$ we get the coefficient bounds obtained by Silverman [10]. Our next theorem is on the partial sums of the functions in $M_n(\alpha)$ which for the case $n=1$ gives Theorem 4(i) by Siverman [10].

THEOREM 2. Let $S_m(z,f)$ denote the mth partial sum of a function f in $M_n(\alpha)$. If $f\in M_n(\alpha)$ and if $1\le n\le 4$, then

$$S_m(z,f)\in M_{n-1}\left(\frac{2\alpha n+1-n}{n+1}\right).$$

To prove this theorem we shall need the following lemmas, the first of which is due to Gasper [5].

LEMMA 1. Let R be the positive root of the equation

$$9t^7 + 55t^6 - 14t^5 - 948t^4 - 3247t^3 - 5013t^2 - 3780t - 1134 = 0.$$

If $-1 < t \leq R \approx 4.5678018$, then

$$\sum_{k=1}^{m} \frac{cosk\theta}{k+t} \geq -\frac{1}{1+t}, \quad m = 1, 2, \ldots .$$

When $t = 1$, Lemma 1 confirms the estimate by Rogosinski and Szegö [8].

LEMMA 2. Let $-1 < t \leq R \approx 4.5678018$. Then

$$Re\left(\sum_{k=2}^{m} \frac{z^{k-1}}{k+t-1}\right) > -\frac{1}{1+t}, \quad z \in U.$$

PROOF. Let $z = re^{i\theta}$, $0 \leq r < 1$, and $0 \leq |\theta| \leq \pi$. Using the minimum principle for harmonic functions and the above Lemma 1, we obtain

$$Re\left(\sum_{k=2}^{m} \frac{z^{k-1}}{k+t-1}\right) > \sum_{k=1}^{m-1} \frac{\cos k\theta}{k+t} \geq -\frac{1}{1+t}.$$

LEMMA 3. Let $p(z)$ be analytic in U, $p(0) = 1$, and $Re\ p(z) > 1/2$ in U. Then for any function F, analytic in U, the function $p * F$ takes values in the convex hull of the image of U under F.

The above Lemma 3 is a well-known and celebrated result that can be derived from Herglotz' representation for $p(z)$ in U.

PROOF OF THEOREM 2. Let $f \in M_n(\alpha)$ be of the form (1.1). Then we have

$$Re\left(1 + \sum_{k=2}^{\infty} k\binom{k+n-1}{n} a_k z^{k-1}\right) > \alpha \tag{2.3}$$

or

$$Re\left(1 + \sum_{k=2}^{\infty} \frac{2kn}{n+1}\binom{k+n-1}{n} a_k z^{k-1}\right) > \frac{2\alpha n + 1 - n}{n+1}.$$

For the mth partial sum of f we can write

$$\begin{aligned}(D^{n-1}S_m(z,f))' &= 1 + \sum_{k=2}^{m} k\binom{k+n-2}{n-1} a_k z^{k-1} \\ &= \left(1 + \sum_{k=2}^{\infty} \frac{2kn}{n+1}\binom{k+n-1}{n} a_k z^{k-1}\right) * \left(1 + \sum_{k=2}^{m} \frac{n+1}{2(k+n-1)} z^{k-1}\right).\end{aligned}$$

For $1 \le n \le 4$, it is clear by Lemma 2 that

$$Re\left(1+\sum_{k=2}^{m}\frac{n+1}{2(k+n-1)}z^{k-1}\right)>\frac{1}{2}\,,\quad z\in U.$$

Now an application of Lemma 3 to $(D^{n-1}S_m(z,f))'$ concludes the theorem.

REMARK. We think that Theorem 2 also holds for $n \ge 5$, but we are unable to verify this. One way to check this is to see if Lemma 2 can be extended to larger values of t, $t > R$.

In our next theorem we investigate the operator $\Lambda(f) = Re\{z^{-1}(z^n f'(z))^{(n-1)}\}$ acting on $M_n(\alpha)$.

THEOREM 3. If $n \ge 1$ and if $f \in M_n(\alpha)$, then

$$Re\frac{(z^n f'(z))^{(n-1)}}{z} > n!\,(2\alpha - 1 + 2(1-\alpha)log\ 2).$$

The result is sharp.

PROOF. For $n \ge 1$ let $f \in M_n(\alpha)$ be of the form (1.1). Then we have

$$Re\left(1+\frac{1}{2(1-\alpha)}\sum_{k=2}^{\infty}k\binom{k+n-1}{n}a_k z^{k-1}\right)>\frac{1}{2}\,,\quad z\in U.$$

Write

$$\frac{(z^n f'(z))^{(n-1)}}{n!z}=\left(1+\frac{1}{2(1-\alpha)}\sum_{k=2}^{\infty}k\binom{k+n-1}{n}a_k z^{k-1}\right)*\left(1+2(1-\alpha)\sum_{k=2}^{\infty}\frac{1}{k}z^{k-1}\right).$$

By a result of Hallenbeck [6] (Remark 3, page 291), it is obvious that

$$Re\left(1+2(1-\alpha)\sum_{k=2}^{\infty}\frac{1}{k}z^{k-1}\right)>(2\alpha-1)+2(1-\alpha)log\ 2\,,\quad z\in U.$$

Finally an application of Lemma 3 completes the proof. For sharpness, let f be the function $f_x(z)$ defined by (2.1).
The case $M_1(0)$ can be found in [12], and the case $M_1(\alpha)$ is obtained by Ali [3] and Silverman [10].

The next lemma, due to Fejĕr [4], will be used to prove our theorem on the convolution of functions in $M_n(\alpha)$. A sequence $\{c_k\}_{k=0}^{\infty}$ of non-negative real numbers is said to be a convex null sequence if $c_k \to 0$ as $k \to \infty$, and $c_o - c_1 \ge c_1 - c_2 \ge \ldots \ge c_{k-1} - c_k \ge \ldots \ge 0$.

LEMMA 4. Let $\{c_k\}_{k=0}^{\infty}$ be a convex null sequence. Then the function

$$p(z)=\frac{c_o}{2}+\sum_{k=1}^{\infty}c_k z^k\,,\quad z\in U,$$

is analytic and $Rep(z) > 0$ in U.

THEOREM 4. Let f and g be in $M_n(\alpha)$. Then $f * g \in M_n(\beta)$ if

$$\beta = \frac{n(2\alpha+1)+4\alpha-1}{2(n+1)} \geq \alpha. \tag{2.4}$$

PROOF. For $c_o = 1$ and

$$c_k = \frac{n+1}{(k+1)\binom{k+n}{n}}, \quad k \geq 1,$$

we see that $\{c_k\}_{k=0}^{\infty}$ is a convex null sequence. Therefore, by Lemma 4, we have

$$Re\left(1+\sum_{k=2}^{\infty}\frac{n+1}{k\binom{k+n-1}{n}}z^{k-1}\right) > \frac{1}{2}. \tag{2.5}$$

Let $g(z) = z + \Sigma_{k=2}^{\infty} b_k z^k$ be in $M_n(\alpha)$. By (2.3) we have

$$Re\left(1+\sum_{k=2}^{\infty}k\binom{k+n-1}{n}b_k z^{k-1}\right) > \alpha. \tag{2.6}$$

Now we convolve (2.5) and (2.6) and apply Lemma 3 to obtain

$$Re\left(1+(n+1)\sum_{k=2}^{\infty}b_k z^{k-1}\right) > \alpha$$

or

$$Re\frac{g(z)}{z} = Re\left(1+\sum_{k=2}^{\infty}b_k z^{k-1}\right) > \frac{n+\alpha}{n+1}$$

or

$$Re\left(\frac{g(z)}{z} - \frac{2\alpha+n-1}{2(n+1)}\right) > \frac{1}{2}.$$

Since $Re(D^n f)' > \alpha$, we once again use Lemma 3 to obtain

$$Re\left((D^n f)' * \left(\frac{g(z)}{z} - \frac{2\alpha+n-1}{2(n+1)}\right)\right) > \alpha$$

or

$$Re\left((D^n f)' * \frac{g(z)}{z}\right) > \frac{n(2\alpha+1)+4\alpha-1}{2(n+1)} = \beta.$$

Using the fact that $(D^n(f*g))' = (D^n f)' * (g(z)/z)$ concludes the theorem.

For the case $n = 0$ and $0 \leq \alpha < 1$, Theorem 4 gives the corresponding result in Theorem 2 by Ahuja [1], and for the case $n = 1$ and $\alpha = 0$ it is Theorem 3 by Singh and Singh [12].

It is clear from the proof of the above Theorem 4 that the following more general result holds.

THEOREM 5. Let $f \in M_n(\alpha)$ and let $g \in \mathcal{A}$ so that

$$Re\frac{g(z)}{z} > \frac{n+\alpha}{n+1}, \quad z \in U.$$

Then $f * g \in M_n(\beta)$ where β is given by (2.4).

REFERENCES

1. AHUJA, O.P., Univalent functions whose derivatives have a positive real part, Rend. di Mat. Serie VII, 1 (2) (1982), 173-187.
2. ALEXANDER, J.W., Functions which map the interior of the unit circle upon simple regions, Ann. Math. 17 (1915), 12-22.
3. ALI, R.M., On a subclass of starlike functions, Rocky Mount. J. Math. 24 (2) (1994), 447-451.
4. FEJĚR, L., Uber die positivitat von summen, die nach trigonometrischen order Legendreschen funktionen fortscheiten, Acta Litt. Ac. Sc. Szeged (1925), 75-86.
5. GASPER, G., Nonnegative sums of cosine, ultraspherical and Jacobi polynomials, J. Math. Anal. Appl. 26 (1969), 60-68.
6. HALLENBECK, D.J., Convex hulls and extreme points of some families of univalent functions, Trans. Amer. Math. Soc. 192 (1974), 288-292.
7. NOOR, K.I., On some applications of the Ruscheweyh derivatives, Math. Japon. 36 (1991), 869-874.
8. ROGOSINSKI, W. & SZEGŐ, Uber die Abschimlte Vol potenzreihen die in ernein Kreise be schrarikt bleiben, Math. Z. 28 (1928), 73-94.
9. RUSCHEWEYH, S., New criteria for univalent functions, Proc. Amer. Math. Soc. 49 (1975), 109-115.
10. SILVERMAN, H., A class of bounded starlike functions, Internat. J. Math. & Math. Sci. 17 (2) (1994), 249-252.
11. SINGH, R. & SINGH, S., Starlikeness and convexity of certain integrals, Ann. Univ. Mariae Curie-Sklodowska Sect A 35 (1981), 45-47.
12. SINGH, R. & SINGH, S., Convolution properties of a class of starlike functions, Proc. Amer. Math. Soc. 106 (1989), 145-152.

NUMERICAL-SYMBOLIC HYBRID METHOD FOR DEGENERATE PARABOLIC EQUATION RELATED TO POPULATION GENETICS

Kazuo Amano
Department of Mathematics, Josai University
kamano@tansei.cc.u-tokyo.ac.jp

1. Introduction

As is well-known, if given partial differential equation is very simple, we can compute its general solution with arbitrary functions by using arithmetics and elementary calculus. However, most of equations require hard and abstract mathematical technicalities, and explicit and concrete representation of solution turns out to be utterly beyond our reach. It seems that researchers have already given up to construct explicit general solutions; they are either trying to find solutions in abstract function spaces or working out numerical algorithms.

In this paper, we shall show new possibilities of *approximate general solution*; though explicit representation is already at deadlock, approximate one is able to break through obstacles and gives a new viewpoint. In fact, we prove that, for a certain initial value problem, there exists a simple algebraic representation of approximate general solution, *i.e.*, a symbolic combination of additions, subtractions and multiplications of initial data solves the problem. Our procedure of construction of general solution is quite different from classical ones; we use a new type of *numerical-symbolic hybrid method*. Our numerical-symbolic hybrid computation totally depends on LISP and its result is expressed in C language, since the size of desired formula is more than 5.4M bytes. Such a formula is too big for classical pen and paper calculation. It is to be noted that a remarkably fast algorithm is derived from our formula of approximate general solution.

The purpose of this paper is to construct an approximate general solution of initial value problem for degenerate parabolic equation

$$(1.1) \qquad \frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial x^2}(V(x)u) - \frac{\partial}{\partial x}(M(x)u) \qquad (t > 0,\ 0 < x < 1)\ ,$$

which appears in population genetics theory, where $V(x) = \dfrac{x(1-x)}{2N}$, $M(x)$ is a polynomial of x, and $N \in \{1, 2, 3, \cdots\}$ denotes the given population number. For a certain class of coefficient $M(x)$, Crow and Kimura obtained several explicit representations of solutions of (1.1) by using special functions (*cf.* [2]). However, their technicality is too artistic and too delicate to generalize it; in fact, if you perturb the function $M(x)$ slightly, their method simply breaks down. In this paper, we shall show that we can construct approximate general solution of (1.1) for any given $M(x)$.

Section 2 is devoted to preliminary lemmas. In Section 3, we construct an approximate general solution of (1.1) and translate its procedure into symbolic list operations, which is expressed in PASCAL-like REDUCE language ([4]).

2. Preliminaries

We shall simplify the given partial differential equation (1.1) by performing a certain transformation (Lemma 2.2) and give an approximate representation formula (Lemma 2.3).

Lemma 2.1. *For any nonnegative integer n and for any C^{n+1} smooth function $u(t,x)$, we have*

$$\begin{aligned} u(t+h,x+k) &= \sum_{\nu=0}^{n} \frac{1}{\nu!}\Big(h\frac{\partial}{\partial t}+k\frac{\partial}{\partial x}\Big)^{\nu} u(t,x) \\ &+\int_0^1 \frac{(1-\theta)^n}{n!}\Big(h\frac{\partial}{\partial t}+k\frac{\partial}{\partial x}\Big)^{n+1} u(t+\theta h,x+\theta k)\,d\theta \end{aligned} \tag{2.1}$$

Let us define a differential operator A by

$$Au = \frac{1}{2}\frac{\partial^2}{\partial x^2}(V(x)u) - \frac{\partial}{\partial x}(M(x)u)\ , \tag{2.2}$$

where $V(x) = \dfrac{x(1-x)}{2N}$, $M(x)$ is a polynomial of x with real coefficients, and $N \equiv N_e$ denotes the effective population number $\in \{1,2,3,\cdots\}$. As is well-known, (2.2) plays an essential role in population genetics theory (*cf.*, for example, [2]). In fact, if we assume that a pair of alleles α_1 and α_2 are segregating in a Mendelian population, then the probability density $u(t,x)$ of frequency of α_1 at t th generation satisfies the partial differential equation

$$\frac{\partial u}{\partial t} = Au \qquad \text{in} \quad (0,NT)\times(0,1)\ . \tag{2.3}$$

Here T is a positive constant.

Lemma 2.2. *We put*

$$a(x) = \frac{x(1-x)}{2}, \quad b(x) = \frac{1-2x}{2} - NM(x), \quad c(x) = -NM'(x), \tag{2.4}$$

and define a differential operator B by

$$B = \frac{1}{2}a(x)\frac{\partial^2}{\partial x^2} + b(x)\frac{\partial}{\partial x} + c(x)\ . \tag{2.5}$$

If $v(t,x)$ is a solution of the equation

$$\frac{\partial v}{\partial t} = Bv \qquad \text{in} \quad (0,T)\times(0,1) \tag{2.6}$$

then

$$u(t,x)=\exp(-\frac{t}{2N})\,v(\frac{t}{N},x) \tag{2.7}$$

satisfies (2.3).

By virtue of Lemma 2.2, the partial differential equation (2.3) is reduced to (2.6). ¿From view points of numerical analysis and population genetics, it is much easier to solve (2.6) than (2.3). In fact, the rate of fixation term $\exp(-t/2N)$ is separated in (2.7) and the original time scale is changed to a moderate one in the process of reduction.

As will be shown later, we can construct the desired *approximate general solution* of the equation $u_t = Bu$ by using the following lemma recursively.

Lemma 2.3. *Assume that* $u(t,x)\in C^{2,4}([0,\infty)\times[0,1])$ *is a classical solution of degenerate parabolic equation*

$$\frac{\partial u}{\partial t}=Bu \quad (t>0,\ 0<x<1). \tag{2.8}$$

Then we obtain

$$u(t,x)=Lu(t,x)-h^4Ru(t,x) \tag{2.9}$$

when $0<h\ll 1$, $0\le x\pm\sqrt{a(x)}h\le 1$, $0\le x+b(x)h^2\le 1$, *and* $t-h^2\ge 0$. *Here*

$$\begin{aligned} &Lu(t,x)\\ &=\frac{1}{6}u(t,x+\sqrt{a(x)}\,h)+\frac{1}{6}u(t,x-\sqrt{a(x)}\,h)\\ &+\frac{1}{3}u(t,x+b(x)h^2)+\frac{1}{3}u(t-h^2,x)+\frac{h^2}{3}c(x)u(t,x) \end{aligned} \tag{2.10}$$

and

$$\begin{aligned} &Ru(t,x)\\ &=\frac{1}{3}\int_0^1\Big\{\frac{a^2(x)(1-\theta)^3}{12}\Big(\frac{\partial^4 u}{\partial x^4}(t,x+\sqrt{a(x)}\,\theta h)+\frac{\partial^4 u}{\partial x^4}(t,x-\sqrt{a(x)}\,\theta h)\Big)\\ &+b^2(x)(1-\theta)\frac{\partial^2 u}{\partial x^2}(t,x+b(x)\theta h^2)+(1-\theta)\frac{\partial^2 u}{\partial t^2}(t-\theta h^2,x)\Big\}\,d\theta\,. \end{aligned} \tag{2.11}$$

Lemma 2.3 shows that if $u(t,x)$ is a $C^{2,4}$ smooth solution, then $u(t,x)$ is able to be rewritten as

$$\begin{aligned} u(t,x)&=Lu(t,x)+O(h^4)\\ &=\frac{1}{6}u(t,x+\sqrt{a(x)}\,h)+\frac{1}{6}u(t,x-\sqrt{a(x)}\,h)\\ &+\frac{1}{3}u(t,x+b(x)h^2)+\frac{1}{3}u(t-h^2,x)+\frac{h^2}{3}c(x)u(t,x)+O(h^4)\,; \end{aligned} \tag{2.12}$$

recursive use of this formula gives the desired *approximate general solution* of degenerate parabolic equation (2.8). Here, by (2.8) and (2.10), $O(h^4)$ depends on u_{xxxx}, u_{xx} and B^2u and further, a standard method of maximum principle ensures that we can estimate $\dfrac{\partial^k u}{\partial x^k}$ $(0\le k\le 4)$ explicitly (*cf.*, for example, [5]).

3. Construction of Approximate General Solution

In this section, we shall construct approximate general solution of (2.8) and gives its error bound (Theorems 3.1 and 3.2). We also show that our result is able to be translated into symbolic list operations quite easily (Reduce Program).

We note that seeing the phenomenon which (1.1) describes, any boundary condition should not be given on $x = 0,\ 1$, *i.e.*, the boundary $x = 0,\ 1$ should be either *entrance* or *natural boundary* of diffusion process generated by (2.6). So, without loss of practical generality, we may assume that

$$b(x) \begin{cases} \geq 0 & \textit{in a neighborhood of } 0 \textit{ in } [0,1] , \\ \leq 0 & \textit{in a neighborhood of } 1 \textit{ in } [0,1] . \end{cases} \tag{3.1}$$

Furthermore, we may consider

$$c(x) \leq 0 \quad \textit{in } [0,1] \tag{3.2}$$

since if we put

$$w(t,x) = e^{-tc_0}\, u(t,x) \ , \quad c_0 = \sup_{0\leq x\leq 1} c(x)$$

then

$$u_t = Bu \iff w_t = (B - c_0)w \ .$$

Throughout this section, $u(t,x) \in C^{2,4}([0,\infty)\times[0,1])$ is a classical solution of the equation (2.8) and h denotes a sufficiently small fixed positive constant, *i.e.*, $0 < h \ll 1$.

By using (2.12), if we relace $(h^2/3)c(x)u(t,x)$ with

$$\frac{h^2}{3}c(x)\Big(\frac{1}{6}\,u(t,x+\sqrt{a(x)}\,h) + \frac{1}{6}\,u(t,x-\sqrt{a(x)}\,h)$$
$$+ \frac{1}{3}\,u(t,x+b(x)h^2) + \frac{1}{3}\,u(t-h^2,x) + \frac{h^2}{3}\,c(x)u(t,x) + O(h^4)\Big)$$

in (2.12), we obtain

$$\begin{aligned} u(t,x) = {} & \Big(1 + \frac{h^2}{3}c(x)\Big)\Big(\frac{1}{6}\,u(t,x+\sqrt{a(x)}\,h) + \frac{1}{6}\,u(t,x-\sqrt{a(x)}\,h) \\ & + \frac{1}{3}\,u(t,x+b(x)h^2) + \frac{1}{3}\,u(t-h^2,x)\Big) + \frac{h^4}{9}\,c^2(x)u(t,x) + O(h^4) \ . \end{aligned} \tag{3.3}$$

Here, since $c^2(x)|u(t,x)| = \big(NM'(x)\big)^2|u(t,x)|$ is not so small in some cases, it is better not to replace $(h^4/9)\,c^2(x)u(t,x)$ with $O(h^4)$. So, we repeat the above recursive transformation again for (3.3) and get the following

$$\begin{aligned} u(t,x) = {} & \Big(1 + \frac{h^2}{3}c(x) + \frac{h^4}{9}c^2(x)\Big)\Big(\frac{1}{6}\,u(t,x+\sqrt{a(x)}\,h) + \frac{1}{6}\,u(t,x-\sqrt{a(x)}\,h) \\ & + \frac{1}{3}\,u(t,x+b(x)h^2) + \frac{1}{3}\,u(t-h^2,x)\Big) + O(h^4) \ ; \end{aligned} \tag{3.4}$$

this is the key formula of this section.

We define a function $\varepsilon(x)$ by

$$
(3.5) \qquad \varepsilon(x) = \begin{cases} h & \text{if } h^2 \le x \le 1-h^2 \text{ or } x = 0,\ 1 \\ \dfrac{x}{\sqrt{a(x)}} & \text{if } 0 < x < h^2 \\ \dfrac{1-x}{\sqrt{a(x)}} & \text{if } 1-h^2 < x < 1\,. \end{cases}
$$

Here it is to be noted that, by (2.4), (3.1) and (3.5),

$$
(3.6) \qquad x \pm \sqrt{a(x)}\,\varepsilon(x),\ x + b(x)\,\varepsilon^2(x) \in [0,1]
$$

holds for any $0 \le x \le 1$.

Lemma 3.1. *For* $0 < h \ll 1$,

$$
(3.7) \qquad h^2 \le x \le 1-h^2 \quad \textit{implies} \quad \frac{h^2}{4} \le x \pm \sqrt{a(x)}\,h \le 1 - \frac{h^2}{4}\,.
$$

Lemma 3.2. *For* $0 < h \ll 1$,

$$
(3.8) \qquad \frac{h^2}{4} \le y \le 1 - \frac{h^2}{4} \quad \textit{implies} \quad \varepsilon(y) \ge \frac{h}{\sqrt{2}}
$$

By L_ε, we denote a difference operator

$$
(3.9) \qquad L_\varepsilon f(t,x) = \begin{cases} \Big(1 + \dfrac{\varepsilon^2(x)}{3}c(x) + \dfrac{\varepsilon^4(x)}{9}c^2(x)\Big) & \\ \times\Big(\dfrac{1}{6}f(t,x+\sqrt{a(x)}\,\varepsilon(x)) + \dfrac{1}{6}f(t,x-\sqrt{a(x)}\,\varepsilon(x)) & \\ \quad + \dfrac{1}{3}f(t,x+b(x)\varepsilon^2(x)) + \dfrac{1}{3}f(t-\varepsilon^2(x),x)\Big) & \text{if } t \ge h^2 \\ f(t,x) & \text{otherwise}\,. \end{cases}
$$

We define functions $p_k(t,x;s,y)$, $k = 0,1,2,\cdots$ as follows :

$$
(3.10) \qquad p_0(t,x;s,y) = \begin{cases} 1 & \text{if } (s,y) = (t,x) \\ 0 & \text{otherwise}\,, \end{cases}
$$

$$
(3.11) \quad p_1(t,x;s,y) = \begin{cases} \dfrac{1}{6}\Big(1 + \dfrac{\varepsilon^2(x)}{3}c(x) + \dfrac{\varepsilon^4(x)}{9}c^2(x)\Big) & \text{if } (s,y) = (t, x \pm \sqrt{a(x)}\,\varepsilon(x)) \\ \dfrac{1}{3}\Big(1 + \dfrac{\varepsilon^2(x)}{3}c(x) + \dfrac{\varepsilon^4(x)}{9}c^2(x)\Big) & \text{if } (s,y) = (t, x + b(x)\,\varepsilon^2(x)) \\ \dfrac{1}{3}\Big(1 + \dfrac{\varepsilon^2(x)}{3}c(x) + \dfrac{\varepsilon^4(x)}{9}c^2(x)\Big) & \text{if } (s,y) = (t - \varepsilon^2(x), x) \\ 0 & \text{otherwise} \end{cases}
$$

and $p_{k+1}(t,x;s,y)$ $(k \geq 1)$ is a function such that

$$\int_D f(s,y)\, p_{k+1}(t,x;ds,dy) = \int_D L_\varepsilon f(s,y)\, p_k(t,x;ds,dy) \tag{3.12}$$

$$\left(i.e., \quad \sum_{(s,y)\in D} f(s,y)\, p_{k+1}(t,x;s,y) = \sum_{(s,y)\in D} L_\varepsilon f(s,y)\, p_k(t,x;s,y) \right)$$

for any function $f(t,x)$ defined in $D \equiv [0,\infty] \times [0,1]$. Here (3.12) is well-defined, since $\sqrt{a(x)}\,\varepsilon(x) \neq |b(x)|\,\varepsilon^2(x)$ for $0 < h \ll 1$. It is clear that, in (3.12), $p_{k+1}(t,x;s,y)$ is uniquely determined by $p_k(t,x;s,y)$.

(3.10) implies

$$u(t,x) = \int_D p_0(t,x;ds,dy)\, u(s,y)\ . \tag{3.13}$$

Since $\varepsilon^4(x) = O(h^4)$, (3.4) and (3.11) give

$$u(t,x) = \int_D p_1(t,x;ds,dy)\, u(s,y) + O(h^4)\ . \tag{3.14}$$

Furthermore, by (3.4), (3.12) and (3.14), we have

$$\begin{aligned} u(t,x) &= \int_D p_1(t,x;ds,dy)\, u(s,y) + O(h^4) \\ &= \int_D p_1(t,x;ds,dy)(L_\varepsilon u(s,y) + O(h^4)) + O(h^4) \\ &= \int_D p_2(t,x;ds,dy)\, u(s,y) + O(h^4)r + O(h^4) \\ &= \int_D p_2(t,x;ds,dy)(L_\varepsilon u(s,y) + O(h^4)) + O(h^4)r + O(h^4) \\ &= \int_D p_3(t,x;ds,dy)\, u(s,y) + O(h^4)r^2 + O(h^4)r + O(h^4)\ , \end{aligned} \tag{3.15}$$

where

$$r = \sup_{0\leq x\leq 1} \left(1 + \frac{\varepsilon^2(x)}{3}c(x) + \frac{\varepsilon^4(x)}{9}c^2(x)\right) = 1 + O(h^4) \tag{3.16}$$

by virtue of (3.2) and (3.5). We can repeat this procedure inductively.

Combining the above results, we obtain the following

Theorem 3.1. *Assume that $u(t,x) \in C^{2,4}([0,\infty) \times [0,1])$ is a classical solution of degenerate parabolic equation (2.8). Then we have*

$$u(t,x) = \int_{0\leq s\leq t,\ 0\leq y\leq 1} p_k(t,x;ds,dy)\, u(s,y) + O(h^4)\sum_{\nu=0}^{k-1} r^\nu \tag{3.17}$$

for any $k = 0,1,2,\cdots$. Here $O(h^4)$ is independent of k.

In (3.17), if we replace $u(s,y)$, $0 \leq s < h^2$ with given initial data $\phi(y)$, we get the following

Theorem 3.2. *Assume that* $u(t,x) \in C^{2,4}([0,\infty) \times [0,1])$ *is a classical solution of degenerate parabolic equation (2.8) satisfying initial condition*

$$u(0,x) = \phi(x) \qquad (0 < x < 1) .$$

Then we have

$$\begin{aligned} u(t,x) = \int_{0 \le s < h^2,\ 0 \le y \le 1} & p_k(t,x;ds,dy)\,\phi(y) + O(h^2)\,r^k \\ (3.18) \qquad & + O(1)\,r^k \sum_{\ell=0}^{k_0} \binom{k}{\ell} \left(\frac{1}{3}\right)^{\ell} \left(\frac{2}{3}\right)^{k-\ell} + O(h^4) \sum_{\nu=0}^{k-1} r^{\nu} \end{aligned}$$

for any $k \gg 1$, *where* k_0 *is the smallest integer* $\ge 2t/h^2$. *Here* $O(1)$, $O(h^2)$ *and* $O(h^4)$ *are independent of* k .

Hence, we obtain the desired *approximate general solution*

$$u(t,x) \sim \int_{0 \le s < h^2,\ 0 \le y \le 1} p_k(t,x;ds,dy)\,\phi(y) \tag{3.19}$$

for $1 \ll k \ll 1/h^4$. For a given $0 < h \ll 1$, we have only to choose a certain $k \gg 1$ which minimizes

$$h^2 r^k + r^k \sum_{\ell=0}^{k_0} \binom{k}{\ell} \left(\frac{1}{3}\right)^{\ell} \left(\frac{2}{3}\right)^{k-\ell} + h^4 \sum_{\nu=0}^{k-1} r^{\nu} .$$

If we identify a linear combination of the form

$$\int_{0 \le s \le t,\ 0 \le y \le 1} p_k(t,x;ds,dy)\,u(s,y) = q_1 u(t_1,x_1) + q_2 u(t_2,x_2) + \cdots + q_n u(t_n,x_n)$$

with a list

$$((q_1\ t_1\ x_1)\ (q_2\ t_2\ x_2)\ \cdots\ (q_n\ t_n\ x_n))$$

of LISP type, then the essential part of Theorem 3.2 is translated into the following list operations.

Reduce Program.

```
procedure hybrid_method(t, x, n);
begin;
   list_in := {{1, t, x}};
   list_tmp := {};
   while n > 0 do
   <<
      while length(list_in) > 0 do
      <<
```

```
        tmp := first(list_in);
        p := first(tmp);
        s := first(rest(tmp));
        y := first(rest(rest(tmp)));
        q := (1+h**2*c(y)/3+h**4*c(y)**2/9)*p;
        if domain_p(s) neq 0 then
        <<
           list_tmp := cons({q/6, s, y+sqrt(a(y))*h}, list_tmp);
           list_tmp := cons({q/6, s, y-sqrt(a(y))*h}, list_tmp);
           list_tmp := cons({q/3, s, y+b(y)*h**2}, list_tmp);
           list_tmp := cons({q/3, s-h**2, y}, list_tmp)
        >>
        else
           list_tmp := cons({p, s, y}, list_tmp);
        list_in := rest(list_in)
     >>;
     list_in := list_tmp;
     list_tmp := {};
     n := n-1
  >>;
  return list_in
end;
```

Here `domain_p(t)` is a function defined in $[0, \infty)$ such that

$$\texttt{domain\_p(t)} = \begin{cases} 1 & \text{if } t \geq h^2 \\ 0 & \text{otherwise .} \end{cases}$$

Functions $a(x)$, $b(x)$ and $c(x)$ are coefficients defined by (2.4). To be more explicit, the parameter h should be modified in each step of main loop of the above Reduce Program so that

$$y \pm \sqrt{a(y)}\, h,\ y + b(y)\, h^2 \in [0, 1]$$

is valid and also, we should identify any pair of list $(q_i\ t_i\ x_i)$ and $(q_j\ t_j\ x_j)$ with $(q_i + q_j\ t_i\ x_i)$ when $(t_i\ x_i) = (t_j\ x_j)$.

References

[1] K. Amano, Numerical-symbolic hybrid method for biharmonic Dirichlet problem, (*to appear*).

[2] J. F. Crow and M. Kimura, An Introduction to Population Genetics Theory, *Burgess Publishing Company*, 1970.
[3] W. J. Ewens, Mathematical Population Genetics, *Springer-Verlag*, 1989.
[4] A. C. Hearn, REDUCE User's Manual, version 3.5, *RAND Publication*, CP78(Rev. 7/94), 1994.
[5] O. A. Oleinik and E. V. Radkevich, Second Order Equations with Nonnegative Characteristic Form, *Amer. Math. Soc., Province, Rhode Island and Plenum Press, New York*, 1973.

DEPARTMENT OF MATHEMATICS, JOSAI UNIVERSITY, SAKADO, SAITAMA, JAPAN
E-mail: kamano@tansei.cc.u-tokyo.ac.jp

A NOTE ON NONLINEAR FREE-SURFACE FLOW UNDER A SLUICE GATE. PRELIMINARY REPORT

J. Asavanant[1] and J.-M. Vanden-Broeck[2]

ABSTRACT

The problem of the flow under a sluice gate has been studied by many investigators in the last five decades (see Benjamin[1], Fangmeier and Strelkoff[2], Larock[3], Chung[4], Cheng, Liggett and Liu[5] and others). Solutions were obtained for large values of the downstream Froude number or by approximating the upstream free surface by a rigid lid. It is the purpose of this report to investigate numerically the complete nonlinear problem. The problem is solved by using a series truncation technique. The flow domain is mapped conformally onto the unit semicircle. A complex velocity function is constructed in the form of a series expansion. The unknown coefficients in the expansion are determined by a collocation procedure. Accurate solutions are presented for the flow under a inclined sluice gate. It is assumed that the flow leaves tangentially at both ends of the gate.

Correspondent: Jack Asavanant
Correspondent's telephone: (62) 218-5163
Correspondent's FAX: (62) 255-2287
Correspondent's e-mail: asavanan@cms.wisc.edu

[1] Department of Mathematics, Chulalongkorn University, Bangkok 10330, Thailand.
[2] Department of Mathematics and Center for the Mathematical Sciences, University of Wisconsin-Madison, Madison, Wi 53705, USA.

A Note on Nonlinear Free-Surface Flow under a Sluice Gate. Preliminary Report.

Let us consider the steady irrotational free-surface flow of an incompressible and inviscid fluid under an inclined sluice gate (see Fig.1). The gate inclination is denoted by γ. The $x-$ axis is along the bottom AD and the $y-$ axis is directed vertically upwards through the midpoint of the gate. As $|x| \to \infty$, the flows approach uniform streams with velocity $\tilde{U}$ and depth $\tilde{H}$ upstream and velocity U and depth H downstream. The flow is subcritical upstream and supercritical downstream. The downstream Froude number is defined as

$$F = \frac{U}{\sqrt{gH}}. \tag{1}$$

Here g is the acceleration of gravity. The flow is assumed to leave tangentially at both ends of the gate.

We introduce dimensionless variables by taking $\left(Q^2/g\right)^{\frac{1}{3}}$ as the unit length and $(Qg)^{\frac{1}{3}}$ as the unit velocity. Here Q denotes the discharge of the flow. We define the potential function ϕ, the streamfunction ψ, the complex potential function $f = \phi + i\psi$, and the complex velocity function $\zeta = df/dz$. Here $z = x + iy$. Without loss of generality, we choose $\phi = 0$ so that $\phi_b = -\phi_c$. Here ϕ_b and ϕ_c are the values of ϕ at the separation points B and C respectively. The bottom AD is a streamline on which $\psi = 0$. The flow region in the $f-$ plane is the strip $0 \le \psi \le 1$.

In terms of the dimensionless variables, the constant pressure condition on the free surface can be written

$$|\zeta|^2 + 2y = 3 \quad \text{on AB and CD.} \tag{2}$$

We map the flow region in the $f-$ plane onto the upper half of the unit circle in the complex $t-$ plane by the transformation

$$f = \frac{2}{\pi}\log\left(\frac{1+t}{1-t}\right). \tag{3}$$

The free surfaces AB, CD and the gate BC are mapped onto the circumference of the half unit circle, whereas, the bottom AD of the channel goes onto the real diameter. We use the notation $t = re^{i\sigma}$ so that the free surfaces and the gate are described by $r = 1$ and $0 \le \sigma \le \pi$. The images of points B and C are $t_2 = -e^{-i\beta}$ and $t_1 = e^{i\beta}$ respectively. Here

$0 \leq \beta \leq \pi/2$. The mathematical problem becomes that of finding ζ as an analytic function of t with appropriate behaviors at $t = 1, t_1$, and t_2

As $\phi \to \infty$, the flow approaches a uniform supercritical stream. Since supercritical flows are characterized by the presence of exponentially decaying terms, the local behavior of ζ as $\phi \to \infty$ is

$$\zeta \sim E + K e^{-\pi\lambda f} \qquad \text{as } \phi \to \infty. \tag{4}$$

Here E and K are constants to be found as part of the solution and λ is the smallest positive root of

$$F^2 \pi\lambda - \tan(\pi\lambda) = 0. \tag{5}$$

It follows from our choice of dimensionless variables that the Froude number F is related to the velocity downstream by

$$F = |\zeta(1)|^{\frac{3}{2}}. \tag{6}$$

Furthermore, the Froude number and the ratio $\tilde{H}/H$ of the depths at infinity are related by

$$\frac{\tilde{H}}{H} = \frac{2}{\sqrt{1 + 8/F^2} - 1} \tag{7}$$

(see Binnie[6]). At the separation points B and C, the behavior of the flow is similar to those of free-streamline problems. Therefore, we expect the flow to behave at these separation points like

$$\zeta \sim G + H[f - (\mp b + i)]^{\frac{1}{2}} \qquad \text{as } f \to (\mp b + i). \tag{8}$$

Here $b + i$ and $-b + i$ denote the values of f at C and B respectively, and G and H are constants to be found as part of the solution.

We solve the problem by following the series truncation procedure used by Vanden-Broeck and Keller[7], Vanden-Broeck[8]. Using (3), (4) and (8), we represent ζ by the expansion

$$\zeta(t) = F^{\frac{2}{3}} \exp\Bigg[A(1-t)^{2\lambda} + B_1\left(t^2 + 1 - 2t\cos\beta\right)^{\frac{1}{2}} + B_2\left(t^2 + 1 + 2t\cos\beta\right)^{\frac{1}{2}}$$

$$-B_1(2 - 2\cos\beta)^{\frac{1}{2}} - B_2(2 + 2\cos\beta)^{\frac{1}{2}} + \sum_{n=1}^{\infty} a_n\left(t^n - 1\right)\Bigg]. \tag{9}$$

The coefficients a_n and the constants A, B_1, B_2 are to be determined so that $\zeta(t)$ satisfies (2) and the kinematic conditions on AD and BC. It can easily be seen that the kinematic condition $v = 0$ on AD is satisfied if we require a_n, A, B_1, B_2 to be real. Differentiating (2) with respect to σ, we obtain

$$u(\sigma)u_\sigma(\sigma) + v(\sigma)v_\sigma(\sigma) - \frac{2}{\pi \sin\sigma}\frac{v(\sigma)}{u^2(\sigma)+v^2(\sigma)} = 0 \quad \text{on AB and CD.} \tag{10}$$

The kinematic condition on BC can be expressed as

$$v(\sigma) = -u(\sigma)\tan\gamma \qquad \text{on BC.} \tag{11}$$

Note that γ is measured clockwise from the negative $x-$ axis. The problem is now to find the coefficients a_n and the constants A, B_1, B_2 in (9) so that (10) and (11) are satisfied. Once ζ is known, we can calculate the profile of the free surface by integrating $\frac{\partial x}{\partial \phi} + i\frac{\partial y}{\partial \phi} = \frac{1}{\zeta}$ along the circumference of the unit circle

$$z(\sigma) = z\left(\frac{\pi}{2}\right) - \frac{2}{\pi}\int_{\frac{\pi}{2}}^{\sigma} \frac{1}{\zeta(s)\sin(s)}ds \qquad \text{for } 0 \le \sigma \le \pi,$$

where

$$z\left(\frac{\pi}{2}\right) = \int_0^i \frac{df}{dt}\frac{1}{\zeta}dt.$$

The upstream Froude number , defined as $\tilde{F} = \frac{\tilde{U}}{\sqrt{g\tilde{H}}}$, can be obtained from

$$\tilde{F} = |\zeta(-1)|^{\frac{3}{2}}.$$

We approximate the problem numerically by truncating the infinite series in (9) after N terms. There are N+5 unknowns, $\gamma, \lambda, A, B_1, B_2$ and $\{a_i\}_{i=1}^N$, for given values of β and F^2. We define the N+3 collocation points

$$\sigma_i = \frac{\pi}{N+3}\left(i - \frac{1}{2}\right), i = 1, \ldots, N+3. \tag{12}$$

For simplicity, we restrict the angle β to the form

$$\beta = \frac{\pi}{N+3}B, \tag{13}$$

where B is an integer smaller than $(N+3)/2$. We obtain N+3 equations by satisfying (10) at $\sigma_i, i=1,\ldots,B-1$ and $i=(N+3)/2+B,\ldots,N+3$, and (11) at $\sigma_i, i=B,\ldots,(N+3)/2$ $+B-1$. Relation (7) provides another equation. The last equation is obtained by relating the Froude number to the upstream and the downstream velocities by using (7) and the conservation of mass $UH=\tilde{U}\tilde{H}$. This yields

$$F^2 = \frac{2|\zeta(1)|^2}{|\zeta(-1)|(|\zeta(1)|+|\zeta(-1)|)}. \tag{14}$$

The system of N+5 nonlinear equations is solved by Newton's method.

The numerical scheme was used to compute solutions for various values of F^2 and β. For F^2 close to one, the coefficients a_n decrease rapidly (see table 1). Most of the results presented here were obtained with $N=400$.

Typical profiles are shown in Figs.1 and 2. As $\beta \to \pi/2$, the length L of the gate tends to zero and the flow reduces to a uniform stream. As $\beta \to 0$, $L \to \infty$. Numerical values of L versus F^2 for $\beta = \frac{\pi}{3}, \frac{2\pi}{5}$ and $\frac{19\pi}{40}$ are shown in Fig.3.

In Fig.4 we present numerical values of the gate inclination γ versus F^2 for various values of β. We define the contraction coefficient C_c as the dimensionless reciprocal of y−coordinate of the point C. In Fig.5 we present numerical values of C_c versus F^2. For F^2=1, the flow reduces to a uniform stream and C_c=1. For each value of F^2, the ratio $\tilde{H}/H$ can be evaluated by using (7).

As F^2 increases, the coefficients a_n decrease less rapidly and more collocation points are needed to obtain accurate solutions. More than 600 collocation points are required to compute accurate solution for $F^2 \geq 1.35$. Therefore we presented only solutions for $1 \leq F^2 < 1.35$.

Acknowledgement

This work was supported by the National Science Foundation.

References

[1]B. Benjamin, "On the flow in channels when rigid obstacles are placed in the stream", J. Fluid Mech. **1**, 227 (1956).

[2]D.D. Fangmeier and T.S. Strelkoff, "Solutions for gravity flow under a sluice gate", J. Eng. Mech. **94**, 153 (1968).

[3]B.E. Larock, "Gravity-affected flow from planar sluice gates", J. Eng. Mech. **96**, 1211 (1969).
[4]Y.K. Chung, "Solution of flow under sluice gates", J. Eng. Mech. **98**, 121 (1972).
[5]A.H-D. Cheng, J.A. Liggett and P.L-F. Liu, "Boundary calculations of sluice and spillway flows", J. Hydraulics Div., ASCE, **107**, No. HY 10, 1163 (1981).
[6]A.M. Binnie, "The flow of water under a sluice-gate", Quart. J. Mech. and Applied Math. **5** part 4, 395 (1952).
[7]J.-M. Vanden-Broeck and J.B. Keller, "Weir flows", J. Fluid Mech. **176**, 283 (1987).
[8]J.-M. Vanden-Broeck, "Flow under a gate", Phys. Fluids **29**, 3148 (1986).

LIST OF FIGURES

F^2	a_1	a_{100}	a_{200}	a_{300}	a_{400}
1.05	0.71×10^{-1}	0.15×10^{-3}	-0.11×10^{-4}	0.11×10^{-7}	-0.50×10^{-9}
1.10	0.19	0.37×10^{-3}	-0.42×10^{-4}	-0.17×10^{-4}	0.89×10^{-8}
1.15	0.38	0.68×10^{-3}	-0.76×10^{-4}	-0.31×10^{-4}	0.45×10^{-7}
1.20	0.65	0.11×10^{-2}	-0.12×10^{-3}	-0.49×10^{-4}	0.12×10^{-6}

Table 1

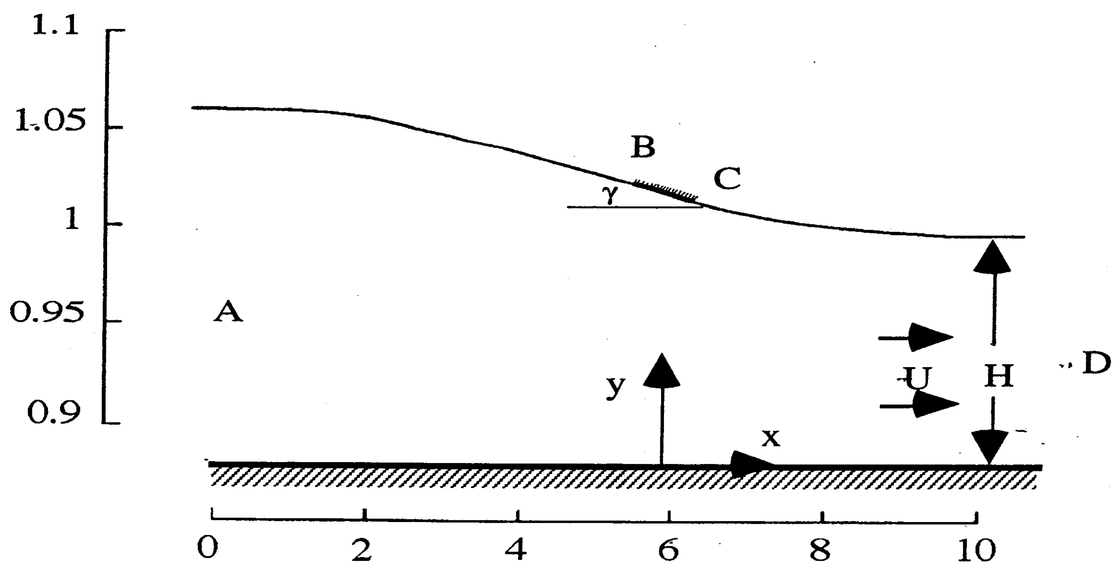


Figure 1

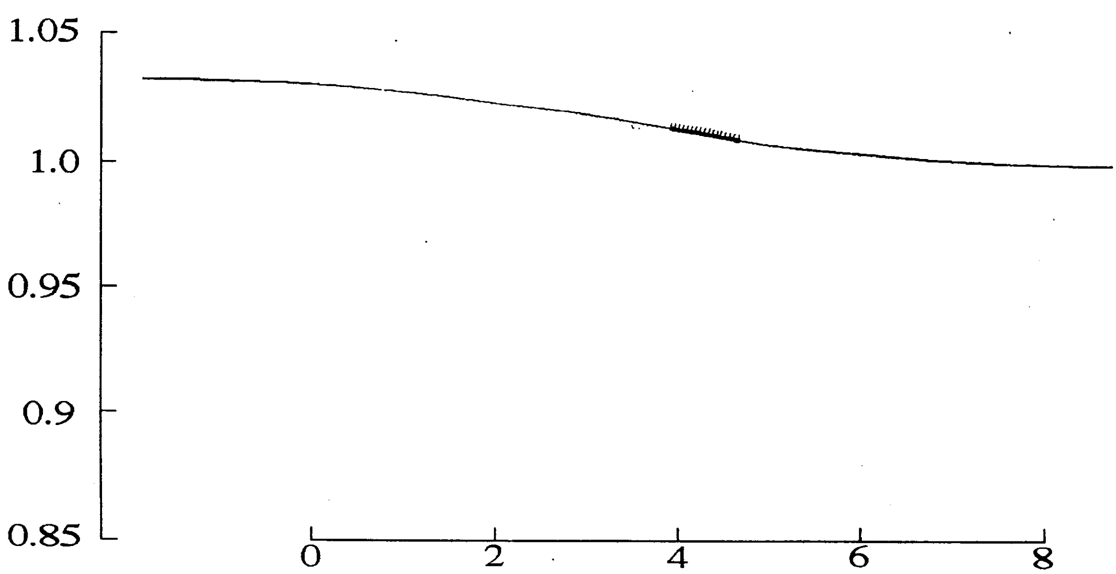


Figure 2

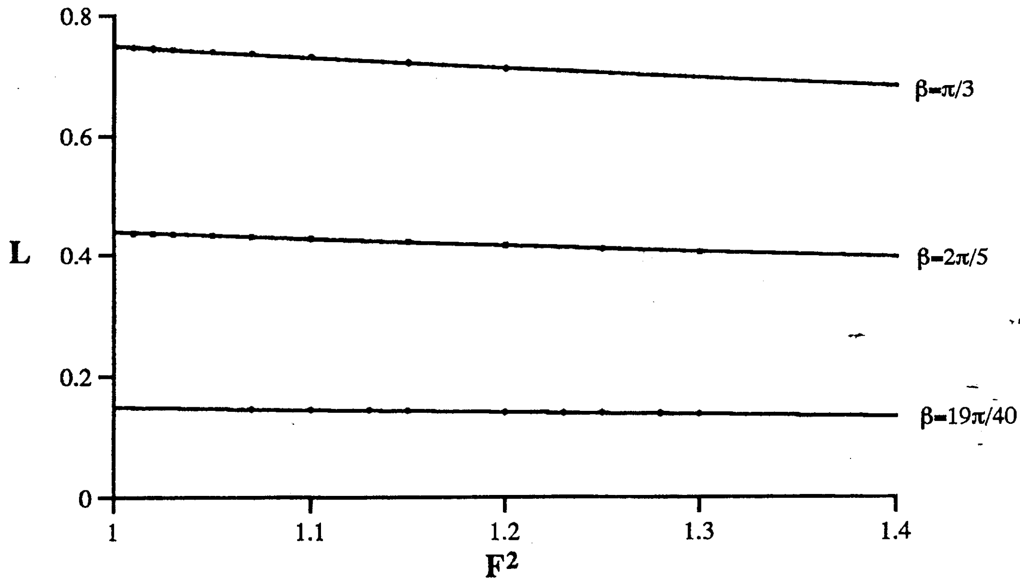


Figure 3

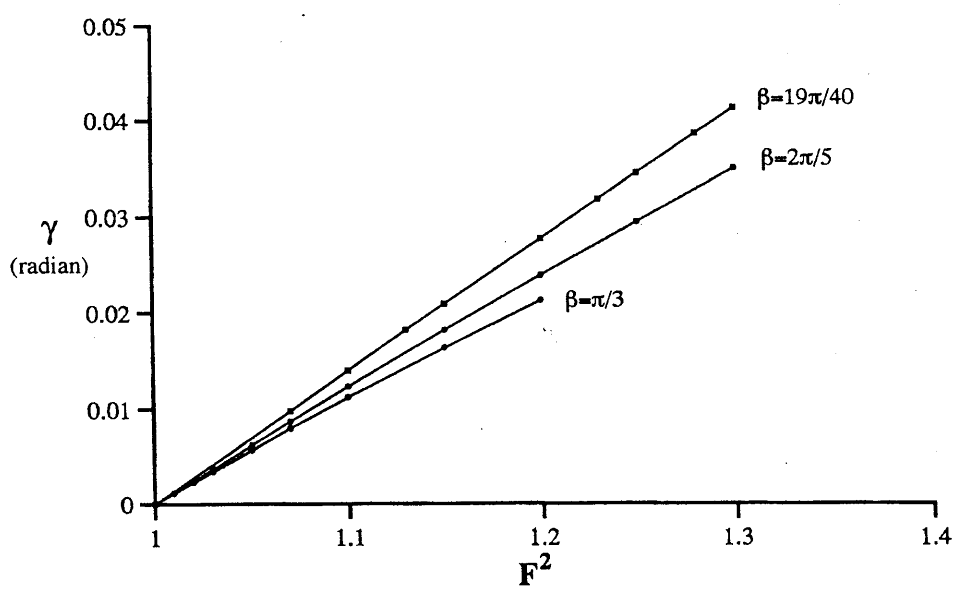


Figure 4

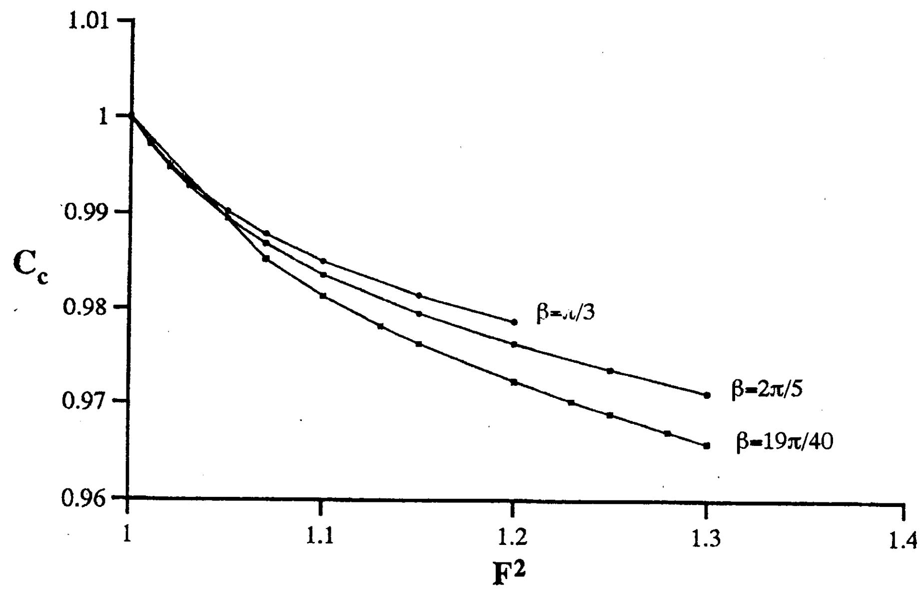


Figure 5

CONVERTING MULTIPOLYNOMIAL SYSTEM INTO EIGENPROBLEM VIA WELL ARRANGED BASIS

Feng Guo-chen†, Zhang Shu-gong† and Huang Kai‡
†Institute of Mathematics, Jilin University, Changchun130023, PRC
‡Department of Mathematics, Beijing University, Beijing,100871, PRC

§1 Introduction

In paper [FWZ], the authors have shown that for a given multipolynomial system one can construct an equivalent system and build up a joint eigenproblem, then one can find all solutions of the given system if one could solve this joint eigenproblem. Especially in the case that the joint eigenproblem can be reduced to an ordinary eigenproblem, things will get easier to deal with. However, apart from the situation that the projection of the solution set of the given system onto each axis is infinite, it is strongly dependent on the construction of the equivalent system that whether the joint eigenproblem can be reduced to an ordinary one. In this paper, an algorithm based on Gauss-Jordanian elimination to construct the equivalent system will be proposed. In detail, based on an autoreduced polynomial system (at the begining) or a so-called Well Arranged Basis (WAB) (in the middle), we construct a WAB with respect to some integer t which generates the same polynomial ideal as the initial one does. Then with this WAB, a criterion, which garantees that the joint eigenproblem formed from this WAB can be reduced to an ordinary one, is checked. If it is not satisfied, next WAB w.r.t. some larger t is constructed based on the former. Proceeding in this way, the algorithm gives either a generalized eigenproblem which can be reduced to an ordinary one or a sequence of WAB's. In the latter case, it will be shown that the sequence is finite and the finally obtained WAB contains a reduced Groebner basis. If the ideal generated by the initial system is zero-dimensional, the algorithm will certainly generate an ordinary eigenproblem for every unknown.

Let N_0^n be the set of all n-tuples with all nonnegative integers as their coordinates. For $i = (i_1, i_2, \cdots, i_n) \in N_0^n$, i_ν, the ν-th coordinate of i, is denoted by $Coor_\nu(i) = i_\nu$. Then a power product $x^i = x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$ and its degree tuple i can be uniquely determined by each other.

Let $i, j \in N_0^n$ be two tuples. Then tuple $i + j$ is defined by $Coor_\nu(i+j) = Coor_\nu(i) + Coor_\nu(j)$, $\nu = 1, 2, \cdots, n$. For a positive integer s, si is defined by $Coor_\nu(si) = sCoor_\nu(i)$. If there is a tuple l such that $i = j + l$, then i is said to be a multiple of j, denoted by $i \gg j$. The ν-th unit tuple $e^{(\nu)}$ is defined by $Coor_\mu(e^{(\nu)}) = \delta_{\mu\nu}$, where $\delta_{\mu\nu}$ is the Kronecker notation.

For n-tuples (Thus power products), we will adopt total degree ordering throughout this paper. A tuple $i = (i_1, i_2, \cdots, i_n)$ is said to be higher than tuple $j = (j_1, j_2, \cdots, j_n)$, denoted by $i > j$, iff the first nonzero entry in $(|i| - |j|, i_1 - j_1, \cdots, i_n -$

$j_n)$ is positive, where $|i| = \sum_{\nu=1}^{n} i_\nu$.

Let $f \in C[x_1, \cdots, x_n]$. By $LP(f)$ we denote the leading power product of f with respect to total degree ordering, $LD(f)$ the degree tuple of $LP(f)$ and $deg(f)$ the total degree of f.

For convenience, we briefly describe the eigenvalue method. Given a polynomial system

$$f_\mu(x) = \sum_j a_j^{(\mu)} x^j = 0, \quad a_j^{(\mu)} \in C, \ \mu = 1, 2, \cdots, \sigma. \tag{1.1}$$

Let $t \geq max_\mu\{deg(f_\mu)\}$ be some integer. Then

$$\left\{ \begin{array}{rll} x^i f_\mu(x) = \sum_j a_j^{(\mu)} x^{i+j} & = 0, & |i| \leq t - deg(f_\mu), \ \mu = 1, \cdots, \sigma. \\ x^i(x_\nu - \lambda_\nu) & = 0, & |i| < t \end{array} \right. \qquad (1.2)_\nu$$

$$\nu = 1, \cdots, n.$$

Constitute a joint eigenproblem. They can also be put into the matrix form

$$\begin{pmatrix} A_{11} & A_{10} \\ A_{01} & A_{00} - \lambda_\nu I \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_0 \end{pmatrix} = 0 \qquad (1.3)_\nu$$

$$\nu = 1, \cdots, n.$$

The following theorems hold true about the above eigenproblem (1.2) or (1.3).

Theorem 1.1 [FWZ] $\{\lambda_\nu = \xi_\nu, \nu = 1, \cdots, n.\}$ is a zero of system (1.1) iff it is a joint eigenvalue of the joint eigenproblem $(1.2)_\nu, \nu = 1, 2, \cdots, n$. □

Theorem 1.2 [FWH] If the zero set of system (1.1) has an irreducible subvariety V_0 for which the projection onto x_ν-axis is α, then α is a regular eigenvalue of the generalized eigenproblem $(1.2)_\nu$ for t large enough. Conversely, if t large enough and α is a regular eigenvalue of $(1.2)_\nu$, then α is the ν-th component of some zero of system (1.1). □

Theorem 1.3 [FWZ] If the ideal generated by $\{f_1, \cdots, f_\sigma\}$ is zero-dimensional and t large enough, then for each $\nu, 1 \leq \nu \leq n$, $(1.2)_\nu$ can be reduced into an ordinary eigenproblem. □

§2 Forming Eigenproblem with WAB

Let $PS = \{f_1, \cdots, f_\sigma\}$ be a system as given in Section 1, F the ideal generated by PS. Suppose $WAB(PS, d, t)$ is a new polynomial system which is constructed based on PS and integer d, t. $WAB(PS, d, t)$ possesses the following properties

1) $WAB(PS, d, t)$ has the same zero set as PS;

2) each polynomial in $WAB(PS, d, t)$ is monic and at most of total degree t;

3) for any $f \in WAB(PS, d, t)$, if $deg(x^i f) \leq t$, then $x^i f$ can be linearly represented by polynomials in $WAB(PS, d, t)$. Such a polynomial system is referred to as Well Arranged Basis which we will discuss in Section 4. We will use $WAB(PS, d, t)$ to form eigenproblem instead of $\{x^i f_\mu \mid |i| \leq t - deg(f_\mu), \ \mu = 1, 2, \cdots, \sigma\}$ in (1.2).

Let $T = \{LD(F) \in N_0^n \mid f \in WAB(PS,d,t)\}$, $T^c = \{i \in N_0^n \mid \mid i \mid \leq t\} \backslash T$. Then for each $i \in T$, there is a polynomial $h_i \in WAB(PS,d,t)$ which can be written as

$$h_i = x^i - \sum_{j \in T^c} a_j^{(i)} x^j \tag{2.1}$$

or equivalently

$$x^i = \sum_{j \in T^c} a_j^{(i)} x^j \quad mod\, F \tag{2.2}$$

while for each $i \in T^c$, if we define

$$a_j^{(i)} = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases} \tag{2.3}$$

then x^i can also be expressed as

$$x^i = \sum_{j \in T^c} a_j^{(i)} x^j \tag{2.4}$$

Let $D(i) = \{j \in T^c \mid a_j^{(i)} \neq 0\}$ for each i, $|i| \leq t$. If S is a subset of $\{i \in N_0^n \mid |i| \leq t\}$, then $D(S)$ is defined as

$$D(S) = \bigcup_{i \in S} D(i) \tag{2.5}$$

For each $\nu \in \{1, 2, \cdots, n\}$, we consider the tuple set

$$S_\nu = \{0, e^{(\nu)}, 2e^{(\nu)}, \cdots, te^{(\nu)}\} \tag{2.6}$$

Then a criterion for checking whether a generalized eigenproblem can be reduced to an ordinary one can be stated as following

Proposition 2.1 (Criterion for EIG). If there exists an $\nu \in \{1, 2, \cdots, n\}$ such that

$$D(e^{(\nu)} + D(S_\nu)) \subset D(S_\nu) \tag{2.7}$$

then the generalized eigenproblem for x_ν formed from PS can be reduced to an ordinary one.

Proof Arrange the tuples in $D(S_\nu)$ as

$$i^{(\tau)} > i^{(\tau-1)} \cdots > i^{(1)} > i^{(0)} = 0 \tag{2.8}$$

Then correspondingly there exists a vector

$$W = (x^{i^{(\tau)}}, x^{i^{(\tau-1)}}, \cdots, x^{i^{(1)}}, 1)^T \tag{2.9}$$

From Criterion (2.7), there exists a matrix $A^{(\nu)} \in C^{(\tau+1)\times(\tau+1)}$ such that

$$x_\nu W = A^{(\nu)} W \quad mod\, F \tag{2.10}$$

which implies that the generalized eigenproblem for x_ν can be reduced to an ordinary one through interchanging rows and columns in $(1.3)_\nu$. □

Now we can state our main algorithm which generates either an eigenproblem which can be reduced to an ordinary one or a Groebner basis for the ideal F. A polynomial set PS is said to be autoreduced if the leading power product of a polynomial in PS is never a multiple of the leading power product of other polynomial in PS.

Algorithm EorG.

input: PS autoreduced

output: an eigenproblem for some ν or a reduced Groebner basis.

$d := 0$

$t := t(PS)$, some integer determined by PS

Repeat

  $PS := WAB(PS, d, t)$

  If there exists $\nu, 1 \leq \nu \leq n$ such that $D(e^{(\nu)} + D(S_\nu)) \subset D(S_\nu)$,

    form the ordinary eigenproblem for x_ν.

  else

    $d := t$

    $t := t(PS)$

until $t \leq d$.

Theorem 2.2 The Algorthm **EorG** will yield either an eigenproblem which can be reduced to an ordinary one at some stage or a reduced Groebner basis for the ideal within finite steps. □

Theorem 2.3 If the ideal F is zero-dimensional, then for any $\nu \in \{1, 2, \cdots, n\}$, Algorithm **EorG** can generate an ordinary eigenproblem for x_ν. □

Once the algorithm yields an ordinary eigenproblem for x_ν, we can easily reach the conclusion that x_ν is algebraic over field C. We can decompose the system (1.1) into some subsystem $\{f_1, \cdots, f_\sigma, x_\nu - \alpha_\mu^{(\nu)}\}$, where $\alpha_\mu^{(\nu)}$'s are the eigenvalues of $A^{(\nu)}$.

As for the case that the projection of the solution set onto every axis is infinite, our algorithm never gerenate an ordinary eigenproblem. In this case, one may deal with the generalized eigenproblem. Some authors have made some progress. Those who are inerested in this field may refer to [H],[HW].

§3 Tuples and Tuple sets

In this section, some notations and results presented in paper [W] will be used.

Suppose $T \subset N_0^n$ is a tuple set. It is said to be autoreduced if no i in T is a multiple of another j in T.

For an autoreduced tuple set T, we can define an n-tuple, $Max(T)$, as

$$Max(T) = (max_1(T), max_2(T), \cdots, max_n(T)) \tag{3.1}$$

where $max_\nu(T) = max\{Coor_\nu(i) \mid i \in T\}$.

Suppose T and T' are two autoreduced tuple sets with tuples arranged in increasing order

$$\begin{array}{l} T : i^{(1)} < i^{(2)} < \cdots < i^{(\delta)}; \\ T' : j^{(1)} < j^{(2)} < \cdots < j^{(\sigma)}. \end{array} \tag{3.2}$$

T is said to be higher than T' if either of the following two cases holds true:

1) there exists some $\mu \leq \delta$ and $\leq \sigma$ such that

$$i^{(\nu)} = j^{(\nu)} \text{ for } \nu < \mu \text{ while } i^{(\mu)} > j^{(\mu)}$$

2) $\delta < \sigma$ and $i^{(\nu)} = j^{(\nu)}$ for $\nu \leq \mu$.

Lemma 3.1 (Lemma 2, [W]) Let T be an autoreduced set and i a tuple which is not a multiple of any tuple in T. Let T' be the autoreduced set obtained by adjoining i to T and then removing all tuples in T which are multiples of i. Then T' is of lower order than T. □

Lemma 3.2 (Lemma 4, [W]) Any sequence of autoreduced sets steadily decreasing in order is finite. □

Associated with an autoreduced tuple set T, we define

$$\begin{array}{rl} X(T) = \{i \in T \mid & \text{there is } \nu \in \{1, \cdots, n\} \text{ such that } Coor_\nu(i) \neq 0 \\ & \text{while } Coor_\mu(i) = 0,\ \mu \neq \nu\} \end{array} \tag{3.3}$$

and

$$t(T) = \begin{cases} |Max(T)|, & Ind(T) < 2, \\ |Max(T)| - Ind(T) + 1, & \text{otherwise} \end{cases} \tag{3.4}$$

where $Ind(T)$=the cardinal number of $X(T)$.

If T is an arbitrary tuple set, we can define an autoreduced subset of T, KT, as

$$KT = \{i \in T \mid \text{there is no } j \in T \text{ such that } i \gg j\} \tag{3.5}$$

§4 Well Arranged Basis

Suppose $PS = \{f_1^{(0)}, f_2^{(0)}, \cdots, f_{\sigma_0}^{(0)}\}$ is a polynomial set. Its leading degree tuple set is denoted by $T(PS)$. F is the polynomial ideal generated by PS and $t \geq max_\nu\{deg(f_\nu^{(0)})\}$ some integer. Based on PS, we now construct a so-called Well Arranged Basis for the ideal F. Let

$$\begin{array}{rcl} PS_0 : & = & PS \\ RS_0 : & = & \{x_j f_\nu^{(0)} \mid 0 < |j| \leq t - deg(f_\nu^{(0)}), \\ & & \quad j \text{ is not a multiple of } LD(f_\mu), \mu < \nu, \nu = 1, 2, \cdots, \sigma.\} \\ BS_1 : & = & PS_0 \bigcup RS_0 \\ I_1 : & = & T(BS_1) \\ PS_1 : & = & GJ(BS_1), \text{ the resulting polynomial set by applying} \\ & & \quad \text{Gauss-Jordanian elimination to } BS_1 \\ J_1 : & = & T(PS_1), \text{ obviously we have } J_1 \supset I_1 \\ K_1 : & = & J_1 \backslash I_1, \text{ the newly occuring leading degree tuple set.} \end{array}$$

If $K_1 \neq \emptyset$, rearrange the polynomial set $\{f \in PS_1 \mid LD(f) \in K_1\}$ as $\{f_1^{(1)}, \cdots, f_{\sigma_1}^{(1)}\}$ with $LD(f_\nu^{(1)}) > LD(f_\mu^{(1)})$, $\nu < \mu$, where σ_1 is the number of elements in K_1. Then let

$$\begin{aligned} RS_1 : &= \{x^j f_\nu^{(1)} \mid 0 < |j| \leq t - deg(f_\nu^{(1)}),\ j \text{ is not a multiple of tuples} \\ &\qquad \text{in } I_1 \text{ or } LD(f_\mu^{(1)}) \text{ with } \mu < \nu,\ \nu = 1, 2, \cdots, \sigma_1\} \\ BS_2 : &= PS_1 \cup RS_1 \\ I_2 : &= T(BS_2) \\ PS_2 : &= GJ(BS_2) \\ J_2 : &= T(PS_2) \\ K_2 : &= J_2 \backslash I_2 \end{aligned}$$

Repeat the same procedure as above when $K_2 \neq \emptyset$. Doing in this manner, we get a sequence of polynomial sets

$$PS_0,\ PS_1,\ \cdots,\ PS_\mu,\ \cdots \tag{4.1}$$

and a sequence of tuple sets

$$I_1 \subset J_1 \cdots \subset I_\mu \subset J_\mu \subset \cdots \tag{4.2}$$

Since for all $\mu, I_\mu \subset J_\mu \subset \{i \in N_0^n \mid |i| \leq t\}$, there must exist some $\bar{\mu}$ such that $I_{\bar{\mu}} = J_{\bar{\mu}}$. Hence the above process is finitely terminated. We call the finally obtained polynomial set $\overline{PS} := PS_{\bar{\mu}}$ a Well Arranged Basis with respect to t of the ideal F (cf. [W]). It is evident that the WAB $\overline{PS}$ depends on the integer t and the initial polynomial set PS.

For any polynomial set, when applying Gauss-Jordanian elimination to it, the resulting one possesses the property that the leading power product of some polynmial never occur in other polynomials. Such a polynomial set is said to be well arranged. The WAB $\overline{PS}$ is well arranged.

For any well arranged polynomial set H, we can define an autoreduced subset of it

$$K(H) = \{h \in H \mid LD(h) \in KT(H)\} \tag{4.3}$$

To discuss the properties of the WAB defined as above, we need the following

Lemma 4.1 (cf. [H], Proposition 3) Let $\{g_1, \cdots, g_\sigma\}$ be a polynomial set, $t \geq max_\nu\{deg(g_\nu)\}$ some integer. Then the polynomial sets

$$\begin{aligned} \{x^j g_\nu \ \mid 0 \leq |j| \leq t - deg(g_\nu),\ j \text{ is not a multiple of} \\ LD(g_\nu) \text{ with } \mu < \nu,\ \nu = 1, \cdots \alpha.\} \end{aligned} \tag{4.4}$$

and

$$\{x^j g_\nu \mid 0 \leq |j| \leq t - deg(g_\nu),\ \nu = 1, \cdots, \alpha.\} \tag{4.5}$$

can be linearly expressed (over C) by each other. □

By Lemma 4.1, the polynomial set

$$H = \bigcup_{\delta=0}^{\bar{\mu}} \{x^j f_\nu^{(\delta)} \mid 0 \le |j| \le t - deg(f_\nu^{(\delta)}),\ \nu = 1, \cdots, \sigma_\delta\} \tag{4.6}$$

and $\overline{PS}$ can be linearly expressed by each other.

Corollary 4.2 Any linear combination of polynomials in H is either a zero polynomial or a polynomial with leading degree tuple belonging to $T(\overline{PS})$. Hence $T(\overline{PS}) = T(H)$. □

Let

$$G = \bigcup_{\delta=0}^{\bar{\mu}} \{f_1^{(\delta)}, \cdots, f_{\sigma_\delta}^{(\delta)}\} \tag{4.7}$$

It is easy to see that for any $i \in T(\overline{PS})$, there exists at least one $j \in T(G)$ such that $i \gg j$ and vise versa.

Lemma 4.3 For each $f \in \overline{PS}$ holds relation

$$f = \sum_{\nu=0}^{\bar{\mu}} \sum_{\delta=1}^{\sigma_\nu} g_\delta^{(\nu)} f_\delta^{(\nu)},\quad deg(g_\delta^{(\nu)} f_\delta^{(\nu)}) \le deg(f) \tag{4.8}$$

Proof Reduce f with respect to G. Suppose

$$f = \sum_{\nu=0}^{\bar{\mu}} \sum_{\delta=1}^{\sigma_\nu} g_\delta^{(\nu)} f_\delta^{(\nu)} + R(x),\ deg(g_\delta^{(\nu)} f_\delta^{(\nu)}) \le deg(f)$$

We claim that $R(x) = 0$. If not, then $R(x) = f - \sum_{\nu=0}^{\bar{\mu}} \sum_{\delta=1}^{\sigma_\nu} g_\delta^{(\nu)} f_\delta^{(\nu)}$ is a linear combination of polynomials in H. By Corollary 4.2, $LD(R) \in T(\overline{PS})$, $LD(R)$ therefore must be a multiple of some tuple in $T(G)$. This contradicts to the defintion of $R(x)$. □

Corollay 4.4 For any x^i and $f \in \overline{PS}$ with $deg(x^i f) \le t$ holds

$$x^i = \sum_{h \in \overline{PS}} a_h h,\ a_h \in C \tag{4.9}$$

□

Proposition 4.5 Suppose $K(\overline{PS}) = \{h_1, \cdots, h_\beta\}$ Then for any $f \in \overline{PS}$, there exist $g_\nu \in C[x_1, \cdots, x_n]$ with $deg(g_\nu h_\nu) \le deg(f)$ such that

$$f = \sum_{\nu=1}^{\beta} g_\nu h_\nu \tag{4.10}$$

Proof Reduce f w.r.t. $K(\overline{PS})$. Suppose $R(x) = f - \sum_{\nu=1}^{\beta} g_\nu h_\nu$, $deg(g_\nu h_\nu) \le deg(f)$. Then Corollary 4.4 implies $R(x)$ is a linear combination of polynomials in $\overline{PS}$. Hence it must be zero. □

The above propostion says that $K(\overline{PS})$ is a system of generators for the ideal F. So, we call $K(\overline{PS})$ a reduced Well Arranged Basis for F

Theorem 4.6 Suppose PS is an autoreduced polynomial set, $\overline{PS}$ is a WAB w.r.t. t obtained through the procedure as before, then

$$KT(PS) > KT(\overline{PS}) \tag{4.11}$$

Proof It is easy to see $KT(\overline{PS}) = KT(G)$. Let $T_0 = KT(PS)$, T_1 be the tuple set obtained by adding $LD(f_1) \in T(G)\backslash T_0$ into T_0 then removing all multiples of $LD(f_1)$. Thus by Lemma 3.1, $T_0 > T_1$. Then adding $LD(f_2) \in T(G)\backslash\{T_0 \cup T_1\}$ into T_1 and removing multiples of $LD(f_2)$. We get T_2. Again by Lemma 3.1, $T_0 > T_1 > T_2$ Since $T(G)$ is finite, Doing in this way, within finitely many steps for some τ we get

$$KT(PS) = T_0 > T_1 > \cdots > T_\tau = KT(G) = KT(\overline{PS}). \tag{4.12}$$

□

For convenience , we simply denote $t(KT(PS))$ by $t(PS)$ hereafter.

Theorem 4.7 If PS is a WAB with respect to t and $t \geq t(PS)$, then $K(PS)$ is a reduced Groebner basis for the ideal generated by PS.

Proof Assume $K(PS) = \{h_1, \cdots, h_\alpha\}$. By Theorem 6.2 of Buchberger's [B], we need only to show that for any $h_\mu, h_\mu \in K(PS)$, the S-polynomial of h_μ, h_ν can be reduced to zero with respect to $K(PS)$. We consider two possibility.

1) both of $LP(h_\mu)$ and $HP(h_\nu)$ belong to $X(KT(PS))$, then

$$LCM(LP(h_\mu), LP(h_\nu)) = LP(h_\mu)LP(h_\nu). \tag{4.13}$$

Hence by Lemma 6.4 of Buchberger's [B], the S-polynomial of h_μ and h_ν can be reduced to zero with respect to $K(PS)$.

2) at least one of $LP(h_\mu)$ and $LP(h_\nu)$ does not belong to $X(KT(PS))$. Define a tuple m by

$$Coor_\nu = \begin{cases} Coor_\nu(Max(KT(PS))) - 1, & \text{there exists } i \in X(KT(PS)) \\ & \text{such that } Coor_\nu(i) \neq 0 \\ Coor_\nu(Max(KT(PS))), & \text{else} \end{cases} \tag{4.14}$$

Then there exists δ, dependent on h_μ and h_ν, such that

$$LD(H_\mu), LD(h_\nu) < m + e^{(\delta)} \tag{4.15}$$

Thus

$$\begin{aligned} deg(LCM(LP(h_\mu), LP(h_\nu))) &\leq |m + e^{(\delta)}| \\ &\doteq |Max(KT(PS))| - Ind(KT(PS)) + 1 \\ &= t(PS) \leq t \end{aligned} \tag{4.16}$$

Suppose x^i and x^j satisfy

$$S - polynomial(h_\mu, h_\nu) = x^i h_\mu - x^j h_\nu \tag{4.17}$$

where $deg(x^i h_\mu) = deg(x^j h_\nu) \leq t$. Then by Corollary 4.4 and Proposition 4.5, the S-polynomial of h_μ and h_ν can be reduced to zero with respect to $K(PS)$.

□

In the following we give the algorithm to construct the WAB.

sub algorithm WAB( )

$WAB(PS, d, t)$

input: PS := the initial polynomial set(WAB or an autoreduced polynomial set)

t := some integer with respect to which the WAB being constructed

$$d := \begin{cases} max\{deg(f) \mid f \in PS\}, PS \text{ is } WAB \\ 0, PS \text{ is an autoreduced polynomial set.} \end{cases}$$

output: a WAB with respect to t.

$$\begin{aligned} RS : &= \{x^j f \mid d - deg(f) < |j| \leq t - deg(f), j \text{ is not multiple} \\ &\qquad \text{of } LD(g) \text{ with } LD(g) > LD(f), f, g \in K(PS)\} \\ BS : &= PS \bigcup RS \\ I : &= T(BS) \\ PS : &= GJ(BS) \\ J : &= T(PS) \\ K : &= J \backslash I \end{aligned}$$

while $K \neq \emptyset$, do

$$\begin{aligned} RS : &= \{x^j f \mid 0 < |j| \leq f - deg(f), j \text{ is not the multiple} \\ &\qquad \text{of tuples in } I, \text{ or } LD(g) \text{ with } LD(g) > LD(f), \\ &\qquad LD(g), LD(f) \in K, f, g \in PS\} \\ BS : &= PS \cup RS \\ I : &= T(BS) \\ PS : &= GJ(BS) \\ J : &= T(PS) \\ K : &= J \backslash I \end{aligned}$$

All the conclusions similar to Lemma 4.1 through Theorem 4.8 hold true for the WAB generated by Algorithm WAB( ). The prooves, which are similar to those of Lemma 4.9 through Theorem 4.8, are omitted.

Proof of Theorem 2.1 If the first case does not take place, then the Algorithm $EorG$ gives a sequence of WAB's

$$PS_0, PS_1, \cdots, PS_\mu, \cdots \tag{4.18}$$

which satisfy

$$KT(PS_0) > KT(PS_1) > \cdots > KT(PS_\mu) > \cdots \tag{4.19}$$

By Lemma 3.2, the sequence (4.19) is finite. Therefore, there must be some μ such that $KT(PS_{\bar{\mu}-1}) = KT(PS_{\bar{\mu}})$ which implies $t(PS_{\bar{\mu}}) \leq t(PS_{\bar{\mu}-1})$. Hence the algorithm is finitely terminated. Then Theorem 4.7 says that $K(PS_{\bar{\mu}})$ is a reduced Groebner basis for the ideal F. □

Proof of Theorem 2.2 In the case that F is zero-dimensional, the algorithm at most generates a Groebner basis $K(PS_{\bar{\mu}})$ which contains polynomials with $x_\nu^{i_\nu}, \nu = 1, \cdots, n$, for some i_ν, as their leading power products. Thus we have $Max(KT(PS_{\bar{\mu}})) = (i_1, \cdots, i_n)$. Consider all power products with total degree $t = t(PS_{\bar{\mu}}) = \sum_{\nu=1}^{n} i_\nu - n + 1$. Each such a power product must have some (at least one) $x_\nu^{i_\nu}$ as its divisor. Hence

$$T(PS_\mu) \supset \{i \in N_0^n \mid |i| = t(PS_\mu)\} \tag{4.20}$$

We can see that $T^c = \{i \in N_0^n \mid |i| \leq t(PS_\mu)\} \backslash T(PS_\mu) \supset D(S_\nu)$ for any $\nu \in \{1, 2, \cdots, n\}$. It is easy to see the Criterion (2.7) holds true for each ν. □

References

[FWZ] Feng Guo -chen, Wu Wen - da and Zhang Shu - gong. The Eigenvalue Problem Equivalent to Multivariate Polynomial System, Northeastern Math. J. 9(1)(1993) 5-8.

[W] Wu Wen -tsun, On the Construction of Groebner Basis of a Polynomial Ideal Based On Riquier-Janet Theory, MM Research Preprints, No.5,(1990) 5-22.

[H] Huang kai, The Eigenvalue Method for Solving Polynomial System, Ph.D. Thesis, Jilin University (1993).

[B] Buchberger, B,Groebner Basis: An Algorithmic Method in Polynomial Ideal Theory, Research Trends in Multidimensional Systems theory (ed. N. K. Bose), D. Reidel Publishing Company, Dordrecht(1985).

[HW] Huang Yu -zhen and Wu Wen -da, On the Computation of the Kronecker Form of a Special Matrix Pencil, MM Research Preprints, No.7(1992).

[FWH] Feng Guo -chen, Wu Wen -da and Huand Kai, Second Equivalence Theorem of Multivariate Polynomial System and Eigenproblem, Advance in Math. (in China) 22(5)(1993).

Testing for Cointegration and Convergence in Economic Time Series: An Illustration using Matlab

Nozar Hashemzadeh
Economics Department
Radford University
Radford, Virginia 24142
e-mail: nhashemz@ruacad.runet.edu

Abstract: The possibility of change in parameter(s) of a dynamical system has important implications of whether the system is inherently capable of producing an "equilibrium" over time. In a dynamic setting, variations in a system's parameters generate erratic time paths that cannot be predicted on theoretical grounds. As a result, forecasts made from such models will be unreliable and very hazardous. This paper explores the divergent behavior of regional income flows in the United States during the last six decades using a statistically estimated difference equation. The signal processing tools provided in the recent editions of Matlab are used to analyze and model data structure and characteristics.

1 Introduction

Conventional equilibrium theories suggest that any systematic disparity in regional income (wage) flows will be corrected by free trade and labor migration. In many instances, however, regional disparities in income (wages) do not fade over time but become more divergent.

The notion that free trade and efficient markets ensure a steady long-run equilibrium among regional income flows has been challenged vigorously in recent years. Eberts (1989) has suggested that incomplete information and mismatch between worker skills and job requirements and institutional barriers to mobility often lead to incomplete alignment of earnings and thus contribute to persistent differences in regional income levels. Eberts and Stone (1989) have corroborated that it nearly takes a decade for local labor markets to adjust fully to structural shocks.

This paper contends that disequilibrium and nonlinear behavior is not uncommon in economics. Early examples of disequilibrium and nonlinear behavior in economic processes are embodied in the cobweb model of price adjustment and in the works of Metzler (1941), Modigliani (1944) and Samuelson (1948). More recently, Brock and Sayers (1988) have provided tenable evidence in support of nonlinearity in data pertaining to unemployment, employment, industrial production, and pig iron production. Baumol and Wolf (1983), Baumol and Benhabib (1989), and Day and Shafer (1985, 1987) provide other examples of the presence of nonlinearity in economic relationships.

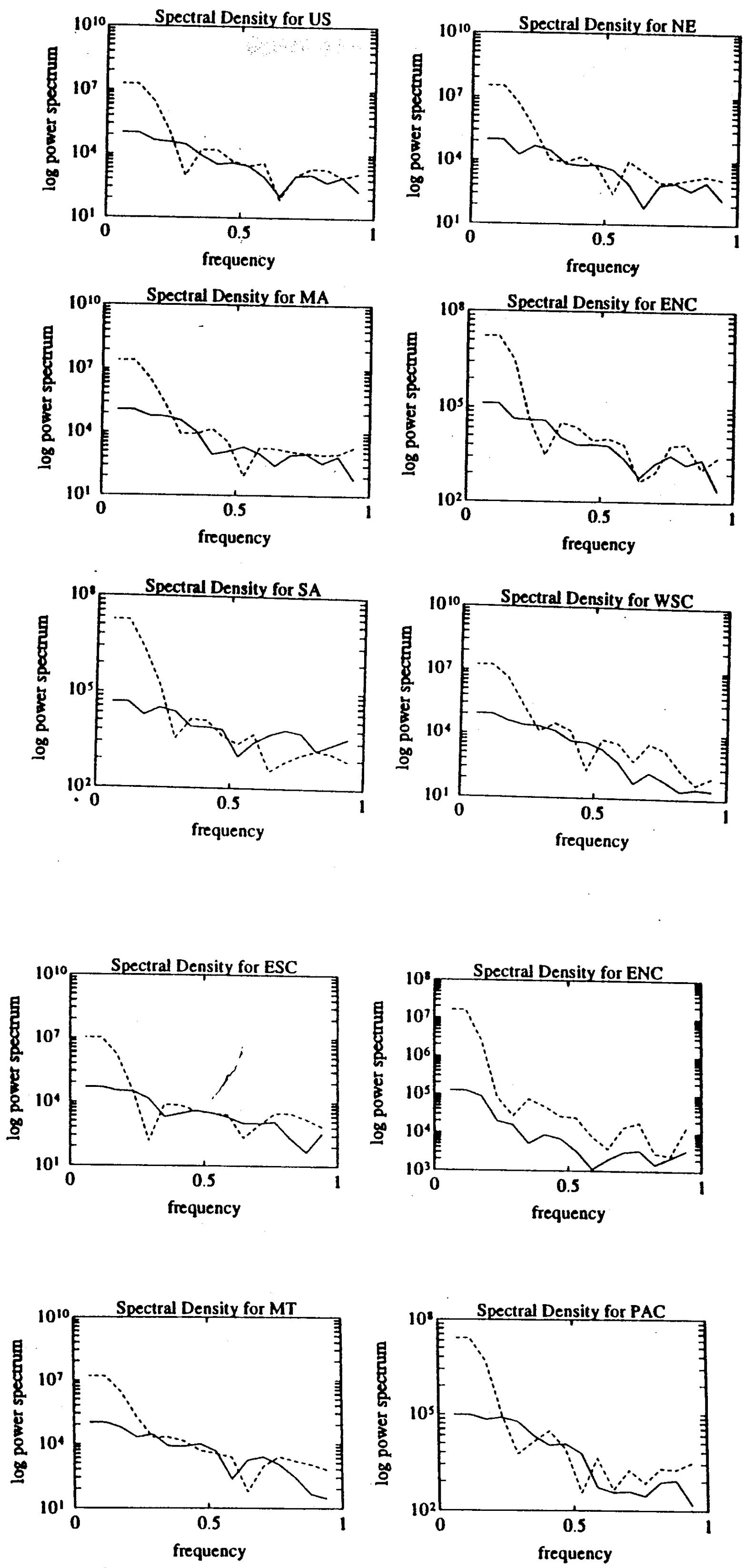
2 Objectives

The primary objective of the present inquiry is to describe and subsequently analyze the dynamics of inter-regional income interactions in an environment in which the system parameter change about their mean values. This approach is motivated by recent advances in the study of nonstationarity and nonlinear dynamical systems by David Ruelle (1988), H.W. Lorenz (1987a, 1987b), and others who have shown that coupling of systems (regions) can decrease the stability of the regional equilibria and may cause persistent fluctuations in wages, income and other economic variables. One of the distinct advantages of this approach is that it combines features of the traditional fixed-parameter(s) linear models of convergence processes with the more realistic nonlinear and nonstationarity processes. The insights gained from exploring the qualitative behavior of a non-linear system holds enormous potentials for improving and reconstructing our methods of model building and forecasting in the regional sciences.

3 Spectral Analysis of Income Data

In the familiar empirical models of competitive equilibria, it is assumed that the regional income flows share a common deterministic trend. Further, it is assumed that the underlying time series are stationary or can be made stationary by detrending the data. Notwithstanding, recent empirical evidence strongly disputes these assumptions. Using per-capita income for nine U. S. regions, we shall demonstrate that the nonlinearities that permeate economic relationships among regions are likely to cause differences in regional incomes (wages) to assume a variety of steady, periodic, aperiodic, or even random-like patterns. As David Ruelle (1988, p. 199), has noted free trade introduces a coupling between different economies with the likelihood of producing "nontrivial time dependence, possibly with sensitive dependence on the initial condition".

Figure 1


Spectral Density for US
Spectral Density for NE
Spectral Density for MA
Spectral Density for ENC
Spectral Density for SA
Spectral Density for WSC
Spectral Density for ESC
Spectral Density for ENC
Spectral Density for MT
Spectral Density for PAC
log power spectrum
frequency
0
0.5
1
10¹⁰
10⁷
10⁴
10¹
10⁸
10⁵
10²
10⁶
10³

We tested the logarithmically transformed data for non-stationarity using the simple Dickey and Fuller (Fuller 1976, Dickey and Fuller 1979) unit root procedure (DF) (results not reported) and spectral analysis. The spectral density plots shown in Figure I reveal the non-stationarity character of the time series. The spectral density functions in Figure I are designed to give a visual understanding of the behavior of income data in the frequency domain. For each region and the U.S., a spectral density function was fitted to equal sections of income data. Notice that the plots of the logarithm of the estimated spectral over the complete frequency range (0, ϖ) are quite different for the two data segments. Furthermore, the evidence from spectral analysis indicates that the income data are characterized by a series of nonlinear patterns. These non-linear patterns are consistent with those reported by James Hamilton (1989) in his study of the time behavior of the U.S. growth national product (gnp).

An obvious ramification of these results is that the non-stationary component must be removed from the data to make it amenable for modelling or forecasting purposes (Granger 1989.

4 The Inverted-U Model

In this section, we try to make the case that convergence patterns observed in percapita income data may be best represented by an autoregressive model that embodies the essential characteristics of the conventional inverted-U model. Patterns of regional income behavior are characterized by both conservative processes (forces that tend to add or prolong income inequalities such as growth poles, differences in labor productivity and returns to scale, and institutional factors) as well as dissipative processes (forces that take away or reduce such inequality such as migration, relocation and trade). These are also the characteristics embedded in the Inverted-U model of income inequality widely used in the development and the regional science literatures (Ram 1991; Lecaillon et al. 1984; Kuznets 1955, 1963; Sahota 1978, 1973; and Amos 1991, 1990). The inverted-U hypothesis implies that income divergence increases, reaches a peak and then decreases as dissipative forces take over.

The empirical model used in our inquiry combines elements of the inverted U-model with some recent developments in the field of non-linear dynamics. Although not identical to the theoretical framework used by Kuznets and others, our fundamental hypothesis is that the amount of income gap (measured here as the differences in the natural log of percapita income) between two neighboring regions in any period is a quadratic function of the differences that prevailed in the previous period. In our model, the time lag arises from the fact that labor and other resources are not instantaneously mobile between regions.

The model embedded in Equation 1, tests the hypothesis that the rate of convergence in wages or income levels is a quadratic function of the amount of differentials in wages or income that existed in the preceding period. In Equation 1, y^{ij}_t is the gap in per-capita incomes between regions i, and j during the current time period, and y^{ij}_{t-1} is the income gap that existed in the last period. When the two income series have converged to an equilibrium, the
$dy^{ij}_t/dt = 0$.

The time lag in Equation 1 was set to one period (year) after inspecting the autocorrelation function and the partial correlation function of the right-hand-side variable and the residuals from the equation.

$$(1)\quad y_t^{ij} = c_1 y_{t-1}^{ij} + c_2 y_{t-1}^{ij\,2} + \xi_t$$

In Equation 1, the first term on the right-hand side represents the influence of the conservative factors (factors that tend to escalate divergence), while the nonlinear term accounts for the dissipative forces (factors that limit divergence). Thus, one would expect c_1 to be positive and c_2 to have a negative sign. The relative importance of these forces and their interactions determine the time evolution of percapita incomes among regions. Careful analysis of these parameters should reveal which of the two forces is more dominant in shaping the time behavior of income gaps between spatially contiguous regions. Note that if the nonlinear parameter in Equation 1 is statistically insignificant ($c_2 = 0$), the quadratic term vanishes and the Equation is reduced to a linear first-order difference equation. As expected, the solution to the reduced form is stable for $|c_1| <1$ and unstable for $|c_1|>1$. In the latter case, the linear first-difference model predicts unbounded divergence which is unrealistic and without precedence in real capitalist regimes (Duncan K, Foley, 1991, p. 89). More importantly, the equation is capable of producing oscillatory behavior for values of $c_1 = 1$.

5 Parameters Estimates

Using per capita income data published by the U. S. Bureau of Economic Analysis, we estimated the parameters of Equation 1 for the New England and the

Mid-Atlantic regions for the years 1929-1991. These estimates have been corrected for significant autocorrelation using Matlab and the Yule-Walker procedure.

Least Squares Estimates of: $\log(y^{ij}_t) = c_1\log(y^{ij}_{t-1}) - c_2\log(y^{ij}_{t-1}{}^2) + \varepsilon_t$

1929-1988				
Variable	c Value	Std Error	T RATIO	App. Prob
Intercept	-0.00082878	0.00338896	-0.245	0.8077
$y^{ij}_{t-1})$	3.22671642	1.14560424	2.817	0.0068
$y^{ij}_{t-1}{}^2$	-2.16755815	1.07632646	-2.014	0.0490

Total Rsq: 0.8867

6 Some Complex Dynamics

What types of dynamics are implied by the empirical analogue of equation 6?. Specifically, does the model generate time paths that approach a steady state with the passage of time? To proceed with these and other questions the following definitions are in order. Using William A. Brock's simple definitions (1991, p. 258), a dynamic system is said to behave erratically if "when exogenous stochastic shocks are shut off, the resulting deterministic dynamics do not converge to a limit cycle or a fixed point as time tends to infinity". Furthermore, a dynamical process is said "to be chaotic if it displays sensitive dependence on initial conditions".

An important question here is whether one can predict the asymptotic behavior of the system in the long run over a range of parameters values. Experiments with quadratic equations (Blackburn et al, 1989) confirm that analytic answers to the above question cannot be easily deduced except for highly restrictive cases. Therefore, in order to understand the complex dynamical behavior of equation II, numerical solutions are derived for ranges of parameters values. The results are presented in the bifurcation diagrams in Figures 1, 2, 3, and 4. These diagrams are designed to show an overall view of the system's response to marginal changes in initial conditions.

The horizontal axis in Figures 3, 4, and 5 are the values of the parameter of the linear term in equation II while the vertical axes are the asymptotic values of y^{ij}_t after 30 to 120 iterations. For fixed values of c_2, the limit values of y^{ij}_t are computed and plotted against c as this parameter is gradually varied by an increment of 0.02 around its estimated mean value of 2.66. These figures demonstrate whether y^{ij}_t converges to a fixed value, two or more periodic values or becomes unpredictable as parameters are changed within one standard deviation of their estimated mean values.

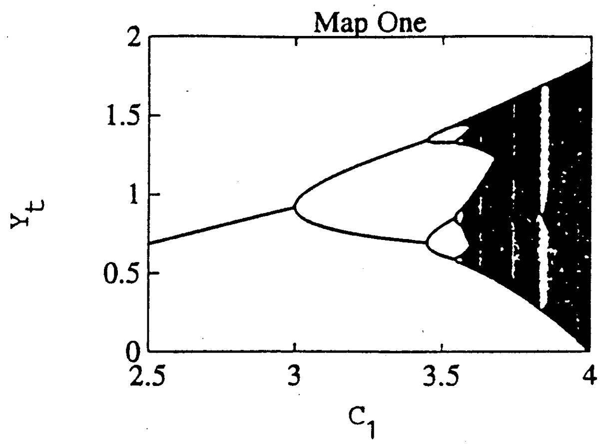


Figure 2: Bifurcation graph for
$y^{ij}_t = 3.226716y^{ij}_{t-1} - 2.16755{y^{ij}_{t-1}}^2$

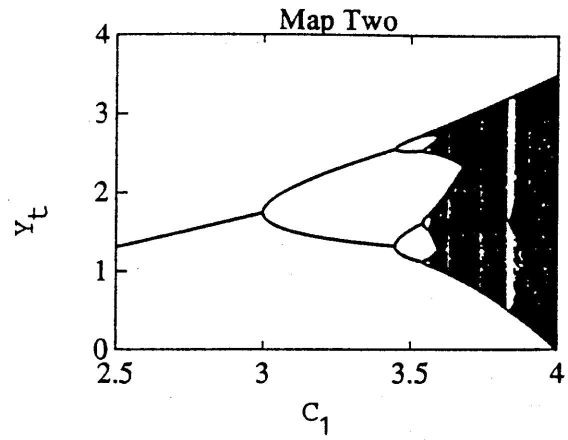


Figure 3: Bifurcation graph for
$y^{ij}_t = 2.077692y^{ij}_{t-1} - 2.16755{y^{ij}_{t-1}}^2$

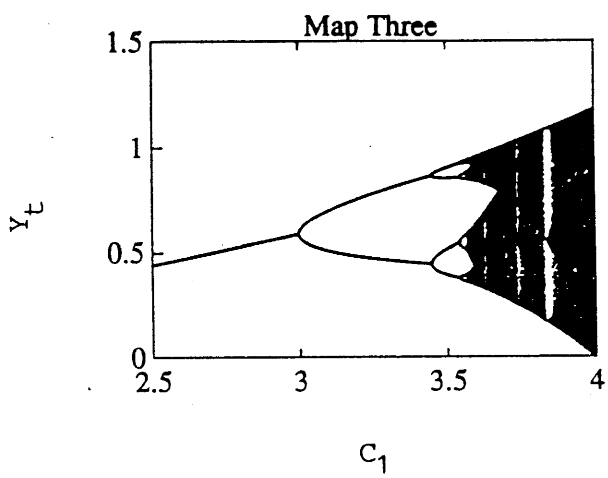


Figure 4: Bifurcation graph for
$y^{ij}_t = 4.3723206y^{ij}_{t-1} - 2.16755{y^{ij}_{t-1}}^2$

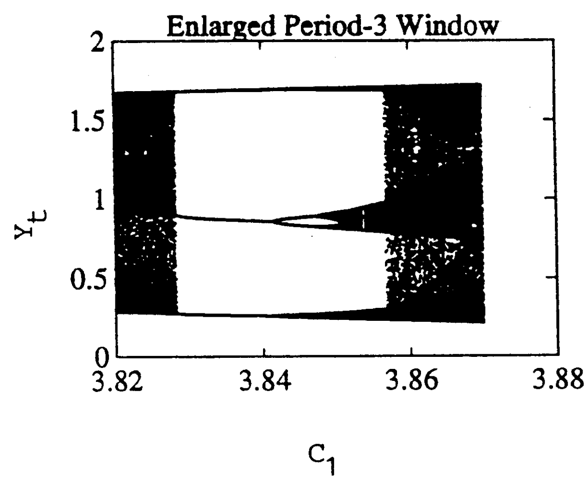


Figure 5: Bifurcation graph for period window from Figure 1.

Close scrutiny of the above figures as well as Figure 4 reveal several important features of the type of dynamics implied by the empirical analogue of equation II. First, the value of the parameter on the linear term--c_1 in equation II-- is very critical as it solely determines the time path of y^{ij}_t. Second, the model produces an "oscillatory equilibrium" rather than a complete convergence of the income series for many values of that lie in the neighborhood of its mean value estimate. Third, it is discovered that the parameter of the quadratic term in equation II is simply a scaling factor. Variations in the value of this parameter simply change the magnitude of y^{ij}_t but not its time path. Most importantly, it is discovered that Equation 1 shares the same generic properties as the widely used logistic equation. Specifically, the bifurcation points, the chaos bands, and odd-period windows occur at precisely at the same values of the growth parameter in the logistic equation.

Note that for a broad range of starting values for c_2 [from 2.5 to 3 approximately], the iterations of the quadratic equation converge to a single equilibrium (fixed point). However, as the value of c_2 is increased from 3 to 4, a cobweb type phenomenon begins to characterize the equilibrating process. In this range, lower order even cycles are followed by higher order even cycles, chaos bands odd period cycles (windows) and then again even period cycles. Note that unlike even period cycles which seem to follow other even period cycles, odd period cycles emerge without precedence and vanish unexpectedly. The appearance of low odd period cycles, as Sharkovsky (1964) and Li and Yorke (1975) have shown, are evidence that other cycles with other periods have occurred as well. Moreover, these odd period cycles are apparently "super attractive" as they tend to absorb countless number of other cycles. As a result, odd-period cycles surface as hollow "windows". This situation is particularly noticeable in the period-3 windows in Figures 1 through 3. To get a better sense of the amount of complexity present in a period-3 window, a segment of Figure I, starting at $c_1 = 3.82$ and ending at $c_1 = 3.87$ has been enlarged and shown in Figure 4. Amazingly, it is discovered that the period-3 band is not completely hollow as it appears, but immersed in period doubling effects with evidence of cycles of periods 2, 6, 8, 16, and so on as well as its own period-3 bands. Whereas these other cycles are not readily visible in the period-3 window, Figure 4 confirms their existence.

While there are many choices of c_2 for which the iterates of equation 6 produce predictable outcomes, chaotic regions (the dense parts in Figures 1-3) become the norm as the value of c_2 is varied from about 3.5 to 4.0. In this interval, excepting the odd-period windows, the state of y^{ij}_t becomes virtually unknown shortly after it has been perturbed. However, this type of completely unpredictable behavior should not be misinterpreted as *random motion*. This dichotomy between predictable chaotic behavior, typical in non-linear dynamical systems, and random behavior which characterizes stochastic processes(s) is an important matter (Day and Shafer, 1985, p. 289; Keen, 1989, p.57; Kesley, 1988, p. 21-26, Nathan.S. Balke, 1991). In looking back over our figures, it is apparent that chaotic behavior is not merely the outcome of some chance phenomena but embedded in the structure of the underlying system itself. By contrast, randomness results from the action and interaction of some unknown stochastic processes. Moreover, the error in prediction associated with a deterministic--albeit a chaotic system, is very different from that of a stochastic system. In a stochastic model, the error in prediction is a linear function of time, while the error in prediction in a chaotic regime grows exponentially with time. This consideration severely limits our ability to forecast economic time-series using the conventional standard statistical methods.

7 Conclusion

The possibility of change in parameter(s) of a dynamical system has important implications of whether the system is inherently capable of producing an "equilibrium" over time. The use of variable parameter(s) models in regional research is justified because regional economies are subject to both internal and external interventions and change in economic policy. Besdies it is known that intervention often lead to changes in the parameters of a dynamical system. When intervention affect parameter values or the system's initial conditions, so much so that it is pushed into its chaotic range, prediction becomes impossible.
The paradigm of non-linear dynamics continues to provide much needed insight into the analysis of economic time-series and other experimental dynamical systems. The discoveries made from examining the time evolution of these systems have provided scientists from diverse fields with a new and somewhat unified perspective on random versus nonrandom behavior. It is only a matter of time before these new approaches are successfully adapted and applied to finite samples with small number of data points.

Although there may be greater uncertainty with long-term forecasting of a time series when the underlying process is non-linear, there is little doubt that non-linear models provide a truer picture of reality than do the conventional fixed-parameter linear models.

REFERENCES

[1] Amos, Orley M."Unbalanced Regional Growth and Regional Income Inequality in the Latter Stages of Development." Regional Science and Urban Economics, 18, 1988, pp.549-566

[2] Amos, Orley M. "An Inquiry into the Causes of Increasing Regional Income Inequality in the United States." The Review of Regional Studies

[3] Amos, Orley M. "Substate and SMSA Personal Income Inequality and Regional Development." The Review of Regional Studies, vol. 16, Spring 1986, pp. 23- 30.

[4] Amos, Orley, M. "Divergence of Per Capita Real Gross State Product by Sector - 1963-1986," The Review of Regional Studies, 21, 1991, pp. 221-234.

[5] Amos, Orley, M, "Growth Pole Cycles,": A Synthesis of Growth Pole and Long Wave Theories," The Review of Regional Studies, 20, 1990, pp. 37-48.

[6] Baumol, William J., and E. N. Wolff, "Feedback from Productivity Growth in R and D, Scandinavian Journal of Economics, 85, 1983, pp. 145-157."

[7] Benhabib, J. and K. Nishimura, "Competitive Equilibrium Cycles," Journal of Economic Theory, 1985, 35, 284-306.

[8] Brock, William A., David A. Hsieh, and Blake LeBaron, NonLinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence, The MIT Press, Cambridge, Mass., 1991.

[9] Brock, William A. and C. L. Sayers, "Is the Business Cycle Characterized by Deterministic Chaos?" Journal of Monetary Economics, July 1988, pp. 71 -90.

[10] Brock William A."Causality, Chaos, Explanation and Prediction In Economics and Finance." in Casti, J., and Karlqvist, A., eds., *Beyond Belief: Randomness, Prediction, and Explanation in Science*, Boca Raton, FL: CRC Press, 1991, pp.230 - 279.

[11] Day, R. H. and Wayne Shafer, "Ergodic Fluctuations in Deterministic Economic Models," Journal of Economic Behavior and Organization, September 1987, 8 3, 339-61.

[12] Day, Richard H. and Wayne Shafer, 1985, "Keynesian Chaos," Journal of Macroeconomics, 7, 277-295.

[13] Dickey, D.A. and W.A. Fuller,"Distribution of the Estimators for Autoregressive Time Series with a Unit Root." Journal of the American Statistical Association, 1979, No:74, pp.427-431.

[14] Eberts, Randall W., "Accounting for the Recent Divergence in Regional Wage Differentials," Economic Review, Federal Reserve Bank of Cleveland, 1989 Quarter 3, Vol. 25, No. 3, 14-25.

[15] Eberts, Randall W. and Joe A. Stone, "Wage and Employment Response to Shocks in Local Labor Markets," monograph presented at the Western Economic Association Meetings, Lake Tahoe, NV, July 1989.

[16] Fuller, Wayne A., *Introduction to Statistical Time Series*, John Wiley & Sons, Inc., 1976.

[17] Hamilton, James, D. "A New Approach to the Economic Analysis of Time Series and the Business Cycle," Econometrica, March 1989, pp. 357 - 384.

[18] Keen, Linda, "Julia Sets," *Proceedings of Symposia in Applied Mathematics*, American Mathematical Society, Volume 39, 1989.

[19] Kesley, David,"The Economics of Chaos or the Chaos of Economics," Oxford Economic Papers, 40, 1988, pp. 1-31.

[20] Li, T. Y. and J. A. Yorke, "Period Three Implies Chaos," American Mathematical Monthly, 82, 1975, 985-92.

[21] Lorenz, H. W., "International Trade and the Possible Occurrence of Chaos," Economics Letters, 23, 1987a, No. 2, 135-138.

[22] Lorenz, H. W., "Strange Attractors in a Multisector Business Cycle Model," Journal of Economic Behavior and Organization, 1987b, 8, pp. 397-411.

[23] Kuznets, Simon, "Economic Growth and Income Inequality," American Economic Review, March 1955, pp. 1 - 28.

[24] Metzler, L., "The Nature and Stability of Inventory Cycles," The Review of Economic Studies, 1941.

[25] Modigliani, F., "Liquidity Preference and the Theory of Interest and Money," Econometrica, 12, January 1944, 45-88.

[26] Mosekilde, Erik, Erik Larsen, and John Sterman. " Coping With Complexity: Deterministic Chaos in Human Decisionmaking Behavior". in Casti, J., and Karlqvist, A., eds., *Beyond Belief: Randomness, Prediction, and Explanation in Science*, Boca Raton, FL: CRC Press, 1991, pp. 199-229.

[27] Ram, Arti,"Kuznets's Inverted-U Hypothesis: Evidence from a Highly Developed Country.", Southern Economic Journal, April 1991, pp. 1112- 1123.

[28] Ruelle, David, "Can Nonlinear Dynamic Help Economists", in Philip W. Anderson, Kenneth J. Arrow, and David Pines, eds., The Economy as an Evolving Complex System, SFI Studies in the Science of Complexity, Vol V, Addison-Wesley Publishing Compnay, 1988,Inc., pp. 195-203.

[29] Samuelson, P. A., "Foundations of Economic Analysis," Cambridge, MA, Harvard University Press, 1948.

[30] Sharkovsky, A., N., "Co-existence of Cycles of a Continuous Mapping of a Line into Itself," Ukranian Math. Z.., 16, 1964, pp. 61-71.

[31] Sahota, Gian S., "Theories of Personal Income Distribution: A Survey." Journal of Economic Literature, March 1978, pp. 1 - 55.

SYMMETRIC FUNCTION CALCULUS ON COMPUTER

Hiroki HASHIGUCHI and **Naoto NIKI**

Department of Industrial Management and Engineering, Science University of Tokyo

hiro@ms.kagu.sut.ac.jp and niki@ms.kagu.sut.ac.jp

Abstract

This paper introduces some algorithms on transition between bases for the ring of symmetric functions, and propose a symbolic multiplication algorithm for monomial symmetric functions which are the most fundamental basis. Those algorithms are implemented on computers in computer algebra languages, *Mathematica* and *Reduce*,

1. Introduction

Consider the ring $\Xi = \mathbf{Q}[x_1, x_2, \ldots, x_n]$ of polynomials in n independent variables $x_1, x_2, \ldots, x_n$. The symmetric group S_n acts on this ring by permuting the variables, and a polynomial is symmetric if it is invariant under this action. The family of symmetric polynomials forms a subring of Ξ and is a graded ring:

$$\Lambda_n = \bigoplus_{k \geq 0} \Lambda_n^k, \tag{1}$$

where Λ_n^k consists of the homogeneous symmetric polynomials of degree k, together with the zero polynomial.

In the theory of symmetric functions, the number n of variables is usually irrelevant, and it is often more convenient to consider that n is large enough or formally infinite. By this abstraction, we have the graded ring

$$\Lambda = \bigoplus_{k \geq 0} \Lambda^k \tag{2}$$

called the ring of symmetric functions in place of the phrase "symmetric polynomials".

A partition λ is any sequence $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_n)$ of nonnegative integers such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. The number of non-zero elements in λ is called the length and is denoted by $l(\lambda)$. The weight $|\lambda|$ of λ is defined as $|\lambda| = \sum_{i\geq 1} \lambda_i$; conversely, λ is called a partition of $|\lambda|$. The cardinality of the set $\Pi_k = \{\lambda \mid \lambda \vdash k\}$ is denote by $q(k)$, where $\lambda \vdash k$ means that λ is a partition of k. For the sake of convenience, a partition is expressed with the multiplicity π_i of a positive integer λ_i, likes as $\lambda = (\lambda_1^{\pi_i}, \lambda_i^{\pi_2}, \ldots, \lambda_s^{\pi_t})$. Here we note $l(\lambda) = \sum_{i=1}^{t} \pi_i$.

Each Λ^k is a $q(k)$-dimensional vector space spanned by several standard bases including monomial symmetric functions, elementary symmetric functions, complete symmetric functions, power sums and Schur functions. Any standard basis of degree k as a set has a one-to-one mapping to the set Π_k of partitions. The algebraic structure of the ring was discussed in Macdonald [MD].

It is one of the most fundamental problems to transform one basis to another, see Booker [BK], Bratley and McKay [B&M], Mckay [MK], Hashiguchi and Niki [H&N] and Remmel and Whitney [R&W]. In representation theory of symmetric group, the transformation of power sums into a linear combination of Schur functions is essential for calculating the characters of the group. In the distribution theory of multivariate statistics, Kendall and Stuart [K&S] have suggested the utility of transition of power sums into monomial symmetric functions in order to obtain the moments of moment distributions; and Nakagawa and Niki [N&N] have designed and coded a package of procedures for that purpose.

We begin with definitions and prelimi..aries in Section 2. In Section 3 and 4, we make a brief survey on the transition algorithms for symmetric functions from the view point of symbolic computation. Section 3 is devoted to symbolic multiplication of monomial symmetric functions, and Section 4 gives transition algorithms from one basis to another.

2. Bases of symmetric functions

2.1. Monomial symmetric functions

For any sequence $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in (\mathrm{N}_0)^n$ of nonnegative integers, we put $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_n$ and we let x^α denote the monomial $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$. For any partition $\lambda \vdash k$ of length $l(\lambda) \leq n$, let $m_n(\lambda)$ denote the following symmetric polynomial:

$$m_n(\lambda) = \sum x^\alpha \in \Lambda_n^k \tag{3}$$

where the sum is taken over all distinct permutations α of λ. If $l(\lambda) > n$ then $m_n(\lambda) \equiv 0$. We write $m(\lambda) \in \Lambda^k$, sometimes in the form $m(\lambda_1, \lambda_2, \ldots, \lambda_n)$, for sufficiently large or formally infinite n.

When a partition $\lambda \vdash k$ is written with multiplicities as $(\lambda_1^{\pi_1}, \lambda_2^{\pi_2}, \ldots, \lambda_t^{\pi_t})$, an augmented (monomial) symmetric function $a(\lambda)$ is defined by

$$a(\lambda) = \pi_1! \pi_2! \cdots \pi_t!\ m(\lambda) = \pi(\lambda)\ m(\lambda) \in \Lambda^k. \tag{4}$$

2.2. Elementary symmetric functions

For each $r \geq 0$ the rth elementary symmetric function e_r is the sum of all the products of r distinct variables, *i.e.*, $e_0 = 1$ and for $r \geq 1$

$$e_r = \sum_{i_1 < i_2 < \cdots < i_r} x_{i_1} x_{i_2} \cdots x_{i_r} = m(1^r). \tag{5}$$

For any partition $\lambda = (\lambda_1, \lambda_2, \ldots) \vdash k$, we define $e(\lambda)$ by

$$e(\lambda) = e_{\lambda_1} e_{\lambda_2} \cdots e_{\lambda_{l(\lambda)}} \in \Lambda^k. \tag{6}$$

2.3. Complete symmetric functions

We define the rth complete symmetric function h_r as the sum of all the monomials of total degree r,

$$h_r = \sum_{\lambda \vdash r} m(\lambda). \tag{7}$$

In particular $h_0 = 1$ and $h_1 = e_1$. For any partition $\lambda = (\lambda_1, \lambda_2, \ldots) \vdash k$, we define $h(\lambda)$ by

$$h(\lambda) = h_{\lambda_1} h_{\lambda_2} \cdots h_{\lambda_{l(\lambda)}} \in \Lambda^k. \tag{8}$$

2.4. Power sums

For each r the rth power sum is defined by

$$p_r = \sum x_i^r = m(r) \tag{9}$$

and for any $\lambda = (\lambda_1, \lambda_2, \ldots) \vdash k$ the power sum $p(\lambda)$ is defined by

$$p(\lambda) = p_{\lambda_1} p_{\lambda_2} \cdots p_{\lambda_{l(\lambda)}}. \tag{10}$$

2.5. Schur functions

For any partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n) \vdash k$ with $l(\lambda) \leq n$, we define Schur polynomial by

$$s_n(\lambda) = \frac{\det\,(x_i^{\lambda_j + n - j})}{\det\,(x_i^{n-j})} \quad (1 \leq i, j \leq n) \in \Lambda_n^k. \tag{11}$$

Since the set $\{m_n(\lambda) \mid \lambda \vdash k\}$ forms a $\mathbf{Q}$-basis for Λ_n^k, there exist integers $K_{\lambda\mu}$ such that

$$s_n(\lambda) = \sum_{\mu \vdash k} K_{\lambda\mu}\, m_n(\mu) \in \Lambda_n^k, \tag{12}$$

where each $K_{\lambda\mu}$ is called a *Kostaka number* which is the number of *Young tableaux* of shape λ and of weight μ. Hashiguchi and Niki [H&N] proposed a series of algorithms to obtain $\{K_{\lambda\mu} \mid \mu \vdash |\lambda|\}$ for given λ. Using Kostaka numbers for given $\lambda \vdash k$, we define Schur function $s(\lambda) \in \Lambda^k$ by

$$s(\lambda) = \sum_{\mu\vdash k} K_{\lambda\mu}\, m(\mu). \tag{13}$$

For any sequence, which is not necessary to be a partition, definition of Schur polynomial similar to (11), or that of function in the same manner as (13), is possible. The following proposition always lead to the standard form

Proposition 1 For nonnegative integers $p \geq 0$ and $q > 0$,

$$s(\ldots, p, q, \ldots) = -s(\ldots, q-1, p+1, \ldots), \tag{14}$$
$$s(\ldots, p, p+1, \ldots) = 0. \tag{15}$$

2.6. Algebraic structure

Proposition 2

1. All the elements of $\{e_r\}$ are mutually algebraically independent over $\mathbf{Q}$ and so are those of $\{h_r\}$ and $\{p_r\}$, and that

$$\Lambda = \mathbf{Q}[e_1, e_2, \ldots] = \mathbf{Q}[h_1, h_2, \ldots] = \mathbf{Q}[p_1, p_2, \ldots]. \tag{16}$$

2. As λ runs through all partitions of k, the sets $\{m(\lambda) \mid \lambda \vdash k\}$, $\{e(\lambda) \mid \lambda \vdash k\}$, $\{h(\lambda) \mid \lambda \vdash k\}$, $\{s(\lambda) \mid \lambda \vdash k\}$, $\{p(\lambda) \mid \lambda \vdash k\}$ and $\{a(\lambda) \mid \lambda \vdash k\}$ are $\mathbf{Q}$-bases for Λ^k.

3. Multiplication of elements in basis

Let the symbol '$||$' denote a binary operator of concatenation which means that two sequences are combined. In addition, let 0_k denote the sequence of zeros of length k. The symbolic product concerning $e(\lambda)\cdot e(\mu)$, $h(\lambda)\cdot h(\mu)$ and $p(\lambda)\cdot p(\mu)$ is done merely by sorting elements in $\lambda\,||\,\mu$. That of Schur functions discussed in Remmel and Whitney [R&W]. Here, we propose algorithms for the symbolic multiplication of $m(\lambda)\cdot m(\nu)$ through that of $a(\lambda)\cdot a(\nu)$.

3.1. Multiplication of monomial symmetric functions

For any index sequence $T = (t_1, \ldots, t_k)$, we let $\mu_T = (\mu_{t_1}, \ldots, \mu_{t_k})$ and $\mu_0 \equiv 0$.

Algorithm 1 (Multiplication of augment symmetric functions)

Input: $\lambda = (\lambda_1, \ldots \lambda_l), \mu = (\mu_1, \ldots \mu_m)$, $l = l(\lambda) \geq m = l(\mu)$

Output: $r = \sum_{\nu} c[\lambda, \mu]_\nu\, a(\nu)$

1. $r \leftarrow 0$, $i \leftarrow 0$.

2. $L \leftarrow \lambda\,||\,0_{m-i}$

3. Let M be the ordered set $\{1, 2, \ldots, m\}$. $c \leftarrow {}_m\mathrm{C}_i$. We take all the subsets of M of size i and write them $I_1, I_2, \ldots, I_c$. $j \leftarrow 1$.

4. $K \leftarrow 0_l \,||\, (M \backslash I_j)$ where all the elements of the set in parenthesis are in ascending order.

5. $p \leftarrow {}_l\mathrm{P}_i$. We make all the repeated permutations of $0_{l-i} \,||\, (I_j)$ and write them $J_1, J_2, \ldots, J_p$. $k \leftarrow 1$.

6. $J \leftarrow J_k \,||\, 0_{m-i}$. $N \leftarrow \mu_J + \mu_K + L$.

7. We sort the elements of N in descending order: $\nu \leftarrow \mathrm{sort}(N)$.
   $r \leftarrow r + a(\nu)$.

8. $k \leftarrow k + 1$. If $k \leq p$ then go back to 6.
   $j \leftarrow j + 1$. If $j \leq c$ then go back to 4.
   $i \leftarrow i + 1$. If $i \leq m$ then go back to 2.
   Otherwise return r.

Before we show that Algorithm 1 gives the expanded expression of the product of $a(\lambda) \cdot a(\mu)$, namely $r = a(\lambda) \cdot a(\nu)$, we prepare some notations. It is straightforward that the product of $a(\lambda)$ and $a(\mu)$ is of the following form:

$$a(\lambda) \cdot a(\mu) = \sum c_\nu^{(\lambda,\mu)}\, a(\nu), \quad c_\nu^{(\lambda,\mu)} \geq 0. \tag{17}$$

We use the symbol Γ for denoting

$$\Gamma^{(\lambda\,\mu)} = \{\nu \,|\, c_\nu^{(\lambda,\mu)} > 0\}. \tag{18}$$

On the other hand, we denote the result of Algorithm 1 as follows.

$$r = \sum c[\lambda,\, \mu]_\nu\, a(\nu), \quad \Gamma[\lambda,\, \mu] = \{\nu \,|\, c[\lambda,\, \mu]_\nu > 0\}. \tag{19}$$

Lemma 1 *For two partitions λ and μ, we have*

$$\Gamma[\lambda,\, \mu] \subset \Gamma^{(\lambda,\mu)}, \quad c[\lambda,\, \mu]_\nu \leq c_\nu^{(\lambda,\mu)}. \tag{20}$$

Proof. When we see the step 6 of Algorithm 1, we find that the $a(\lambda)$ contains the monomial $x^\lambda = x^L$ and that the $a(\mu)$ the monomial $x^{\nu_J + \nu_K}$. Therefore the $a(\lambda) \cdot a(\mu)$ has the $a(\nu)$ of the power ν sorted $N = \mu_J + \mu_K + L$ in descending order. Hence we can obtain

$$\Gamma[\lambda,\, \mu] \subset \Gamma^{(\lambda,\mu)}, \quad c[\lambda,\, \mu]_\nu \leq c_\nu^{(\lambda,\mu)}. \tag{21}$$

Lemma 2 *For positive integers $n, m, l \in \mathbf{N}$ such that $n \geq l + m$ and $l \geq m$, it holds that*

$$_n\mathrm{P}_m = \sum_{i=0}^{m} {}_m\mathrm{C}_i \cdot {}_m\mathrm{P}_i \cdot {}_{n-l}\mathrm{P}_{m-i}. \tag{22}$$

Proof. If $n \geq l + m$ and $l \geq m$, then the following identity holds clearly because of a combinatorial interpretation.

$$ {}_n\mathrm{C}_m = \sum_{i=0}^{m} {}_l\mathrm{C}_i \cdot {}_{n-l}\mathrm{C}_{m-i} \,. \tag{23} $$

The following expansion of (23) lead us to prove Lemma 2.

$$ \begin{aligned} \frac{n!}{(n-m)!\,m!} &= \sum_{i=0}^{m} \frac{l!}{(l-i)!\,i!} \cdot \frac{(n-l)!}{(m-i)!\,(n-l-m+i)!}, \\ \frac{n!}{(n-m)!} &= \sum_{i=0}^{m} \frac{l!}{(l-i)!} \cdot \frac{m!}{(m-i)!\,i!} \cdot \frac{(n-l)!}{(n-l-m+i)!}, \\ {}_n\mathrm{P}_m &= \sum_{i=0}^{m} {}_l\mathrm{P}_i \cdot {}_m\mathrm{C}_i \cdot {}_{n-l}\mathrm{P}_{m-i} \,. \end{aligned} \tag{24} $$

Theorem 1 *For partitions λ and μ, Algorithm 1 gives the expansion of the $a(\lambda) \cdot a(\mu)$ as the weighted sum of augmented symmetric functions, i.e.,*

$$ a(\lambda) \cdot a(\mu) = \sum_{\nu} c[\lambda,\, \mu]_\nu \; a(\mu). \tag{25} $$

where ν, $c[\lambda,\, \mu]_\nu$ are the results of Algorithm 1.

Proof. By making use of the homomorphisms in Macdonald [MD], we have only to prove Theorem 1 for sufficiently large number n of independent variables.

From Lemma 1, we see that the result r of Algorithm 1 is a part of $a(\lambda) \cdot a(\mu)$. Therefore we show that the number of monomials in $a(\lambda) \cdot a(\mu)$ is equal to that of monomials in r. The former number T_1 is

$$ T_1 = {}_n\mathrm{P}_l \cdot {}_n\mathrm{P}_m \,. \tag{26} $$

On the other hand, the latter T_2 is

$$ T_2 = \sum_{i=1}^{m} {}_m\mathrm{C}_i \cdot {}_l\mathrm{P}_i \cdot {}_n\mathrm{P}_{l+m-i} \,. \tag{27} $$

If we multiply the both sides of (22) by ${}_n\mathrm{P}_l$, we get

$$ {}_n\mathrm{P}_l \cdot {}_n\mathrm{P}_m = \sum_{i=0}^{m} {}_m\mathrm{C}_i \cdot {}_l\mathrm{P}_i \cdot {}_n\mathrm{P}_{l+m-i} \,, \tag{28} $$

which implies that T_1 is equal to T_2.

From Algorithm 1 and Theorem 1, It is easy to derive the algorithm for symbolic multiplication of $m(\lambda) \cdot m(\nu)$, namely,

$$ m(\lambda) \cdot m(\mu) = \sum_{\nu} \frac{c[\lambda, \mu]\pi(\nu)}{\pi(\lambda)\pi(\mu)} m(\nu) \tag{29} $$

If we apply Algorithm 1 to $(2,1^2)$ and $(3,2,1)$, then we have

$$\begin{aligned} & a(2,1^2) \times a(3,2,1) \\ = \; & 2a(5,3,2) + 2a(4,2^2) + 2a(4,3^3) + 6a(4,3,2,1) + 2a(4^2,1^2) \\ & +2a(5,3,1^2) + 2a(3^3,1) + 2a(5,2^2,1) + 2a(3^2,2^2) \\ & +2a(4,2^3) + 3a(3^2,2,1^2) + 2a(4,2^2,1^2) + 2a(3,2^3,1) \\ & +a(5,2,1^3) + a(4,3,1^3) + a(3,2^2,1^3), \end{aligned} \tag{30}$$

which yields, from (29),

$$\begin{aligned} & m(2,1^2) \times m(3,2,1) \\ = \; & m(5,3,2) + 2m(4,2^2) + 2m(4,3^2) + 3m(4,3,2,1) + 4m(4^2,1^2) \\ & +2m(5,3,1^2) + 6m(3^3,1) + 2m(5,2^2,1) + 4m(3^2,2^2) \\ & +6m(4,2^3) + 6m(3^2,2,1^2) + 4m(4,2^2,1^2) + 6m(3,2^3,1) \\ & +3m(5,2,1^3) + 3m(4,3,1^3) + 6m(3,2^2,1^3). \end{aligned} \tag{31}$$

4. Transition algorithm between bases

4.1. Relations between e, h and p

The following relations concerning $\{e_r\}, \{h_r\}$ and $\{p_r\}$ are well-known:

Proposition 3 For any positive integer m,

$$\sum_{r=1}^{m}(-1)^{r-1}p_r e_{m-r} - m e_m = 0, \tag{32}$$

$$\sum_{r=1}^{m} p_r h_{m-r} - m h_m = 0, \tag{33}$$

$$\sum_{r=0}^{m}(-1)^r e_r h_{m-r} = 0. \tag{34}$$

Proposition 3 give us a hint for transition between $\{e(\lambda)\}$, $\{h(\lambda)\}$ and $\{p(\lambda)\}$. Nakagawa [ND] used the relation (32) and made algorithms for $\{e(\lambda)\} \leftrightarrow \{p(\lambda)\}$. From (32) and (34), we can find it easy to make algorithms for $\{h(\lambda)\} \leftrightarrow \{p(\lambda)\}$ and $\{e(\lambda)\} \leftrightarrow \{h(\lambda)\}$ in the similar way.

Let two symbols $\{\phi_r\}_{n\geq 1}$ and $\{\psi_r\}_{r\geq 1}$ satisfy the following relation including those of (32), (33) and (34).

$$\phi_m = \begin{cases} \psi_1 & m = 1, \\ \displaystyle\sum_{r=1}^{m-1} c(m,r)\,\psi_r\,\phi_{m-r} + d(m)\,\psi_m & m \geq 2, \end{cases} \tag{35}$$

where all the coefficients $d(m), c(m,1), \ldots, c(m,m-1)$ are given. We show an algorithm from ϕ to the polynomial of ψ's

Algorithm 2 Procedure ϕ_to_ψ;

Input: m from ϕ_m

Output: polynomial of ψ's

1. If $m = 1$ then return ψ_1, else $result \leftarrow d(m)\psi_m$.
2. For r from 1 to $m-1$ step 1 do

$$result \leftarrow c(m,r) \cdot \phi\_to\_\psi(m-r) \cdot \psi_r.$$

3. return $result$.

4.2. Transition from power sum to Schur functions

We explain an outline of an algorithm from $p(\lambda)$ to a linear combination of $\{s(\lambda)\}$. This transition based on the propositions below is well-known in the representation theory of symmetric group.

Proposition 4 For any partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)$ with $l = l(\lambda)$ and for any integer $r \geq 1$, it holds that

$$\begin{aligned} p_r \cdot s(\lambda) = {} & s(\lambda_1 + r, \lambda_2, \ldots, \lambda_l) + s(\lambda_1, \lambda_2 + r, \ldots, \lambda_l) + \cdots + \\ & s(\lambda_1, \lambda_2, \ldots, \lambda_l + r) + s(\lambda_1, \lambda_2, \ldots, \lambda_l, r) + \\ & s(\lambda_1, \lambda_2, \ldots, \lambda_l, 0, r) + \cdots + s(\lambda_1, \lambda_2, \ldots, \lambda_l, \underbrace{0, \cdots, 0}_{r-1}, r), \end{aligned} \tag{36}$$

$$p_r = s(r) - s(r-1,1) + s(r-2,1^2) - \cdots (-1)^{r-1} s(1^r). \tag{37}$$

We implemented several algorithms on computer according to the Propositions 1 and 4. By using those we have, for example,

$$\begin{aligned} p(4,2^2) &= p_4 \cdot p_2 \cdot \{s(2) - s(1,1)\} \\ &= p_4 \cdot \{s(4) - s(2,2) + s(2,0,2)\} - \\ &\qquad p_4 \cdot \{s(3,1) + s(1,3) + s(1,1,2) + s(1,1,0,2)\} \end{aligned} \tag{38}$$

$$\cdots$$

$$\begin{aligned} &= s(8) + 2s(4^2) - 2s(5,3) + 2s(6,3) - s(7,1) \\ &\qquad -s(6,1^2) - 2s(2^4) + s(5,1^3) + 2s(2^3,1^2) \\ &\qquad\quad -s(4,1^4) - 2s(2^2,1^4) + s(3,1^5) + s(2,1^6) - s(1^8) \end{aligned} \tag{39}$$

Note that each Schur function with a sequence rather than a partition has been already changed into its standard form in (39) including
$s(2,0,2) = -s(2,1,1)$, $s(1,3) = -s(2,2)$.

4.3. Power sum and monomial symmetric function

Niki [NN] designed algorithms for $\{p_\lambda\} \leftrightarrow \{a_\lambda\}$ according to the following proposition.

Proposition 5 For any partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)$ with $l = l(\lambda)$ and any positive integer r, it holds that

$$p_r \cdot a(\lambda) = a(r, \lambda_1, \lambda_2, \ldots, \lambda_l) + a(r + \lambda_1, \lambda_2, \ldots, \lambda_l) + \cdots + a(\lambda_1, \lambda_2, \ldots, r + \lambda_l), \tag{40}$$

$$a(\lambda_1, \lambda_2, \ldots, \lambda_l) = p(\lambda_1)\, a(\lambda_2, \ldots, \lambda_l) - a(\lambda_1 + \lambda_2, \ldots, \lambda_l) - \cdots - a(\lambda_2, \ldots, \lambda_2 + \lambda_l). \tag{41}$$

where the augmented functions on the right hand side may have augments which should be arranged in descending order.

For examples,

$$\begin{aligned}
p(4, 2^2) &= p_4 \cdot p_2 \cdot p_2 \\
&= p_4 \cdot p_2 \cdot a(2) \\
&= p_4 \cdot \{a(4) + a(2^2)\} \\
&= a(8) + a(4^4) + 2a(6, 2) + a(4, 2^2), \qquad (42)\\
a(4, 2^2) &= p(4) \cdot a(2^2) - 2a(6, 2) \\
&= p(4) \cdot \{p(2) \cdot a(2) - a(4)\} - 2\{p(6) - a(8)\} \\
&= 2\,p(8) - p(4^2) - 2p(6, 2) + p(4, 2^2), \qquad (43)
\end{aligned}$$

which yield, from (4),

$$p(4, 2^2) = m(8) + 2m(4^2) + 2m(6, 2) + 2m(4, 2^2), \tag{44}$$

$$m(4, 2^2) = p(8) - \frac{1}{2}p(4^2) - p(6, 2) + \frac{1}{2}p(4, 2^2). \tag{45}$$

4.4. Fundamental theorem in symmetric functions

The algorithm to express any symmetric polynomial in term of e_i's is well-known, see [ND], which can be easily extended to that of symmetric functions.

Algorithm 3

Input: $f = \sum c_\lambda\, m(\lambda)$ in Λ

Output: weighted sum of e's.

1. *result* $\leftarrow$ the constant term of f. $f \leftarrow f -$ *result*.

2. If $f = 0$ then return *result*.

3. Choose $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that $LPP(f) = LPP(e_1^{\alpha_1} e_2^{\alpha_2} \cdots e_n^{\alpha_n})$.

4. Expand $\prod_{i=1}^{n} m(1^i)^{\alpha_i}$ by repeatedly applying Algorithm 1 and the identity (29) into a weight sum of m's and assigin it to V.

   $result \leftarrow result + LC(f) \cdot V$.

   $f \leftarrow f - result$.

   go back to step 2.

Here $LPP(f)$ and $LC(f)$ denote the maximal partition with respect to a lexical order and the coefficient of the $LPP(f)$, respectively. For example, we apply Algorithm 3 to $f = m(3,1) - 3m(2,1^2)$ and get

$$16e(4) - 2e(2^2) - 4e(3,1) + e(2,1^2). \tag{46}$$

Similarly, we can also transform $m(\lambda)$ into a linear combination of $\{h(\lambda)\}$ or $\{p(\lambda)\}$.

References

[BK] A. Booker. Numerical evaluation of symmetric polynomials. *J.ACM*, Vol. **12**, No. **1**, pp. 90–94, 1965.

[B&M] P. Bratley and J. K. S Mckay. Algorithm 305 symmetric polynomials. *Commun. ACM*, Vol. **10**, No. **7**, p. 450, 1967.

[H&N] H. Hashiguchi and N. Niki. Algorithm for enumeration of a class of Young tableaux. *Sūshiki-shori*, Vol. **2**, No. **2**, pp. 28–32, 1993. (in Japanese).

[MD] I. G. Macdonald. *Symmetric Functions and Hall polynomials.* Oxford University Press, Oxford, 1979.

[MK] J. K. S Mckay. On the representations of symmetric polynomials. *Commun. ACM*, Vol. **10**, No. **7**, pp. 428–429, 1967.

[ND] S. Nakagawa. *Symmetric Polynomial Calculus in Distribution Theory of Multivariate Statistics.* PhD thesis, Kyushu Univ., 1994.

[N&N] S. Nakagawa and N. Niki. Change of bases of symmetric polynomials of several systems of quantities and their applications to distribution theory of multivariate statistics. In *Theory and Applications in Computational Statistics.* Scientist Inc., Tokyo, 1991.

[NN] N. Niki. Algorithms on the representation of multisystem symmetric polynomials and their applications in statistics. *Proc. 11th Int. Symp. Computer at the University.*, Vol. **5.1**, pp. 1–6, 1989.

[R&W] J. B. Remmel and R. Whitney. Multiplying Schur functions. *J. Algorithms*, Vol. **5**, pp. 471–487, 1982.

[SG] B. E. Sagan. *The Symmetric Group.* Brooks/Cole, Pacific Grove, CA, 1991.

[SK] S. Skiena. *Implementing Discrete Mathematics.* Addison-Wesly, 1990.

[ST] R. P. Stanley. Theory and applications of plane partitions I, II. *Studies appl. Math.*, Vol. **50**, pp. 167–188, 259–279, 1971.

[K&S] M. G. Kendall and A. Stuart. *Advanced theory of statistics*, Vol. 1. Charles Griffin, 4 edition, 1977.

A MAPLE PROGRAM FOR COMPUTING THE KRONECKER CANONICAL FORM OF A SPECIAL MATRIX PENCIL †

Yu-zhen Huang and Wen-da Wu
Mathematics Mechanization Research Center
Institute of System Sciences, *Academia Sinica*,100080
&
Beijing Municipal Computing Center
Beijing Dongsi Nandajie 249,100005

Abstract *A Maple program for computing the Kronecker canonical form of a special matrix pencil, arising from solving polynomial system, is presented.*

§1 Introduction

For solving polynomial system a new approach is to transform the problem to matrix eigenproblem.[St] In the 0-dimensional case, ordinary matrix eigenvalue problem is obtained. In the positive dimensional case usually general eigenvalue problem with special matrix pencil is resulted. The regular part of the Kronecker canonical form of the pencil,when it exists, is closely related to those subvarieties of the system, whose projections on the corresponding axes are isolated points. When it does not exist, the form can be used to transform the system to another, which is quadratic.

For simplicity we consider the polynomial system over rational field, and the system itself is Groebner basis in total degree ordering and lexicographic.

The general eigenproblem resulted is of the form

$$(\hat{A} - x_\lambda \hat{B})Z = 0 \tag{1.1}$$

where the components of the vector Z are monomials in the unknowns of the system arranged in ascending order, x_λ is one of the unknowns chosen to form the eigenproblem, matrix $\hat{A} \in Q^{m\times n}$, and

$$\hat{B} := [I_m \;\; O]$$

where I_m is the identity matrix of order m and O is $m \times (n-m)$ zero matrix with $n > m$.

The program is written in MAPLE V, and can be modified to compute the Kronecker canonical form of any matrix pencil over rational field.

†The work reported herein has been supported in part by the National Climbing Project "Method for Computer Proving and its Applications" and in part by Beijing Municipal Natural Science Foundation

The notations, computing steps[HW] and examples are given in the following sections.

§2 Notation

(1.1) is equivalent to

$$\hat{K}\hat{Q}^{-1}Z = 0 \tag{2.1}$$

where

$$\hat{K} := \hat{P}(\hat{A} - x_\lambda \hat{B})\hat{Q}$$

is the Kronecker canonical form, $\hat{P}$ and $\hat{Q}$ are nonsingular.

Case 1: regular part exists
the output will be a list containing
(1) A: regular part matrix
(2) p: the square free form of the characteristic polynomial of A
(3) wz: the corresponding vector of A, i.e. subvector of $\hat{Q}^{-1}Z$
(4) $vlist$: the corresponding vector of the irregular part

Case 2: regular part does not exist
the output will be a list containing
(1) $vlist$: the corresponding vector of the irregular part
(2) eq: expressions of u_j in x_λ, Y_i, where $j = 1, \cdots, n, i = 1, \cdots, r$, r is the number of irregular parts in $\hat{K}$, Y_i are polynomials in unknowns of the original system and are considered as the unknowns of the new system.

vlist is in the form

$$[[v_1], [v_2], \cdots, [v_s], [v_{s+1}, v_{s+2}], \cdots, [v_{s+k}, \cdots, v_p]]$$

(1) each v_i is a linear combination of $u_j (j = 1 \cdots n)$ which correspond to the power products in Z
(2) each element of *vlist* is a list, the number of its operands corresponds to the column dimension of the irregular part

§3 Computing Steps

Step 1. [preparation]

(1) define $m \times 1$ matrix Z_1 with components u_i
(2) separate $\hat{A}$ into A and B, A is the first m columns of $\hat{A}$, B is the last $n - m$ columns

(3) wz_1 is the matrix composed by last $n-m$ elements of Z_1
wz is $Z_1 \backslash wz_1$
(4) initialize $vlist; t=1; r_1=n-m;$

Step 2. [column elemination of B and modification of $wz_1, \cdots, wz_t$]

(1) initialize *collist* and *rowlist*
(2) for i from 1 to column dimension of B do
for j from 1 to row dimension of B do
if $B[i,j] \neq 0$ then
$\{rowlist \longleftarrow i;$
$collist \longleftarrow j;$
$wz_k = Q_j^{-1} wz_k (k=1,\cdots,t);$
where Q_j^{-1} is got by replacing the j-th column of unit matrix $I_{coldim\ of B}$ by i-th row of B
$B = BQ_j;$
where Q_j is got by modifying the j-th row of unit matrix $I_{coldim\ of B}$ into
$(-dB[i,1],\cdots,-dB[j,j-i],d,-dB[i,j+1],...,-dB[i,coldim\ of B])$
where $d=B[i,j]^{-1}$; exit }
(3) r_2=rank of B;
(4) form *zerolist*, the members of *zerolist* is the column number of the zero columns of B

Step 3 [form the corresponding vector of the irregular part]

if *zerolist* is not empty then
{ $list_j = [wz_t[l_j], wz_{t-1}[l_j], \cdots, wz_1[l_j]]$,
where l_j is the j-th element in *zerolist*
put the $list_j$ into *vlist*;
$wz_s = wz_s \backslash wz_s[l_j]$,
where $s=1,\cdots,t, j=1,\cdots,$ the number of elements in *zerolist* }
else goto step 4

Step 4 [exits of the program]

(1) if r_2=0 then { return $A, p, wz, vlist$; STOP }
(2) if r_2=column dimension of A then
{B=submatrix(A, 1..r_2,collist); (submatrix is a maple function)
$A = B^T AB;$
$wz_{t+1} = B^T wz;$
$wz_s = wz_s + Awz_{s+1} (s=t,\cdots,1)$
$list_j = [wz_{t+1}[j], wz_t[j], \cdots, wz_1[j]]; \quad j=1,\cdots,r_2$

put $list_j$ into $vlist$}

(3) Return information "There is no regular part!!!"
Return $vlist, eq$;
STOP;

(4) if $r_2 \neq 0$ and $r_2 \neq$ column dimension of A then goto step 5

Step 5 [recursive preparation and formation of A, B]

(1) $crowlist = [1, 2, \cdots, m] \backslash rowlist$

(2) $M = submatrix(B, crowlist, collist)$;
$N = submatrix(A, rowlist, crowlist)$;
$K = submatrix(A, rowlist, rowlist)$;
$B = submatrix(A, crowlist, rowlist)$;
$A = submatrix(A, crowlist, crowlist)$;
$A = A - MN$;
$B = B - MK + AM$;
$K = K + NM$;

Step 6 [recursive preparation and vector modification]

(1) wz_{t+1}=submatrix(wz,rowlist,1..1);
$wz = wz \backslash wz_{t+1}$;
$wz = wz - Mwz_{t+1}$;
$Y = wz$;
$wz_t = wz_t + Kwz_{t+1} + NY$;

(2) for j from $t-1$ to 1 do
$\{Y = AY + Bwz_{j+2}$;
$wz_j = wz_j + Kwz_{j+1} + NY; \}$

(3) $t = t + 1; m = m - r_2; r_1 = r_2$;

(4) goto Step 2

§4 Examples

Example 1. Polynomial system

$$\begin{cases} f_1 = y^2 + x + 2z - 2 \\ f_2 = xy - 2xz + yz - 2z^2 - y + 2z \\ f_3 = x^2 - 2xy + xz - 2yz - x + 2y \end{cases}$$

The Groebner basis of the system is

$$\begin{cases} g_1 = y^2 + x + 2z - 2 \\ g_2 = xy - 2xz + yz - 2z^2 - y + 2z \\ g_3 = x^2 - 3xz - 4z^2 + 4z - x \\ g_4 = 2xz^2 + 2z^3 + 3xz + z^2 - x - 4z + 1 \end{cases}$$

in total degree ordering with $x \succ y \succ z$

$$\hat{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ -\frac{1}{2} & 2 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & -\frac{3}{2} & -1 & 0 \end{bmatrix}$$

$$Z = [\, 1 \;\; z \;\; y \;\; x \;\; z^2 \;\; zy \;\; zx \;\; z^3 \;\; z^2y \,]^T$$

$$x_\lambda = z$$

The program will return a list $[A, p, wz, vlist]$

$$A = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & -\frac{3}{2} \end{bmatrix}$$

$$p = \lambda^2 + \frac{3}{2}\lambda - \frac{1}{2}$$

$$wz = \begin{bmatrix} u_4 + u_2 - 1 \\ u_7 + u_5 - u_2 \end{bmatrix}.$$

$$vlist = [[u_3, u_6, u_9], [u_1, u_2, u_5, u_8]]$$

From the eigenvalues of the A we get isolated solutions $z = \frac{1}{4}(-3 \pm \sqrt{17}), y = 2z, x = 2z - 1$ and from wz we get solution's manifold $x + z - 1 = 0$.

Example 2. Polynomial system

$$\begin{cases} f_1 = x^2 + xz + 3y - 1 \\ f_2 = yz + x + y + 2 \end{cases}$$

Groebner basis is $g_1 = f_1, g_2 = f_2$ in total degree ordering with $x \succ y \succ z$;

$$\hat{A} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & -3 & 0 & -1 & 0 \end{bmatrix}$$

$$Z = [\ 1 \ \ z \ \ y \ \ x \ \ xz \ \ xy\ ]^T$$

$$x_\lambda = x$$

The output is the following:

There is no regular part!!!

$$vlist = [[\ u_1, u_4, -u_5 - 3u_3 + u_1\ ], [\ u_2 + u_4, u_1 - 3u_3, u_4 - 3u_6\ ]]$$

$$eq = \{u_4 = \lambda Y_1, u_5 = -\lambda^2 Y_1 + \lambda Y_2, u_6 = \tfrac{1}{3}\lambda Y_1 - \tfrac{1}{3}\lambda^2 Y_2,$$

$$u_3 = \tfrac{1}{3}Y_1 - \tfrac{1}{3}\lambda Y_2, u_1 = Y_1, u_2 = -\lambda Y_1 + Y_2\}$$

Because of $Y_1 = 1$ then we can get

$$\begin{cases} u_1 = 1, u_2 = -\lambda + Y_2, u_3 = \frac{1}{3} - \frac{1}{3}\lambda Y_2, \\ u_4 = \lambda, u_5 = -\lambda^2 + \lambda Y_2, \\ u_6 = \frac{1}{3}\lambda - \frac{1}{3}\lambda^2 Y_2 \end{cases}$$

In terms of Y_i the Groebner basis becomes

$$g_1' \equiv 0$$
$$g_2' = -\tfrac{1}{3}(\lambda Y_2^2 - (1 + \lambda^2 - \lambda)Y_2 - 2\lambda - 7)$$

then we get solution manifold:

$$\begin{cases} x = \lambda \\ y = \frac{1}{6}(1 - \lambda^2 + \lambda \pm \sqrt{\lambda^4 - 2\lambda^3 + 11\lambda^2 + 26\lambda + 1}\ ) \\ z = \frac{1}{2\lambda}(1 - \lambda^2 - \lambda \mp \sqrt{\lambda^4 - 2\lambda^3 + 11\lambda^2 + 26\lambda + 1}\ ) \end{cases}$$

Reference

[St] H.J.Stetter, Multivariate Polynomial Equations as Matrix Eigenproblems, WSSIAA 2 (1993) 355-371.

[HW] Huang Yuzhen and Wu Wenda, On the Computation of the Kronecker form of a Special Matrix Pencil, MM Research Preprints, No.7, 1992, 31-35

COMPUTATION OF EXPONENT POLYNOMIAL MATRIX A(X) IN THE FORM OF PUISEUX SERIES

Kentaro Katamachi

Institute of Mathematics, University of Tsukuba
Tsukuba-shi 305, Japan
katamati@math.tsukuba.ac.jp

Abstract

It is well-known that e^{At}, with A a constant matrix, can be computed by $\mathcal{L}^{-1}\{(sI-A)^{-1}\}$ where $\mathcal{L}^{-1}$ is Laplace inverse transform. When A has polynomial entries, exact computation with conventional algorithms are impractical. Hence, we consider efficient computation of this exponent in the form of truncated Puiseux series (fractional power series).

The old method is, although approximate, quite practical and seems to be useful for many applications.

For example, let A be 2×2 matrix which has polynomial entries, then the characteristic polynomial $\det(sI-A)$ is expressed as $s^2+a_1(x)s+a_0(x)$. we compute roots of $s^2+a_1(x)s+a_0(x)=0$ using Puiseux series expansion with Sasaki–Kako's method [SK93] , which are eigenvalues λ_1, λ_2. e^{At} can be expressed as $e^{At} = \frac{A-\lambda_2}{\lambda_1-\lambda_2}e^{\lambda_1 t} + \frac{A-\lambda_1}{\lambda_2-\lambda_1}e^{\lambda_2 t}$ provided that $\lambda_1 \neq \lambda_2$.

1 Introduction

The purpose of this paper is to express $e^{A(x)t}$ approximately, where $A(x)$ has polynomial entries, using Puiseux series expansion of eigenvalues of $A(x)$. This corresponds to solving linear ordinary differential equation with one time-independent parameter. When A is a constant matrix, the way of calculating e^{At} is well-known. However when A has polynomial entries, conventional algorithms is impractical. Our algorithm is to calculate $e^{A(x)t}$ approximately using the roots of characteristic polynomial, i.e., eigenvalues in the form of truncated Puiseux series (fractional power series).

Newton polygon method has been a standard method to perform Puiseux expansions for a long time [Pui50]. However, the method is very inefficient and conventionally has been performed by introducing algebraic numbers successively, which makes the method even more inefficient. Although the method proposed by D.Duval keeps away from introducing algebraic numbers if possible [Duv89], its computation is also quite heavy due to exact computation.

On the other hand, Sasaki & Kako proposed the method which uses floating-point arithmetic for the numerical coefficients, without introducing algebraic numbers [SK93]. This algorithm is more efficient than conventional algorithms and can be calculate practical precision, we calculate Puiseux series expansion by the Sasaki-Kako's method. In **2**, we denote the notation and definition. In **3**, we consider calculating the $e^{A(x)t}$ by applying the the method using Laplace inverse transformation. In **4**, we consider calculating the $e^{A(x)t}$ by applying the Sylvester's interpolation formula. In **5**, we show the some numerical examples.

2 Preliminary

I and $\mathbb{C}\langle\cdot\rangle$ denote the identity matrix, a field of Puiseux series over $\mathbb{C}$, respectively. A matrix $A(x)$ with polynomial entries is called a polynomial matrix. The roots of the characteristic polynomial, $\det(sI - A(x))$of this polynomial matrix, can be expressed as Puiseux series, $\lambda_i(x) \in \mathbb{C}\langle x\rangle$ $(i = 1, \ldots, n)$.

Definition 2.1 (Approximate eigenvalues) A truncated Puiseux series $\lambda_i^{(m)}(x)$ which satisfies the following equation is called as m-th order approximate eigenvalues of polynomial matrix $A(x)$, where $\lambda_i(x)$ is an exact eigenvalues of $A(x)$.

$$\lambda_i(x) \equiv \lambda_i^{(m)}(x) \qquad (\mathrm{mod} \quad x^{m+1})$$

3 The method using Laplace inverse transformation

We calculate the exponent of polynomial matrix $A(x)$ with well-known formula

$$e^{At} = \mathcal{L}^{-1}\{(sI - A)^{-1}\}.$$

First, we compute the characteristic polynomial of $A(x)$,

$$\det(sI - A(x)) = s^n + a_1(x)s^{n-1} + \cdots + a_{n-1}(x)s + a_n(x).$$

The resolvent of $A(x)$, $(sI - A(x))^{-1}$ is expressed as

$$(sI - A(x))^{-1} = \frac{P_1 s^{n-1} + P_2 s^{n-2} + \cdots + P_n}{s^n + a_1(x)s^{n-1} + \cdots + a_{n-1}(x)s + a_n(x)}$$

in general. Multiplying both sides of the equation by $\det(sI - A) \cdot (sI - A)$ from left, we obtain the following equations.

$$\begin{aligned} P_1 &= I, \\ P_2 &= AP_1 + a_1 I, \\ &\vdots \\ P_k &= AP_{k-1} + a_{k-1} I, \\ &\vdots \\ P_n &= AP_{n-1} + a_{n-1} I, \\ 0 &= AP_n + a_n I. \end{aligned}$$

In this way, we can calculate the resolvent. We can easily compute Laplace inverse transformation and get $e^{A(x)t}$ by executing the following partial fractions decomposition. Let the roots of the characteristic polynomial $\det(sI - A(x)) = 0$ of $A(x)$, be $\lambda_1, \ldots, \lambda_m$ and let multiplicity of λ_i be $\sigma_1, \ldots, \sigma_m(\sigma_1 +$

$\sigma_2 + \cdots + \sigma_m = n)$, then the partial fractions decomposition of $(sI - A(x))^{-1}$ is

$$(sI - A)^{-1} = \sum_{i=1}^{m} \left\{ \frac{A_{i1}}{(s-\lambda_i)} + \frac{A_{i2}}{(s-\lambda_i)^2} + \cdots + \frac{A_{i\sigma_i}}{(s-\lambda_i)^{\sigma_i}} \right\}, \qquad (1)$$

where

$$\begin{aligned} A_{ij} &= \lim_{s \to \lambda_i} \frac{1}{(\sigma_i - j)!} \cdot \frac{d^{(\sigma_i - j)}}{ds^{(\sigma_i - j)}} \left\{ (s-\lambda_i)^{\sigma_i} \frac{P_1 s^{n-1} + P_2 s^{n-2} + \cdots + P_n}{s^n + a_1(x) s^{n-1} + \cdots + a_{n-1}(x) s + a_n(x)} \right\} \\ &= \frac{1}{(\sigma_i - j)!} \cdot \frac{d^{(\sigma_i - j)}}{ds^{(\sigma_i - j)}} \left\{ \frac{P_1 s^{n-1} + P_2 s^{n-2} + \cdots + P_n}{\prod_{k=1, \neq i}^{m} (s - \lambda_k)^{\sigma_k}} \right\} \Bigg|_{s=\lambda_i} . \end{aligned}$$

Laplace inverse transformation of $\frac{1}{(s-\lambda_i)^l}$ is

$$\mathcal{L}^{-1} \left\{ \frac{1}{(s-\lambda_i)^l} \right\} = \frac{1}{(l-1)!} t^{l-1} e^{\lambda_i t}.$$

From (1) and the above equation we obtain

$$e^{A(x)t} = \sum_{i=1}^{m} \left\{ A_{i1} + A_{i2} t + \cdots + A_{i\sigma_i} \frac{t^{\sigma_i - 1}}{(\sigma_i - 1)!} \right\} e^{\lambda_i t}.$$

We get the exponent of a polynomial matrix $A(x)$. However since we use approximate eigenvalues, we call this exponent as approximate exponent polynomial matrix.

4 The method using Sylvester's interpolation formula

When the characteristic polynomial of polynomial matrix $A(x)$ is square-free i.e., the roots of $\det(sI - A) = 0$, $\lambda_1, \ldots, \lambda_n$ are different to each other, we can compute $e^{A(x)t}$ using the following Sylvester's interpolation formula.

$$e^{A(x)t} = \sum_{k=1}^{n} P_k e^{\lambda_k t},$$

where

$$P_k = \prod_{\substack{j=1 \\ j \neq k}}^{n} \frac{(A - \lambda_j I)}{(\lambda_k - \lambda_j)}.$$

5 Numerical examples

Example 5.1
Let a polynomial matrix $A(x)$ be

$$A(x) = \begin{pmatrix} 1 & x \\ 2 & 0 \end{pmatrix},$$

then its characteristic polynomial is $\det(sI - A(x)) = s^2 - s - 2x$. In this case, we can compute exact eigenvalues $\frac{1 \pm \sqrt{1+8x}}{2}$. However, it is rare that we can compute the eigenvalues exactly. Hence, we calculate the eigenvalues numerically and obtain a Puiseux series expansions: $\lambda_1^{(3)}(x) = 1 + 2x - 4x^2 + 16x^3$, $\lambda_2^{(3)}(x) = -2x + 4x^2 - 16x^3$. Using the Sylvester's interpolation formula, we can compute $e^{A(x)t}$ as follows.

$$\begin{aligned} e^{A(x)t} \equiv & \begin{pmatrix} 2x - 12x^2 + 80x^3 & -x + 4x^2 - 24x^3 \\ -2 + 8x - 48x^2 + 320x^3 & 1 - 2x + 12x^2 - 80x^3 \end{pmatrix} e^{\lambda_1^{(3)}(x)t} \\ & + \begin{pmatrix} 1 - 2x + 12x^2 - 80x^3 & x - 4x^2 + 24x^3 \\ 2 - 8x + 48x^2 - 320x^3 & 2x - 12x^2 + 80x^3 \end{pmatrix} e^{\lambda_2^{(3)}(x)t} \pmod{x^4} \end{aligned}$$

Example 5.2
Let a polynomial matrix $A(x)$ be

$$A(x) = \begin{pmatrix} 1 & 1+x & x^2-1 \\ 2x & 4 & 5x-3 \\ 6 & x^2+1 & 3x-2 \end{pmatrix}$$

The characteristic polynomial of $A(x)$ is

$$\det(sI - A(x)) = s^3 - (3x+3)s^2 - (5x^3+5x^2-8x-3)s - 2x^5 + 11x^3 - 7x^2 - 21x - 1$$

and we can obtain approximate eigenvalues by Puiseux series expansion as follows.

$$
\begin{aligned}
\lambda_1^{(3)}(x) &= 1+2.51984x^{\frac{1}{3}}-0.26457x^{\frac{2}{3}}+x+0.25939x^{\frac{4}{3}}+1.11349x^{\frac{5}{3}}\\
&\quad +0.0695515x^{\frac{7}{3}}+0.421708x^{\frac{8}{3}}+0.330892x^{\frac{10}{3}}+0.0225432x^{\frac{11}{3}}
\end{aligned}
$$

$$
\begin{aligned}
\lambda_2^{(3)}(x) &= 1+(-1.25992+2.18225i)x^{\frac{1}{3}}+(0.132283+0.229122i)x^{\frac{2}{3}}+x\\
&\quad +(-0.26297+0.455477i)x^{\frac{4}{3}}+(-0.556744-0.964308i)x^{\frac{5}{3}}\\
&\quad +(0.0347758-0.0602334i)x^{\frac{7}{3}}+(-0.210854-0.36521i)x^{\frac{8}{3}}\\
&\quad +(-0.165446+0.286561i)x^{\frac{10}{3}}+(0.0112716+0.019523i)x^{\frac{11}{3}}
\end{aligned}
$$

$$
\begin{aligned}
\lambda_3^{(3)}(x) &= 1+(-1.25992-2.18225i)x^{\frac{1}{3}}+(0.132283-0.229122i)x^{\frac{2}{3}}+x\\
&\quad +(-0.26297-0.455477i)x^{\frac{4}{3}}+(-0.556744+0.964308i)x^{\frac{5}{3}}\\
&\quad +(0.0347758+0.0602334i)x^{\frac{7}{3}}+(-0.210854+0.36521i)x^{\frac{8}{3}}\\
&\quad +(-0.165446-0.286561i)x^{\frac{10}{3}}+(0.0112716-0.019523i)x^{\frac{11}{3}}
\end{aligned}
$$

These eigenvalues are different to each other, so we get

$$
\begin{aligned}
e^{A(x)t} &\equiv \frac{(A(x)-\lambda_2^{(3)}(x))(A(x)-\lambda_3^{(3)}(x))}{(\lambda_1^{(3)}(x)-\lambda_2^{(3)}(x))(\lambda_1^{(3)}(x)-\lambda_3^{(3)}(x))}e^{\lambda_1^{(3)}(x)t}\\
&\quad +\frac{(A(x)-\lambda_1^{(3)}(x))(A(x)-\lambda_3^{(3)}(x))}{(\lambda_2^{(3)}(x)-\lambda_1^{(3)}(x))(\lambda_2^{(3)}(x)-\lambda_3^{(3)}(x))}e^{\lambda_2^{(3)}(x)t}\\
&\quad +\frac{(A(x)-\lambda_1^{(3)}(x))(A(x)-\lambda_2^{(3)}(x))}{(\lambda_3^{(3)}(x)-\lambda_1^{(3)}(x))(\lambda_3^{(3)}(x)-\lambda_2^{(3)}(x))}e^{\lambda_3^{(3)}(x)t}\\
&\qquad \pmod{x^4}
\end{aligned}
$$

by using Sylvester's interpolation formula. In fact, we simplify the expression over the field of Puiseux series over $\mathbb{C}$, however these representations are too long, so we omit these representations.

6 Conclusion

We have show in this paper that we can express $e^{A(x)t}$ approximately, where $A(x)$ has polynomial entries, using Puiseux series expansion of eigenvalues of $A(x)$. This corresponds to solving linear ordinary differential equation with one time-independent parameter.

Acknowledgment

The author is greatly indebted to Professor Tateaki Sasaki, Dr. Kitamoto and my colleagues for helpful discussion.

References

[SK93] T.Sasaki and F.Kako, Solving multivariate algebraic equation by Hensel construction. Preprint of Univ. Tsukuba, January 1993. 22 pages.

[Pui50] M.V.Puiseux, Recherches sur les fonctions algebriques. Journal de Mathematiques pures et appliquées,15 pp. 365-480, 1850.

[Duv89] Dominique Duval. Rational Puiseux Expansions. Composito Mathematica,70 pp. 119-154, 1989.

COMPUTATION OF JORDAN DECOMPOSITION IN THE FORM OF POWER SERIES

Takuya Kitamoto
Institute of Mathematics, University of Tsukuba,
Tsukuba-shi 305, Japan
kita@math.tsukuba.ac.jp

Abstract

It is well-known that every matrix M can be decomposed into $M = VJV^{-1}$, Jordan decomposition, where J is in Jordan normal form and V is a full rank matrix. Algorithms to compute V and J are already known. However, when M has polynomial entries, exact computation of them is rather heavy and often impractical even with moderate size of M. In this paper, we consider efficient computation of Jordan decomposition in the form of truncated power series via Hensel construction. Problem is that Jordan normal form can change, according to the modulus to be taken, which makes Hensel construction quite difficult.

For example, let $M = \begin{bmatrix} 1 & y \\ 0 & 1 \end{bmatrix}$, then Jordan normal form of M with moduli y and y^2 are $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Given a matrix with polynomial entries, we define "modified Jordan decomposition" and present an efficient algorithm to compute the modified Jordan decomposition in the form of truncated power series. In the power series expansion of the modified Jordan decomposition $M \equiv V_m J_m V_m^{-1} \pmod{S^n}$, sub-diagonal elements of J_m are not always 0 or 1 and can be monomials of the variables in M. The decomposition has the following features.

- J_m and V_m can be computed via Hensel construction, i.e., there exist constant matrices $\Delta J_m^{(k)}$ and $\Delta V_m^{(k)}$ which satisfy

$$J_m \pmod{y^{k+1}} = J_m \pmod{y^k} + \Delta J_m^{(k)} y^k,$$
$$V_m \pmod{y^{k+1}} = V_m \pmod{y^k} + \Delta V_m^{(k)} y^k.$$

- Let power series expansion of ordinal Jordan decomposition of M be $M \equiv VJV^{-1} \pmod{y^n}$, then V and J can be easily computed from V_m and J_m.

Therefore, we can compute power series expansion of ordinal Jordan decomposition, computing power series expansion of the modified Jordan decomposition. A numerical example is given to show effectiveness of the algorithm.

1 Introduction

Approximate algebra, which is proposed by Sasaki [Sas 88], is being investigated by Sasaki himself and his colleagues. The study already has shown the applicability of the approximate algebra to solving ill-conditioned polynomial equation [ONS 91], control theory [Kit 94c], etc. More application is being expected. In the approximate algebra, an algebraic function is approximated in two ways;

- It is computed in the form of truncated power series.
- Coefficient of the power series is approximated by finite precision floating numbers.

Based on the concept of the approximate algebra, [Kit 94b] shows various approximate operations on a matrix with polynomial entries, e.g., computation of determinant, LU and singular decomposition, etc. In this paper, we focus on yet another approximate operation, Jordan decomposition. Given a matrix M with polynomial entries, we compute a full rank matrix $V^{(m)}$ and the following matrix $J^{(m)}$

$$J^{(m)} = \begin{bmatrix} J_1(y) & & & \\ & J_2(y) & & \\ & & \ddots & \\ & & & J_p(y) \end{bmatrix}, \quad J_k(y) = \begin{bmatrix} \lambda_k(y) & 1 & & & \\ & \lambda_k(y) & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_k(y) \end{bmatrix}, \tag{1}$$

which satisfy

$$M \equiv VJV^{-1} \pmod{y^{m+1}}, \tag{2}$$

where $\lambda_k(y)$ $(1 \le k \le p)$ are mth order polynomials. As is well-known, Jordan decomposition can be easily constructed from eigenvalues and generalized eigenvectors. [Kit 94b][Kit 95] already described the algorithm to compute eigenvalues and ordinal (not generalized) eigenvectors in the form of truncated power series by Hensel construction. Hence, we focus on only the algorithm to compute "generalized" eigenvectors. Hensel construction of the generalized eigenvectors is not straightforward and may break down as is shown later. To avoid this problem, we define "modified Jordan decomposition," which can be computed smoothly by Hensel construction. The modified Jordan decomposition can be easily transformed to the ordinal Jordan decomposition in (1) and (2).

This paper is organized as follows. In **2**, we give definitions and notations. In **3**, we show how Hensel construction of the ordinal Jordan decomposition breaks down with an illustrative example. In **4**, we define the modified Jordan decomposition and describe the algorithm to compute it by Hensel construction. In **5**, we give a numerical example and finally in 6, conclude.

2 Definitions and notations

Definition 1 (polynomial matrix) A matrix with polynomial entries is called a polynomial matrix.

Definition 2 (polynomial vector) A vector with polynomial entries is called a polynomial vector.

In this paper, we denote the identity matrix and determinant of matrix M by E and $\mathrm{Det}(M)$, respectively. In the following definitions, M is a given $n \times n$ polynomial matrix.

Definition 3 (approximate eigenvalue and eigenvector) Let $\lambda^{(m)}$ and $v^{(m)}$ be a polynomial and a polynomial vector which satisfy

$$Mv^{(m)} \equiv \lambda^{(m)} v^{(m)} \pmod{y^{m+1}}.$$

We call $\lambda^{(m)}$ and $v^{(m)}$ as the mth order approximate eigenvalue and eigenvector of M, respectively.

Definition 4 (approximate generalized eigenvector) Let $\lambda^{(m)}$ be a mth order approximate eigenvalue of M. Let the multiplicity of $\lambda^{(m)}$ and the rank of $\left(M - \lambda^{(m)}E \pmod{y^{m+1}}\right)$ be k and l, respectively. If $k > 1$ and $n - k < l$, then there exist at least one set of polynomial vectors $v_i^{(m)}$ $(i = 1, ..., q)$ which satisfy

$$\begin{aligned} (M - \lambda^{(m)}E)v_1^{(m)} &\equiv 0 \pmod{y^{m+1}}, \\ (M - \lambda^{(m)}E)v_{i+1}^{(m)} &\equiv v_i^{(m)} \pmod{y^{m+1}}. \end{aligned}$$

We call $v_i^{(m)}$ $(1 \le i \le q)$ as the approximate generalized eigenvector. A set of $v_i^{(m)}$ $(i = 1, .., q)$ is called a chain of approximate generalized eigenvectors.

Definition 5 (approximate Jordan decomposition) Computation of full rank matrix $V^{(m)}$ and $J^{(m)}$ in (1) which satisfy (2) is called mth order approximate Jordan decomposition. We call $J^{(m)}$ as mth order approximate Jordan normal form.

Definition 6 (negative-power-free) Let X be either a polynomial, polynomial vector or polynomial matrix. If X has no term with a negative power, we say X is negative-power-free.

3 Breakdown of Hensel construction

Let M be the following polynomial matrix.

$$M = \begin{bmatrix} 1 - y - 2y^2 - y^3 & 4y^2 + 4y^3 + y^4 \\ -y^2 & 1 - y + 2y^2 + y^3 \end{bmatrix} \tag{3}$$

Direct substitution shows that 1st order approximate Jordan decomposition of M is

$$V^{(1)} = \begin{bmatrix} 2+y & 1-y \\ 1 & 1-y \end{bmatrix}, \; J^{(1)} = \begin{bmatrix} 1-y & 0 \\ 0 & 1-y \end{bmatrix}. \tag{4}$$

On the other hand, 2nd order approximate Jordan decomposition of M is

$$V^{(2)} = \begin{bmatrix} 2+y & y^{-2} - y^{-1} - 1 \\ 1 & y^{-2} - y^{-1} \end{bmatrix}, \; J^{(2)} = \begin{bmatrix} 1-y & 1 \\ 0 & 1-y \end{bmatrix}. \tag{5}$$

Note that in the two Jordan decompositions, Jordan normal form $J^{(1)}$ and $J^{(2)}$ are different. Moreover, the power series in $V^{(2)}$ starts from the term with negative power y^{-2}. This example shows that Hensel construction of Jordan decomposition $J^{(m)}$ and $V^{(m)}$ breaks down, since the following condition is necessary to perform the Hensel construction of $J^{(m)}$ and $V^{(m)}$ smoothly;

$$\left.\begin{array}{l}\text{There exist constant matrices } \Delta J^{(k)} \text{ and } \Delta V^{(k)} \text{ which satisfy} \\ \quad J^{(m)} \pmod{y^{k+1}} = J^{(m)} \pmod{y^k} + \Delta J^{(k)} y^k, \\ \quad V^{(m)} \pmod{y^{k+1}} = V^{(m)} \pmod{y^k} + \Delta V^{(k)} y^k. \end{array}\right\} \ (1 \le k \le m) \tag{6}$$

4 Modified Jordan decomposition

To avoid the breakdown of Hensel construction in the previous chapter, we modify Jordan decomposition slightly. More concretely, we compute an approximate generalized eigenvector in negative-power-free form.

Let mth order approximate Jordan decomposition of a given polynomial matrix M as follows;

$$M \equiv V^{(m)} J^{(m)} (V^{(m)})^{-1} \pmod{y^{m+1}} \tag{7}$$

To simplify the explanation, we assume that $J^{(m)}$ is in the following form,

$$J^{(m)} = \begin{bmatrix} \lambda^{(m)} & 1 & & & \\ & \lambda^{(m)} & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda^{(m)} \end{bmatrix}, \tag{8}$$

where $\lambda^{(m)}$ is the approximate eigenvalue of M. We define polynomial vectors $v_i^{(m)}$ as

$$\begin{bmatrix} v_1^{(m)} & v_2^{(m)} & \cdots & v_n^{(m)} \end{bmatrix} \stackrel{\text{def}}{=} V^{(m)}, \tag{9}$$

which are approximate generalized eigenvectors. $v_i^{(m)}$ may not be negative-power-free. Hence, we define negative-power-free polynomial vectors $\tilde{v}_i^{(m)}$, using the $v_i^{(m)}$.

Since $v_i^{(m)}$ are generalized eigenvectors, they satisfy

$$(M - \lambda^{(m)} E) v_1^{(m)} \equiv 0 \pmod{y^{m+1}}, \tag{10}$$

$$(M - \lambda^{(m)} E) v_{l+1}^{(m)} \equiv v_l^{(m)} \pmod{y^{m+1}}. \tag{11}$$

First, note that $v_1^{(m)}$ is always negative-power-free. If some entries of $v_1^{(m)}$ start from the term with negative power y^{-k_1}, then we can put $y^{k_1} v_1^{(m)}$ as $v_1^{(m)}$, which clearly satisfies (10). Hence, we define $\tilde{v}_1^{(m)}$ as

$$\tilde{v}_1^{(m)} \stackrel{\text{def}}{=} v_1^{(m)}. \tag{12}$$

Next, suppose that $v_2^{(m)}$ starts from negative power y^{-k_2}, then we define $\tilde{v}_2^{(m)}$ as

$$\tilde{v}_2^{(m)} \stackrel{\text{def}}{=} y^{k_2} v_2^{(m)}. \tag{13}$$

Thus, we obtain

$$\begin{aligned}(M-\lambda^{(m)}E)\tilde{v}_2^{(m)} &\equiv y^{k_2}v_1^{(m)} \pmod{y^{m+1}},\\ &\equiv y^{k_2}\tilde{v}_1^{(m)} \pmod{y^{m+1}}.\end{aligned}$$

In this way, we obtain negative-power-free polynomial vectors $\tilde{v}_2^{(m)},\cdots,\tilde{v}_n^{(m)}$ which satisfy

$$(M-\lambda^{(m)}E)\tilde{v}_{i+1}^{(m)} \equiv y^{k_{i+1}-k_i}\tilde{v}_i^{(m)} \pmod{y^{m+1}} \quad (i=1,...,n-1) \tag{14}$$

Putting the above equation in the following form,

$$M[\tilde{v}_1^{(m)},\cdots,\tilde{v}_n^{(m)}] \equiv [\tilde{v}_1^{(m)},\cdots,\tilde{v}_n^{(m)}]\begin{bmatrix}\lambda^{(m)} & y^{k_2} & & & \\ & \lambda^{(m)} & y^{k_3-k_2} & & \\ & & \ddots & \ddots & \\ & & & \ddots & y^{k_n-k_{n-1}} \\ & & & & \lambda^{(m)}\end{bmatrix}, \tag{15}$$

we obtain

$$M \equiv \tilde{V}^{(m)}\tilde{J}^{(m)}(\tilde{V}^{(m)})^{-1} \pmod{y^{m+1}}, \tag{16}$$

where

$$\tilde{V}^{(m)} = \begin{bmatrix}\tilde{v}_1^{(m)} & \tilde{v}_2^{(m)} & \cdots & \tilde{v}_n^{(m)}\end{bmatrix}, \tag{17}$$

$$\tilde{J}^{(m)} = \begin{bmatrix}\lambda^{(m)} & y^{k_2} & & & \\ & \lambda^{(m)} & y^{k_3-k_2} & & \\ & & \ddots & \ddots & \\ & & & \ddots & y^{k_n-k_{n-1}} \\ & & & & \lambda^{(m)}\end{bmatrix}. \tag{18}$$

We call the decomposition in (16),(17) and (18) as the modified Jordan decomposition. In the decomposition, $\tilde{J}^{(m)}$ and $\tilde{V}^{(m)}$ are negative-power-free.

To decide k_{i+1} in (14), we compute rank of $\left(M-\lambda^{(m)}E \pmod{y^{l+1}}\right)$ $(0 \le l \le m)$, transforming $\left(M-\lambda^{(m)}E \pmod{y^{m+1}}\right)$ into upper triangular form by row operations Note that the rank of $\left(M-\lambda^{(m)}E \pmod{y^{l+1}}\right)$ and $\left(\tilde{J}^{(m)}-\lambda^{(m)}E \pmod{y^{l+1}}\right)$ should be the same for all $(0 \le l \le m)$. Because we have

$$M-\lambda^{(m)}E \equiv \tilde{V}^{(m)}(\tilde{J}^{(m)}-\lambda^{(m)}E)(\tilde{V}^{(m)})^{-1} \pmod{y^{l+1}} \tag{19}$$

and $\tilde{V}^{(m)}$ is a full rank matrix. Hence, we can decide k_{i+1} so that the rank of $\left(M-\lambda^{(m)}E \pmod{y^{l+1}}\right)$ agrees with that of $\left(\tilde{J}^{(m)}-\lambda^{(m)}E \pmod{y^{l+1}}\right)$ for all $(0 \le l \le m)$.

We give an illustrative example. Let M be the polynomial matrix in (3). Since

$$\mathrm{Det}(xE-M) = (x-1+y)^2, \tag{20}$$

2nd order approximate eigenvalue is $\lambda^{(2)} = 1-y$. We compute the rank of

$$M-\lambda^{(2)}E \equiv \begin{bmatrix}-2y^2 & 4y^2\\ -y^2 & 2y^2\end{bmatrix} \pmod{y^3}. \tag{21}$$

Adding ($1/2 \times$ first row) to the second row of the matrix, we obtain

$$\begin{bmatrix} -2y^2 & 4y^2 \\ 0 & 0 \end{bmatrix} \pmod{y^3}. \tag{22}$$

Hence $\tilde{J}^{(2)}$ in (18) is

$$\tilde{J}^{(2)} \equiv \begin{bmatrix} 1-y & y^2 \\ 0 & 1-y \end{bmatrix} \pmod{y^3}. \tag{23}$$

Therefore, we obtain $k_2 = 2$ and

$$\begin{aligned} (M - \lambda^{(m)}E)\tilde{v}_1 &\equiv 0 \pmod{y^3}, \\ (M - \lambda^{(m)}E)\tilde{v}_2 &\equiv y^2\tilde{v}_1 \pmod{y^3}. \end{aligned}$$

Using the above equations, we perform Hensel construction of $\tilde{v}_1$ and $\tilde{v}_2$. We describe the above procedure formally as algorithms.

Algorithm 1 (Computation of approximate Jordan decomposition)

Input : $n \times n$ polynomial matrix M
Output : An mth order approximate Jordan decomposition $J^{(m)}$ and $V^{(m)}$.

*Step*1 Compute mth order approximate eigenvalues $\lambda_1^{(m)}, \lambda_2^{(m)}, \cdots, \lambda_n^{(m)}$ and put

$$\Lambda \leftarrow \{\lambda_1^{(m)}, \cdots, \lambda_n^{(m)}\}$$

*Step*2 Pick one of the approximate eigenvalue out of Λ and put it as $\lambda^{(m)}$. If the multiplicity of $\lambda^{(m)}$ is one, then remove $\lambda^{(m)}$ from Λ and compute the associated approximate eigenvector with the algorithm in [Kit 95]. Otherwise, remove all $\lambda^{(m)}$ from Λ and apply **algorithm 2** and compute the associated generalized approximate eigenvectors.

*Step*3 Construct approximate Jordan decomposition $J^{(m)}$ and $V^{(m)}$ from the approximate eigenvalues and generalized eigenvectors obtained in *Step*1 and *Step*2.

Algorithm 2 (Computation of approximate generalized eigenvectors)

Input : $n \times n$ polynomial matrix M and mth order approximate eigenvalue $\lambda^{(m)}$
Output : The approximate generalized eigenvectors associated to the eigenvalue

*Step*1 Compute the rank of $\left(M - \lambda^{(m)}E \pmod{y^{m+1}}\right)$ with row operations. For each chain of the generalized eigenvectors $\tilde{v}_i^{(m)}$ $(1 \leq i \leq q)$, decide k_i in (14) and (15). Put $l \leftarrow 0$.

*Step*2 For each chain of the generalized eigenvectors, put 0th order approximate eigenvector $\tilde{v}_i^{(0)}$ as

$$\tilde{v}_i^{(0)} \leftarrow \left[\begin{array}{cccc} v_{0,i,1} & v_{0,i,2} & \cdots & v_{0,i,n} \end{array} \right]^T \tag{24}$$

and decide $v_{0,i,j}$ so that the equations

$$(M - \lambda^{(m)}E)\tilde{v}_i^{(l)} \equiv 0 \pmod{y^{l+1}} \tag{25}$$

$$(M - \lambda^{(m)}E)\tilde{v}_{i+1}^{(l)} \equiv y^{k_{i+1}-k_i}\tilde{v}_i^{(m)} \pmod{y^{l+1}} \tag{26}$$

are satisfied (note that some of the variable $v_{0,i,j}$ may remain undecided).

*Step*3 For each chain of the generalized eigenvectors, put $\Delta\tilde{v}_i^{(l+1)}$ and $\tilde{v}_i^{(l+1)}$ as

$$\Delta\tilde{v}_i^{(l+1)} \leftarrow \left[\begin{array}{cccc} v_{l+1,i,1} & v_{l+1,i,2} & \cdots & v_{l+1,i,n} \end{array} \right]^T \tag{27}$$

$$\tilde{v}^{(l+1)} \leftarrow \tilde{v}_i^{(l)} + \Delta\tilde{v}_i^{(l+1)} \tag{28}$$

and decide $v_{k,i,j}$ so that (25) and (26) are satisfied.

*Step*4 If $m = l+1$ then for each chain of the generalized eigenvectors, substitute appropriate constant numbers into undecided variable $v_{k,i,j}$ in $\tilde{v}_i^{(m)}$ (undecided variables are free parameters) and compute approximate generalized eigenvectors $v_i^{(m)} = y^{-k_i}\tilde{v}_i^{(m)}$ and return $v_i^{(m)}$. Otherwise put $l \leftarrow l+1$ and go to *Step*3.

5 Numerical example

Let M be the following matrix, where $M_{i,j}$ denotes (i,j)th element of M.

$$\begin{array}{lll}
M_{1,1} = 1 - \frac{7}{2}y^2 + \frac{7}{2}y^4 + y^5, & & M_{1,2} = -y + y^2 + \frac{7}{2}y^3 + y^4, \\
M_{1,3} = y + \frac{5}{2}y^2 - \frac{5}{2}y^3 - y^4, & & M_{1,4} = 3y - 9y^3 - 9y^4 - 2y^5, \\
M_{2,1} = -y + \frac{1}{2}y^2 + \frac{3}{2}y^3 + y^4 + \frac{1}{2}y^5 - \frac{1}{2}y^6, & & M_{2,2} = 1 + 2y^2 + \frac{1}{2}y^3 + \frac{1}{2}y^4 - \frac{1}{2}y^5, \\
M_{2,3} = y - \frac{3}{2}y^2 - y^4 + \frac{1}{2}y^5, & & M_{2,4} = -3y - 5y^2 - 6y^3 - 2y^4 + y^6, \\
M_{3,1} = 2y - y^2 - \frac{11}{2}y^3 - y^4 + \frac{3}{2}y^5, & & M_{3,2} = -2y - 4y^2 - y^3 + \frac{3}{2}y^4, \\
M_{3,3} = 1 - y + 3y^2 + \frac{5}{2}y^3 - \frac{3}{2}y^4, & & M_{3,4} = 6y + 15y^2 + 10y^3 - y^4 - 3y^5, \\
M_{4,1} = -\frac{1}{2}y^2 + \frac{1}{4}y^3 + y^4 + \frac{1}{4}y^5, & & M_{4,2} = \frac{1}{2}y^2 - y^3 + \frac{1}{4}y^4, \\
M_{4,3} = \frac{1}{2}y^2 - \frac{3}{4}y^3 - \frac{1}{4}y^4, & & M_{4,4} = 1 - y - \frac{3}{2}y^2 - 3y^3 - \frac{5}{2}y^4 - \frac{1}{2}y^5.
\end{array}$$

We will compute 2nd order approximate Jordan decomposition of M.

Trace of algorithm 1

*Step*1 Computing 2nd order approximate eigenvalues of M with the algorithm in [Kit 95], we obtain

$$\lambda_1^{(2)} = 1 + y,\ \lambda_2^{(2)} = \lambda_3^{(2)} = \lambda_4^{(2)} = 1 - y.$$

Hence we put

$$\Lambda \leftarrow \{1+y, 1-y, 1-y, 1-y\}$$

*Step*2 We pick $1+y$ out of Λ and put $\lambda^{(2)} \leftarrow 1+y$. Since the multiplicity of $\lambda^{(2)}$ is one, we apply the algorithm in [Kit 95] and obtain the associated approximate eigenvector

$$v_1^{(2)} = \begin{bmatrix} 1, & 0, & 1+y, & 0, \end{bmatrix}^T.$$

*Step*2 We pick $1-y$ out of Λ and put $\lambda^{(2)} \leftarrow 1-y$. Since the multiplicity of $\lambda^{(2)}$ is not one, we remove all $1-y$ from Λ (hence Λ becomes $\{\}$) and apply **algorithm 2.**

Trace of algorithm 2

*Step*1 Transforming $\left(M - \lambda^{(m)}E \pmod{y^{m+1}}\right)$ into upper triangular form, we obtain

$$\begin{bmatrix} y-\frac{7}{2}y^2 & -y+y^2 & y+\frac{5}{2}y^2 & 3y \\ 0 & 0 & 2y+4y^2 & 4y^2 \\ 0 & 0 & 0 & y^2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \pmod{y^3}.$$

Therefore approximate Jordan normal form $J^{(2)}$ is

$$J^{(2)} \equiv \begin{bmatrix} 1+y & 0 & 0 & 0 \\ 0 & 1-y & y & 0 \\ 0 & 0 & 1-y & y^2 \\ 0 & 0 & 0 & 1-y \end{bmatrix} \pmod{y^3}$$

and $k_2 = 1$, $k_3 = 3$. Hence we have the chain of the generalized eigenvectors, $\tilde{v}_1^{(m)}, \tilde{v}_2^{(m)}$ and $\tilde{v}_3^{(m)}$, which satisfy

$$\begin{aligned} (M - \lambda^{(m)}E)\tilde{v}_1^{(m)} &\equiv 0 \pmod{y^{m+1}}, \\ (M - \lambda^{(m)}E)\tilde{v}_2^{(m)} &\equiv y\tilde{v}_1^{(m)} \pmod{y^{m+1}}, \\ (M - \lambda^{(m)}E)\tilde{v}_3^{(m)} &\equiv y^2\tilde{v}_2^{(m)} \pmod{y^{m+1}}. \end{aligned}$$

We put $l \leftarrow 0$.

*Step*2 Put $\tilde{v}_i^{(0)}$ as in (24) and decide $v_{0,i,j}$ so that (25) and (26) are satisfied. The equations (25) and (26) are automatically satisfied and any of $v_{0,i,j}$ is not decided.

*Step*3 Put $\Delta\tilde{v}_i^{(1)}$ and $\tilde{v}_i^{(1)}$ as in (27) and (28) and decide $v_{k,i,j}$ so that (25) and (26) are satisfied and obtain

$$\begin{aligned} v_{0,1,1} &= v_{0,2,3} - 3v_{0,1,4}, & v_{0,1,2} &= v_{0,2,3}, & v_{0,1,4} &= 0, \\ v_{0,2,1} &= v_{0,2,2} - 3v_{0,2,4}, & v_{0,3,1} &= v_{0,3,2} - 3v_{0,3,4}, & v_{0,3,3} &= 0. \end{aligned}$$

Hence,

$$\begin{aligned} v_1^{(1)} &= \begin{bmatrix} v_{0,2,3} + v_{1,1,1}y, & v_{0,2,3} + v_{1,1,2}y, & v_{1,1,3}y, & v_{1,1,4}y \end{bmatrix}^T, \\ v_2^{(1)} &= \begin{bmatrix} v_{0,2,2} - 3v_{0,2,4} + v_{1,2,1}y, & v_{0,2,2} + v_{1,2,2}y, & v_{0,2,3} + v_{1,2,3}y, & v_{0,2,4} + v_{1,2,4}y \end{bmatrix}^T, \\ v_3^{(1)} &= \begin{bmatrix} v_{0,3,2} - 3v_{0,3,4} + v_{1,3,1}y, & v_{0,3,2} + v_{1,3,2}y, & v_{1,3,3}y, & v_{0,3,4} + v_{1,3,4}y \end{bmatrix}^T. \end{aligned}$$

*Step*4 Since $l+1<2$, we put

$$l \leftarrow 0+1$$

and go to *Step*3.

*Step*3 Put $\Delta\tilde{v}_i^{(2)}$ and $\tilde{v}_i^{(2)}$ as in (27) and (28) and decide $v_{k,i,j}$ so that equations (25) and (26) are satisfied and obtain

$$\begin{array}{l} v_{0,2,2}=\frac{6}{5}v_{1,1,4}+\frac{2}{5}v_{1,2,1}-\frac{2}{5}v_{1,2,2}+\frac{6}{5}v_{1,2,4}, \quad v_{0,2,3}=\frac{2}{5}v_{1,1,1}-\frac{2}{5}v_{1,1,2}+\frac{6}{5}v_{1,1,4}, \\ v_{0,2,4}=0, \quad v_{0,3,2}=\frac{34}{5}v_{1,1,4}+\frac{2}{5}v_{1,3,1}-\frac{2}{5}v_{1,3,2}+\frac{6}{5}v_{1,3,4}, \\ v_{0,3,4}=2v_{1,1,4}, \quad v_{1,1,1}=v_{1,1,2}+2v_{1,1,4}, \\ v_{1,1,3}=0, \quad v_{1,2,1}=7v_{1,1,4}+v_{1,2,2}-3v_{1,2,4}+\frac{5}{2}v_{1,3,3}, \quad v_{1,2,3}=v_{1,1,2}. \end{array}$$

Hence,

$$v_1^{(2)} = \begin{bmatrix} 2v_{1,1,4}+(v_{1,1,2}+2v_{1,1,4})y+v_{2,1,1}y^2 \\ 2v_{1,1,4}+v_{1,1,2}y+v_{2,1,2}y^2 \\ v_{2,1,3}y^2 \\ v_{1,1,4}y+v_{2,1,4}y^2 \end{bmatrix},$$

$$v_2^{(2)} = \begin{bmatrix} 4v_{1,1,4}+v_{1,3,3}+(7v_{1,1,4}+v_{1,2,2}-3v_{1,2,4}+\frac{5}{2}v_{1,3,3})y+v_{2,2,1}y^2 \\ 4v_{1,1,4}+v_{1,3,3}+v_{1,2,2}y+v_{2,2,2}y^2 \\ 2v_{1,1,4}+v_{1,1,2}y+v_{2,2,3}y^2 \\ v_{1,2,4}y+v_{2,2,4}y^2 \end{bmatrix},$$

$$v_3^{(2)} = \begin{bmatrix} \frac{4v_{1,1,4}+2v_{1,3,1}-2v_{1,3,2}+6v_{1,3,4}}{5}+v_{1,3,1}y+v_{2,3,1}y^2 \\ \frac{34v_{1,1,4}+2v_{1,3,1}-2v_{1,3,2}+6v_{1,3,4}}{5}+v_{1,3,2}y+v_{2,3,2}y^2 \\ v_{1,3,3}y+v_{2,3,3}y^2 \\ 2v_{1,1,4}+v_{1,3,4}y+v_{2,3,4}y^2 \end{bmatrix}.$$

*Step*4 Since $l+1=2$, we put

$$v_{1,1,2}=1, \; v_{1,1,4}=-1, \; v_{1,3,2}=1$$

and all other undecided variables as 0, obtaining

$$\tilde{v}_1^{(2)}=\begin{bmatrix} -2-y \\ -2+y \\ 0 \\ -y \end{bmatrix}, \; \tilde{v}_2^{(2)}=\begin{bmatrix} -4-7y \\ -4 \\ -2+y \\ 0 \end{bmatrix}, \; \tilde{v}_3^{(2)}=\begin{bmatrix} -\frac{6}{5} \\ -\frac{36}{5}+y \\ 0 \\ -2 \end{bmatrix}.$$

Hence, approximate generalized eigenvectors $v_1^{(2)}, v_2^{(2)}, v_3^{(2)}$ are

$$v_1^{(2)}=\begin{bmatrix} -2-y \\ -2+y \\ 0 \\ -y \end{bmatrix}, \; v_2^{(2)}=\begin{bmatrix} -\frac{4}{y}-7 \\ -\frac{4}{y} \\ -\frac{2}{y}+1 \\ 0 \end{bmatrix}, \; v_3^{(2)}=\begin{bmatrix} -\frac{6}{5y^3} \\ -\frac{36}{5y^3}+\frac{1}{y^2} \\ 0 \\ -\frac{2}{y^3} \end{bmatrix}.$$

Thus, we obtain 2nd order approximate Jordan decomposition $J^{(2)}$ and $V^{(2)}$ as

$$V^{(2)}=\begin{bmatrix} 1 & -2-y & -\frac{4}{y}-7 & -\frac{6}{5y^3} \\ 0 & -2+y & -\frac{4}{y} & -\frac{36}{5y^3}+\frac{1}{y^2} \\ 1+y & 0 & -\frac{2}{y}+1 & 0 \\ 0 & -y & 0 & -\frac{2}{y^3} \end{bmatrix}, \; J^{(2)}=\begin{bmatrix} 1+y & 0 & 0 & 0 \\ 0 & 1-y & 1 & 0 \\ 0 & 0 & 1-y & 1 \\ 0 & 0 & 0 & 1-y \end{bmatrix}.$$

6 Conclusion

[Kit 94b] described various approximate operations on a polynomial matrix. This paper proposed yet another approximate operation, approximate Jordan decomposition. We showed that in the ordinal form of Jordan decomposition, Hensel construction may break down and proposed the modified Jordan decomposition. Hensel construction can be performed smoothly in the modified Jordan decomposition, which can be easily transformed to the ordinal Jordan decomposition. We gave a numerical example to show effectiveness of the algorithm.

Acknowledgments

The author is greatly indebted to Professor Tateaki Sasaki for helpful discussion.

References

[Kit 94a] T. Kitamoto "Approximate Eigenvalues, Eigenvectors and Inverse of a Matrix with Polynomial Entries," Jpn. J. Indus. App. Math., Vol. 11, pp. 73–85, 1994.

[Kit 94b] T. Kitamoto "Approximate Operations on Matrix," Winter Workshop on Computer Algebra, Jan. Tokyo, (Proc. of Extended Abstracts, pp. 16–20), 1994.

[Kit 94c] T. Kitamoto "Application of approximate algebraic computation to optimal control (in Japanese)," to appear in EICS, 1994.

[Kit 95] T. Kitamoto "On Puiseux Expansions of Approximate Eigenvalues and Eigenvectors," submitted to Jpn. J. Indus. App. Math. 1995.

[ONS 91] M. Ochi, M-T. Noda and T. Sasaki "Approximate GCD of Multivariate Polynomials and Application to Ill-conditioned Algebraic Equation," J. Inf. Process., Vol. 14, pp. 292–300, 1991.

[Sas 88] T. Sasaki "Approximate Algebraic Computation (in Japanese)," Suriken-Kokyuroku (Collection of Research Reports, Research Institute of Mathematical Study, Kyoto University), Vol. 676, pp. 307–319, 1988.

[SK 93] T. Sasaki and F. Kako "Solving Multivariate Algebraic Equation by Hensel Construction," preprint (Univ. of Tsukuba and Nara Women Univ.), 22 pages, Jan. 1993.

[SN 89] T. Sasaki and M-T. Noda "Approximate Square-free Decomposition and Root-finding of Ill-conditioned Algebraic Equation," J. Inf. Process., Vol. 12, pp. 159–168, 1989.

[SSKS 91] T. Sasaki, M. Suzuki, M. Kolar and M. Sasaki "Approximate Factorization of Multivariate Polynomials and Absolute Irreducibility Testing," Jpn. J. Indus. App. Math., Vol. 8, pp. 357–375, 1991.

SEPARATION OF CLOSE ROOTS BY LINEAR FRACTION TRANSFORMATION

Hidetsune KOBAYASHI
Dept. of Math, Nihon University

Hideo SUZUKI
Dept. of Comp. Science, Tokyo Polytechnic College

1 Introduction

We have some examples of algebraic equation having a group of close solutions. And sometimes a numerical calculation cannot separate these solutions. In such case, we cannot see if a solution obtained by a numerical calculation is a solution with some multiplicity or some solutions being observed as one solution. In fact, we saw the multiplicity obtained by a numerical calculation with Zeuthen's rule(see [1],[2]) was false when the solutions were very close to each other.

Hence, to calculate the multiplicity, we have to separate those solutions very close to each other.

In this paper, we present a method to separate those solutions very close to each other by using a linear fractional transformation. Of course the transformation can be used in case having more than two variables, but now we state only one variable case.

We also show that we can derive an equation such that

1. of degree 1 less than the original one.
2. solutions are approximately equal to solutions of the original one.

We call this procedure to get a new equation *pseudo-localization*(see [3]), since localization at a point eliminates from functions the zero at the point.

2 Linear fractional transformations

We consider the solutions to one variable algebraic equation $f(x) = 0$. To introduce a linear fractional transformation, we homogenize $f(x)$ with a homogeneous coordinate (X, Y) as

$$F(X,Y) = Y^{\deg f} f(X/Y) \tag{1}$$

Let $G(U, V)$ be the homogeneous polynomial obtained by the projective transformation

$$\begin{cases} U &= aX + bY \\ V &= cX + dY \end{cases}$$

We write the inverse of this projective transformation as

$$\begin{cases} X &= a'U + b'V \\ Y &= c'U + d'V \end{cases}$$

Then $G(U, V)$ can be written as follows.

$$G(U,V) = F(a'U + b'V,\ c'U + d'V)$$

If we choose the projective transformation properly, then the point $V = 0$ does not correspond to the point which is a zero of $F(X, Y)$. We employ the affine coordinate $u = U/V$, then the point $V = 0$ is the point at infinity with respect to this coordinate u. The equation

$$G(u, 1) = 0 \text{ where } u = U/V$$

has the solutions that correspond to the original equation $f(x) = 0$ and each corresponding solution has the same multiplicity. Thus, we say that a properly chosen linear fractional transformation preserves the multiplicity.

3 Separation of close solutions

To make a situation simple, at first, we assume that the solutions of $f(x)$ are all real numbers. Of course, we can extend our discussion to the general case having complex numbers. We write the solutions of $f(x) = 0$ as $\alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s$ $(r + s = n)$. Then we have

$$f(x) = \prod_{i=1}^{r}(x - \alpha_i) \prod_{j=1}^{s}(x - \beta_j) \qquad (r + s = n)$$

In this expression , we assume that $\alpha_1, \alpha_2, \cdots, \alpha_r$ are the solutions very close to each other and we assume $\beta_1, \beta_2, \cdots, \beta_s$ are solutions fairly remote from those close solutions. We may assume for a properly chosen real number α and a small positive number ϵ, the close solutions are contained in the interval $(\alpha - \epsilon, \alpha + \epsilon)$

A linear fractional transformation $u = (ax - b)/(cx - d)$ is solved for x. That is $x = (ud - b)/(uc - a)$. We obtain the next expression by substituting this into $f(x)$.

$$(\frac{1}{uc - a})^n \prod_{i=1}^{r}(d - \alpha_i c) \prod_{i=1}^{r}(u - \frac{b - \alpha_i a}{d - \alpha_i c}) \times \prod_{j=1}^{s}(d - \beta_j c) \prod_{j=1}^{s}(u - \frac{b - \beta_j a}{d - \beta_j c})$$

Now we put $h(u)$ as

$$h(u) = \prod_{i=1}^{r}(u - \frac{b - \alpha_i a}{d - \alpha_i c}) \times \prod_{j=1}^{s}(u - \frac{b - \beta_j a}{d - \beta_j c})$$

Here we note that 1. when we use a computer language performing symbolic computation, we can execute this calculation without any numerical error, 2. the size of the coefficients of this expression can be made almost the same level as the coefficients of the original equation.

The solutions corresponding to the close solutions become $(b - \alpha_i a)/(d - \alpha_i c)$ for $i = 1, \ldots, r$. The solutions corresponding other solutions become $(b - \beta_j a)/(d - \beta_j c)$ for $j = 1, \ldots, s$. Next, we analyze numerical errors transformed by the linear fractional transformation.

Let u_i be an approximation to a solution $\tilde{u}_i$ of $h(u) = 0$. And let $\tilde{x}_i$ be the solution of $f(x) = 0$ corresponding to $\tilde{u}_i$. We write errors as $\Delta u_i = u_i - \tilde{u}_i$, $\Delta x_i = x_i - \tilde{x}_i$. Then we have the following relation.

$$\Delta x_i = x_i - \tilde{x}_i = \frac{ad - bc}{(\tilde{u}_i c - a)(u_i c - a)} \Delta u_i \tag{2}$$

In the next subsections, we test following four kinds of linear fraction transformation φ having specified values of a, b, c, d.

3.1 Approximate solution of the original equation is not needed

1. $a = 1, b = \alpha, c = 1, d = \alpha - \gamma$ ($\gamma = k\epsilon, k \geq 1$)

   In this case, the type of linear fractional transformation and the correspondence of the linear fractional transformation are as follows.

   $$\varphi(\infty) = 1,\ \varphi(\alpha) = 0,\ \varphi(\alpha - \gamma) = \infty,\ u = \varphi(x) = \frac{x - \alpha}{x - \alpha + \gamma},\ x = \varphi^{-1}(u) = \frac{(\alpha - \gamma)u - \alpha}{u - 1}$$

   The solutions corresponding to close solutions and the other solutions are as follows respectively.

   $$\frac{c_i}{c_i + k}\quad (c_i\epsilon = \alpha_i - \alpha, |c_i| < 1)\ \ i = 1, \cdots, r,\quad \frac{\beta_j - \alpha}{\beta_j - \alpha + \gamma}\quad j = 1, \cdots, s$$

   The interval $(\alpha - \epsilon, \alpha + \epsilon)$ containing close roots is expanded to $(\frac{-1}{k-1}, \frac{1}{k+1})$.

2. $a = 0, b = \gamma, c = 1, d = \alpha - \gamma$ ($\gamma = k\epsilon, k \geq 1$)

   In this case, the type of linear fractional transformation and its correspondence are as follows.

   $$\varphi(\infty) = 0,\ \varphi(\alpha) = -1,\ \varphi(\alpha - \gamma) = \infty,\ u = \varphi(x) = \frac{-\gamma}{x - \alpha + \gamma},\ x = \varphi^{-1}(u) = \frac{(\alpha - \gamma)u - \gamma}{u}$$

   The solutions corresponding to close roots and other solutions are as follows respectively.

   $$\frac{-k}{k + c_i}\quad (c_i\epsilon = \alpha_i - \alpha, |c_i| < 1)\ \ i = 1, \cdots, r,\quad \frac{-\gamma}{\beta_j - \alpha + \gamma}\quad j = 1, \cdots, s$$

   The interval $(\alpha - \epsilon, \alpha + \epsilon)$ is expanded to $(-1 - \frac{1}{k-1}, -1 + \frac{1}{k+1})$.

3.2 Approximate solution of the original equation is needed

3. $a = \beta_l - \alpha + \gamma, b = (\beta_l - \alpha + \gamma)\alpha, c = \beta_l - \alpha, d = (\beta_l - \alpha)(\alpha - \gamma)$
   ($\gamma = k\epsilon, k \geq 1, |\alpha - \beta_l| > |\alpha - \beta_j|$)

   In this case, the type of linear fractional transformation and correspondence of the linear fractional transformation are as follows.

   $$\varphi(\beta_l) = 1, \varphi(\alpha) = 0, \varphi(\alpha - \gamma) = \infty$$

   $$u = \varphi(x) = \frac{(\beta_l - \alpha + \gamma)(x - \alpha)}{(\beta_l - \alpha)(x - \alpha + \gamma)},\ x = \varphi^{-1}(u) = \frac{(\beta_l - \alpha)(\alpha - \gamma)u - \alpha(\beta_l - \alpha + \gamma)}{(\beta_l - \alpha)(u - 1) - \gamma}$$

   The solutions corresponding to close roots and the other solutions are as follows respectively.

   $$\frac{(\beta_l - \alpha + \gamma)c_i}{(\beta_l - \alpha)(c_i + k)}\quad (c_i\epsilon = \alpha_i - \alpha, |c_i| < 1)\quad i = 1, \cdots, r$$

   $$1\quad (j = l),\ \frac{(\beta_l - \alpha + \gamma)(\beta_j - \alpha)}{(\beta_l - \alpha)(\beta_l - \alpha + \gamma)}\ (j \neq l)\qquad j = 1, \cdots, s$$

   The interval $(\alpha - \epsilon, \alpha + \epsilon)$ is expanded to $(\frac{-(\beta_l - \alpha + \gamma)}{(\beta_l - \alpha)(k-1)}, \frac{\beta_l - \alpha + \gamma}{(\beta_l - \alpha)(k+1)})$.

4. $a = \gamma, b = \beta_l\gamma, c = \beta_l - \alpha, d = (\beta_l - \alpha)(\alpha - \gamma)$ $(\gamma = k\epsilon, k \geq 1, |\alpha - \beta_l| > |\alpha - \beta_j|)$

The linear fractional transformation and its correspondence are as follows.

$$\varphi(\beta_l) = 0, \varphi(\alpha) = -1, \varphi(\alpha - \gamma) = \infty$$

$$u = \varphi(x) = \frac{(x - \beta_l)\gamma}{(\beta_l - \alpha)(x - \alpha + \gamma)}, \quad x = \varphi^{-1}(u) = \frac{(\beta_l - \alpha)(\alpha - \gamma)u - \beta_l\gamma}{(\beta_l - \alpha)u - \gamma}$$

The solutions corresponding to close roots and the other are as follows respectively.

$$\frac{-(\beta_l - \alpha_i - c_i\epsilon)k}{(\beta_l - \alpha)(k + c_i)} \quad (c_i\epsilon = \alpha_i - \alpha, |c_i| < 1) \quad i = 1, \cdots, r$$

$$0 \quad (j = l), \quad \frac{(\beta_j - \beta_l)\gamma}{(\beta_l - \alpha)(\beta_j - \alpha + \gamma)} \quad (j \neq l) \quad j = 1, \cdots, s$$

The interval $(\alpha - \epsilon, \alpha + \epsilon)$ is expanded to $(-1 - \frac{\beta_l - \alpha + \gamma}{(\beta_l - \alpha)(k - 1)}, -1 + \frac{\beta_l - \alpha + \gamma}{(\beta_l - \alpha)(k + 1)})$.

3.3 Errors transformed by linear fractional transformations

Solutions corresponding to close solutions of the original equation are transformed as follows.

Transformation 1,3 : Close roots are mapped to points spread around 0.

Transformation 2,4 : Close roots are mapped to points spread around -1.

Therefore, if we take the value of k (it is about 10 to 1000) chosen suitably, the solutions to the transformed equation which correspond to the close solutions of the original equation are separated greatly.

Next, we describe the errors in these four linear fractional transformations.

Proposition 3.1 *Let u_i be an approximate solution to the solution $\tilde{u}_i$ of $h(u) = 0$. The solution of $f(x) = 0$ corresponding to $\tilde{u}_i$ is denoted by $\tilde{x}_i$. We write errors as $\Delta u_i = u_i - \tilde{u}_i$, $\Delta x_i = x_i - \tilde{x}_i$. Then we have the relation $|\Delta x_i| \fallingdotseq \gamma|\Delta u_i|$*

[Proof] Substituting special values to a, b, c, d sown as above for the expressions (2), the following are obtained.

Transformation 1. $\Delta x_i = x_i - \tilde{x}_i = \frac{1}{(\tilde{u}_i - 1)(u_i - 1)}\gamma\Delta u_i$

Transformation 2. $\Delta x_i = \frac{1}{u_i\tilde{u}_i}\gamma\Delta u_i$

Transformation 3. $\Delta x_i = \frac{-(\beta_l - \alpha + \gamma)(\beta_l - \alpha)\gamma}{\{(\beta_l - \alpha)(\tilde{u}_i - 1) - \gamma\}\{(\beta_l - \alpha)(u_i - 1) - \gamma\}}\Delta u_i$

Transformation 4. $\Delta x_i = \frac{(\beta_l - \alpha)(\alpha - \gamma - \beta_l)}{\{(\beta_l - \alpha)u_i - \gamma\}\{(\beta_l - \alpha)\tilde{u}_i - \gamma\}}\gamma\Delta u_i$

Here $\gamma = k\epsilon \ll 1$ and $|u_i|, |\tilde{u}_i| \ll 1$ (transformation 1 and 3) and $u_i, \tilde{u}_i \fallingdotseq -1$ (transformation 2 and 4), the above equations are written as

Transformation 1. $\Delta x_i = \frac{1}{(\tilde{u}_i - 1)(u_i - 1)}\gamma\Delta u_i \fallingdotseq \gamma\Delta u_i$

Transformation 2. $\Delta x_i = \frac{1}{u_i\tilde{u}_i}\gamma\Delta u_i \fallingdotseq \gamma\Delta u_i$

Transformation 3. $\Delta x_i \fallingdotseq \dfrac{-(\beta_l - \alpha + \gamma)(\beta_l - \alpha)\gamma}{(\beta_l - \alpha + \gamma)^2} \Delta u_i \fallingdotseq -\gamma \Delta u_i$

Transformation 4. $\Delta x_i \fallingdotseq \dfrac{(\beta_l - \alpha)(\beta_l - \alpha)}{(\beta_l - \alpha)^2} \gamma \Delta u_i \fallingdotseq \gamma \Delta u_i$

Therefore

$$|\Delta x_i| \fallingdotseq \gamma |\Delta u_i| \tag{3}$$

□

We see the following properties hold.

Four types of linear fractional transformations have the relation $|\Delta x_i| \fallingdotseq \gamma |\Delta u_i|$. From this equation, we see that an approximate solution u_i having error Δu_i is mapped back to x_i, and its error is γ times Δu_i. In general, γ is a very small positive number. We can say that the linear fractional transformation is useful because the error of the solution to the original equation is γ times the error that arise while calculating to obtain a solution in the transformed space. Moreover when ϵ is smaller (the separation in the original space becomes difficult in numerical calculation), γ becomes smaller, and the separation in the new space gives more precise solutions.

4 Pseudo localization

The pseudo localization is a method composing an equation with one less degree than the original equation and the new equation contains approximations of all close solutions. We assume a group of close solutions are remote from the other solutions. Moreover, we assume the solutions other than close solutions are calculated in good accuracy.

If such pseudo localization procedure is applied repeatedly, we can obtain an equation of lower degree.

We consider the four linear fractional transformations given in the preceding section.

4.1 Approximate solution of the original equation is not needed

1. $a = 1, b = \alpha, c = 1, d = \alpha - \gamma, (\gamma = k\epsilon, k \geq 1)$

   In this case, close solutions are mapped around 0 and the distance of each pair of close solutions are expanded greatly. Moreover, the other solutions are mapped in a small interval containing 1 . Because, solutions not corresponding to close solutions of the original equation are transformed as follows.

   $$\frac{\alpha - \beta_j}{\alpha - \beta_j - \gamma} = \frac{1}{1 - \frac{\gamma}{\alpha - \beta_j}} = \frac{1}{1 - \frac{k\epsilon}{\alpha - \beta_j}}$$

   Here, β_j is located far from the close solutions. Even if we choose k from 10 to 1000 , we have $|k\epsilon/(\alpha - \beta_j)| << 1$. Hence we see above value is very close to 1.

   A new equation having one less degree is obtained by removing $u - 1$ from the transformed expression $h(u)$. That is

   $$h(u) = g_1(u)(u-1) + h(1)$$

   Since $|h(1)|$ is very small, we see $h(u) \fallingdotseq g_1(u)(u-1)$, and $g_1(u)$ gives whole solutions corresponding to the close solutions of the original equation. In fact, we only have to calculate $g_1(u)$ by synthetic division.

2. $a = 0, b = \gamma, c = 1, d = \alpha - \gamma, (\gamma = k\epsilon, k \geq 1)$

In this case, close solutions are spread near -1. The other solutions are mapped around 0. Because, these solutions remote from close solutions are mapped as

$$\frac{\gamma}{\beta_j - \alpha + \gamma} = \frac{k\epsilon}{\beta_j - \alpha + k\epsilon}$$

Since β_j is remote from close solutions, even if we take k from 10 to 1000, we have $|k\epsilon/(\beta_j - \alpha + k\epsilon)| \fallingdotseq 0$.

A new equation having 1 less degree is obtained formally removing u from transformed expression $h(u)$. That is

$$h(u) = g_1(u)u + h(0)$$

$g_1(u)$ of this equation gives an equation to be obtained.

4.2 Approximate solution of the original equation is needed

Any approximate solution can be used to define a linear fractional transformation, but here we use the solution with maximum absolute value.

3. $a = \beta_l - \alpha + \gamma, b = (\beta_l - \alpha + \gamma)\alpha, c = \beta_l - \alpha, d = (\beta_l - \alpha)(\alpha - \gamma)$
$(\gamma = k\epsilon, k \geq 1, |\alpha - \beta_l| > |\alpha - \beta_j|)$

In this case, close solutions are spread near 0. A specific solution other than close solutions become just 1.

An equation with one less degree is obtained by removing $u - 1$ from the transformed expression $h(u)$. That is

$$h(u) = g_1(u)(u - 1)$$

$g_1(u)$ gives the polynomial to be obtained.

4. $a = \gamma, b = \beta_l\gamma, c = \beta_l - \alpha, d = (\beta_l - \alpha)(\alpha - \gamma), (\gamma = k\epsilon, k \geq 1, |\alpha - \beta_l| > |\alpha - \beta_j|)$

In this case, close solutions are spread around 1. A specific solution other than close solutions become just 0.

An equation with one less degree is obtained by removing u from $h(u)$. That is

$$h(u) = g_1(u)u$$

4.3 Errors of pseudo localization

Transformations 1,2 can be defined without using approximation solutions at all. Transformations 3,4 use approximate solution other than close solutions. Therefore, it is necessary to analyze the errors of for transformations 1, 2 and transformations 3, 4 separately.

We write a polynomial obtained by the linear transformations as

$$h(u) = \prod_{i=1}^{r}(u - \tilde{\alpha}_i)\prod_{j=1}^{s}(u - \tilde{\beta}_j)$$

Here $\tilde{\alpha}_i = \varphi(\alpha_i), \tilde{\beta}_j = \varphi(\beta_j)$. Now we have

Proposition 4.1 *Let $\tilde{\alpha}_{i0}$ be a solution of $h(u) = 0$ that corresponds to one of the close solutions. Let $\tilde{\alpha}_{i0} + \Delta\tilde{\alpha}_{i0}$ be a solution of $g_1(u) = 0$ corresponding to $\tilde{\alpha}_{i0}$. Then we have the relation*

$$|\Delta\tilde{\alpha}_{i0}| \fallingdotseq \left|\frac{k(c_{i0}+k)^{n-2}}{-\prod_{i\neq i0,i=1}^{r}(c_{i0}-c_i)\prod_{j=1}^{s}(\alpha+c_{i0}\epsilon-\beta_j)+(c_{i0}+k)^{n-1}\epsilon^s}\right|\epsilon^s$$

Here $c_i\epsilon = \alpha_i - \alpha, c_{i0}\epsilon = \alpha_{i0} - \alpha$.

Especially, the following relation is obtained when we put $\alpha + c_{i0}\epsilon - \beta_j \fallingdotseq \alpha - \beta_j$ for very small positive ϵ.

$$|\Delta\tilde{\alpha}_{i0}| \fallingdotseq \left|\frac{k(c_{i0}+k)^{n-2}}{\prod_{i\neq i0,i=1}^{r}(c_{i0}-c_i)\prod_{j=1}^{s}(\alpha-\beta_j)}\right|\epsilon^s \tag{4}$$

[Proof] Here, we give a proof for the transformation 1. For the transformation 2, the proof is similar.

$$h'(u) = \sum_{l=1}^{r}\prod_{i\neq l,i=1}^{r}(u-\tilde{\alpha}_i)\prod_{j=1}^{s}(u-\tilde{\beta}_j)+\prod_{i=1}^{r}(u-\tilde{\alpha}_i)\sum_{l=1}^{s}\prod_{j\neq l,j=1}^{s}(u-\tilde{\beta}_j)$$

The following expressions are obtained from this expression.

$$\begin{aligned} h(1) &= \prod_{i=1}^{r}(1-\tilde{\alpha}_i)\prod_{j=1}^{s}(1-\tilde{\beta}_j) = \prod_{i=1}^{r}\left(\frac{1}{\alpha_i-\alpha+\gamma}\right)\prod_{j=1}^{s}\left(\frac{1}{\beta_j-\alpha+\gamma}\right)\times\gamma^n \\ h'(\tilde{\alpha}_{i0}) &= \prod_{i\neq i0,i=1}^{r}(\tilde{\alpha}_{i0}-\tilde{\alpha}_i)\prod_{j=1}^{s}(\tilde{\alpha}_{i0}-\tilde{\beta}_j) \\ &= \frac{1}{(\alpha_{i0}-\alpha+\gamma)^{n-1}}\prod_{i\neq i0,i=1}^{r}\frac{\alpha_{i0}-\alpha_i}{\alpha_i-\alpha+\gamma}\prod_{j=1}^{s}\frac{\alpha_{i0}-\beta_j}{\beta_j-\alpha+\gamma}\times\gamma^{n-1} \end{aligned}$$

Moreover, the following are obtained noting $h(\tilde{\alpha}_{i0}) = 0$ because $g_1(u) = (h(u) - h(1))/(u - 1)$.

$$g_1(\tilde{\alpha}_{i0}) = \frac{-h(1)}{\tilde{\alpha}_{i0}-1}, \quad g_1'(\tilde{\alpha}_{i0}) = \frac{h'(\tilde{\alpha}_{i0})(\tilde{\alpha}_{i0}-1)+h(1)}{(\tilde{\alpha}_{i0}-1)^2}$$

On the other hand, $g_1(u)$ is developed as follows.

$$g_1(\tilde{\alpha}_{i0}+\Delta\tilde{\alpha}_{i0}) = g_1(\tilde{\alpha}_{i0}) + g_1'(\tilde{\alpha}_{i0})\Delta\tilde{\alpha}_{i0} + \cdots = 0$$

Therefore, if $\Delta\tilde{\alpha}_{i0}$ is very small, we have

$$\Delta\tilde{\alpha}_{i0} = -\frac{g_1(\tilde{\alpha}_{i0})}{g_1'(\tilde{\alpha}_{i0})} \fallingdotseq \frac{k(c_{i0}+k)^{n-2}}{-\prod_{i\neq i0,i=1}^{r}(c_{i0}-c_i)\prod_{j=1}^{s}(\alpha+c_{i0}\epsilon-\beta_j)+(c_{i0}+k)^{n-1}\epsilon^s}\epsilon^s$$

□

On the other hand, as for the errors of the pseudo localization using the transformations 3,4, we have only to examine the influence when β_l in the linear fractional transformation is shifted slightly.

Proposition 4.2 *Let $\tilde{\alpha}_{i0}$ be a solution of $h(u) = 0$. Let $\tilde{\alpha}_{i0} + \Delta\tilde{\alpha}_{i0}$ be the solution of $g_1(u) = 0$ corresponding to $\tilde{\alpha}_{i0}$. Then we have the following relation*

$$|\Delta\tilde{\alpha}_{i0}| \doteqdot \left|\frac{L_1 k (c_{i0}+k)^{n-2}}{L_2[L_3 \prod_{i\neq i0, i=1}^{r} (\frac{c_{i0}-c_i}{\alpha_i - e\beta_l}) \prod_{j=1}^{s} (\frac{\alpha + c_{i0}\epsilon - \beta_j}{\beta_j - e\beta_l}) + (c_{i0}+k)^{n-1}\epsilon^s]}\right| \epsilon^s$$

Here $c_i\epsilon = \alpha_i - \alpha, c_{i0}\epsilon = \alpha_{i0} - \alpha, \beta_l + \Delta\beta_l = e\beta_l, L_1 = e\beta_l - \alpha - c_{i0}\epsilon, L_2 = \alpha - e\beta_l, L_3 = -(\alpha - e\beta_l - k\epsilon)^{n-1}$.

Especially, we have the following relation when we put $\alpha + c_0\epsilon - \beta_j \doteqdot \alpha - \beta_j$ for a very small ϵ.

$$|\Delta\tilde{\alpha}_{i0}| \doteqdot 2^{s-1}\left|\frac{(1-e)\beta_l}{\alpha - e\beta_l}\right|\left|\frac{(c_{i0}+k)^{n-2}k}{\prod_{i\neq i0, i=1}^{r} (c_{i0}-c_i) \prod_{j=1}^{s} (\alpha - \beta_j)}\right|\epsilon^s$$

[Proof] Omitted.

5 Examples of numerical calculations

Example 1.We consider the following equation.

$$\begin{aligned} f(x) &= \prod_{i=-3, i\neq 0}^{5} (x + \frac{i}{100000000})(x-1001)(x-10001)(x+5001)(x+50001) \\ &= x^{12} + 44000.00000009x^{11} - 3.4506600199604 \times 10^8 x^{10} + \cdots \end{aligned}$$

We apply the linear fractional transformation (transformation 1) to this polynomial. As a result, the following expression is obtained.

$$h(u) = u^{12} - 2.031995781907879u^{11} - 0.8316267067174223u^{10} + \cdots$$

On the other hand, the linear fractional transformation (transformation 3) is applied by using the approximate solutions $-50000, 10000, -5000, 1000$. In this case, the new expression after the linear fractional transformation is as follows.

$$h(u) = u^{12} - 2.031995781903815u^{11} - 0.8316267067140958u^{10} + \cdots$$

We can solve these expression by a numerical calculation. The solutions are transformed by the inverse of the linear fractional transformation with the algebraic computation without making errors. As a result, we have the following table of close solutions of $f(x) = 0$. Here we note that since the imaginary parts are about $10^{-25} \sim 10^{-30}$, we regarded these values as 0 and omitted.

Table 1:Numerical solutions

No.	Transformation 1	Transformation 3
1	-.5000000000000000e-7	-.5000000000000001e-7
2	-.4000000000000000e-7	-.4000000000000000e-7
3	-.3000000000000001e-7	-.3000000000000000e-7
4	-.2000000000000000e-7	-.2000000000000001e-7
5	-.9999999999999999e-8	-.1000000000000000e-7
6	0.1000000000000000e-7	0.9999999999999999e-8
7	0.2000000000000002e-7	0.2000000000000001e-7
8	0.3000000000000002e-7	0.3000000000000001e-7

Next, we apply a pseudo localization procedure by using transformation 1 to this expression $h(u)$. We denote $g_i(u)$ the polynomial obtained after eliminating i solutions. The following polynomials are obtained.

$$
\begin{aligned}
g_1(u) &= u^{11} - 1.031995781907879u^{10} - 1.863622488625301u^9 + \cdot s \quad (\textit{eliminating } u-1) \\
g_3(u) &= u^9 + 0.9680042180921211u^8 - 0.9276140524410589u^7 + \cdots \quad (\textit{eliminating } (u-1)^3) \\
g_4(u) &= u^8 + 1.968004218092121u^7 + 1.040390165651062u^6 + \cdots \quad (\textit{eliminating } (u-1)^4)
\end{aligned}
$$

In the same way, if the transformation 3 is applied, we have the similar expressions.

We can solve these expressions by numerical calculation. The solution is transformed back by the inverse of the linear fractional transformation with the algebraic computation. As a result, we have the following table of close solutions. Here we note that the imaginary part becomes about 10^{-35}. We regarded this value as 0 and omitted.

Table 2:Numerical solutions

Transformation 1			
No.	$g_1(u)$	$g_3(u)$	$g_4(u)$
1	-.5000000000000000e-7	-.5000000000000000e-7	-.4999999997870959e-7
2	-.3999999999999997e-7	-.3999999999999995e-7	-.4000000070306422e-7
3	-.3000000000000003e-7	-.3000000000000002e-7	-.2999999260997139e-7
4	-.2000000000000000e-7	-.2000000000000001e-7	-.2000003277255018e-7
5	-.9999999999999999e-8	-.9999999999999999e-8	-.9999940875907692e-8
6	0.1000000000000001e-7	0.9999999999999996e-8	0.9999856063601379e-8
7	0.1999999999999999e-7	0.2000000000000000e-7	0.2000017998336977e-7
8	0.3000000000000001e-7	0.2999999999999999e-7	0.2999993064138946e-7
Error	10^{-35}	10^{-21}	10^{-14}

Table 3:Numerical solutions

Transformation 3			
No.	$g_1(u)$	$g_3(u)$	$g_4(u)$
1	-.5000000000000000e-7	-.5000000000000004e-7	-.500000000000235e-7
2	-.4000000000000001e-7	-.3999999999999988e-7	-.3999999999922522e-7
3	-.3000000000000000e-7	-.3000000000000008e-7	-.3000000000814309e-7
4	-.2000000000000001e-7	-.1999999999999998e-7	-.1999999996388769e-7
5	-.1000000000000000e-7	-.1000000000000000e-7	-.1000000006514906e-7
6	0.9999999999999999e-8	0.1000000000000000e-7	0.1000000015860604e-7
7	0.2000000000000001e-7	0.1999999999999998e-7	0.1999999980167443e-7
8	0.2999999999999999e-7	0.3000000000000002e-7	0.300000000764256e-7
Error	10^{-40}	10^{-25}	10^{-17}

Example 2. We consider the 50th chebyshev's polynomial $T_{50}(x)$.

$$f(x) = T_{50}(x) = 0$$

$f(x) = 0$ has solutions within the range from -1 to $+1$ and has adjacent solutions at both ends. Therefore, we assume $\alpha = 1, \epsilon = \frac{1}{10}$ and separate the close solutions in this range. This example shows the case where the solutions other than close solutions are very near to close solutions. Applying the linear fractional transformation 1 to $f(x)$, we obtain the polynomial $h(u)$. We write $g_i(u)$ the polynomial obtained by eliminating i solutions. In the following table, we presented only the real parts of the solutions because imaginary parts are about 10^{-30} and are negligible. Exact solutions are as follows. 0.9995065603657316 , 0.9955619646030800 , 0.9876883405951377 , 0.9759167619387474 , 0.9602936856769431 , 0.9408807689542255 , 0.9177546256839811

Table 4:Numerical solutions

$h(u)$	$g_{10}(u)$	$g_{20}(u)$	$g_{25}(u)$
0.9995065603657316	0.9995065603657316	0.9995065684387976	0.9997116722154092
0.9955619646030800	0.9955619646030800	0.9955619577276302	0.9953481236931984
0.9876883405951377	0.9876883405951377	0.9876883413839018	0.9877279255957727
0.9759167619387474	0.9759167619387474	0.9759167619262864	0.9759155042743626
0.9602936856769431	0.9602936856769431	0.9602936856769555	0.9602936895876784
0.9408807689542255	0.9408807689542255	0.9408807689542255	0.9408807689540377
0.9177546256839811	0.9177546256839811	0.9177546256839811	0.9177546256839811

Next, we apply the linear fractional transformation 3 to $f(x)$. We give $\beta_l = -1$(5 times), -0.9(5 times), -0.8(5 times), -0.7(5 times), -0.6(5 times) as approximation solution outside the range.

Table 5:Numerical solutions

$h(u)$	$g_{10}(u)$	$g_{20}(u)$	$g_{25}(u)$
0.9995065603657316	0.9995065603657316	0.9995065603657316	0.9995065607397773
0.9955619646030800	0.9955619646030800	0.9955619646030800	0.9955619642321532
0.9876883405951377	0.9876883405951377	0.9876883405951377	0.987688340654307
0.9759167619387474	0.9759167619387474	0.9759167619387474	0.9759167619370948
0.9602936856769431	0.9602936856769431	0.9602936856769431	0.9602936856769473
0.9408807689542255	0.9408807689542255	0.9408807689542255	0.9408807689542255
0.9177546256839811	0.9177546256839811	0.9177546256839811	0.9177546256839811

References:

[1] H. Kobayashi and H. Suzuki , *The multiplicity of a solution of a system of algebraic equations* , Proc. of the 1992 Intnl Workshop on Math. Mech., pp 53-64.

[2] H. Kobayashi , H. Suzuki and Y. Sakai , *The multiplicity of a solution of a system of algebraic equations II* , Proc. of the 1994 Winter Workshop on Comp. Alg., pp 11-15.

[3] H. Kobayashi , H. Suzuki and Y. Sakai , *Separation of closed roots by linear fraction transformation* , Sushiki-shori vol.2 no.2 pp.2-7 (1993)

EXACT SOLITARY WAVE SOLUTIONS OF NONLINEAR WAVE AND EVOLUTION EQUATIONS

Li Zhibin

Department of Mathematics. Lanzhou University. Lanzhou. 730000. PR China

E-mail: LIL@bepc2.ihep.ac.cn

Abstract A systematic and formal approach is proposed for obtaining travelling wave solutions of nonlinear wave and evolution equations that are essentially of a localized nature. The algorithm is suited to solving nonlinear equations by any symbolic manipulation program. This method is illstrated by constructing exact solutions of a great deal of equations.

1. Introduction

Several methods of obtaining solitary wave solutions have been developed since the inverse scatering technique was established by Gardner *et al* [1]. Searching for exact solutions of nonlinear partial differential equation (PDE) directly became more and more attractive due to the availability of computer symbolic system. Indeed, *Macsyma, Maple, Mathematica* and the like allow one to perform the tedious algebraic calculation.

The original idea behind the direct method goes back to Hiroda [2], who systematically solved large classes of evolution and wave equations by bilinear transformation. Hereman *et al* [3] established a staightforward algorithm for constructing solitary wave solutions from the mixing of the real exponential travelling wave solutions of the underling linear equation. This direct algebraic method has beeen used to construct solitary wave solutions of coupled systems [4]. Seeking travelling wave solutions, one can immediately reduce the nonlinear PDE into a nonlinear ordinary diffenrential equation (ODE) and then use basic ideas from the geometrical theory of ODE. In particular, Painlevé analysis [5] helps in classifying the possible singularities of the solution of the ODE. Partly based on this type of singularity analysis, Takaoka [6] solved the KdV equation with an extra fifth-order dispersive term (Karachwa equation). Restricting himself to hyperbolic-type solutions, Takaoka achieved his goal. With the aid of tanh functions, Lan *et al* [7] obtained travelling wave solutions of a complicated nonlinear wave equation directly. His technique is based on the *a priori* assumption that the solutions can be written in tems of such functions.

In this paper we report a systematized approach for exactly solving nonlinear evolution and wave equations leading to mainly hyperbolic-type solitary waves. The idea of our method that can easily be used in a wider context is to reduce nonlinear evolution and wave equation(s) to polynomial equations. Solving the polynomial equations by means of a mechanization method which has been developed by Wu [8] [9] [10] in recent years yields exact solitary wave solutions of nonlinear evolution and wave equations.

To keep this paper self-contained we will briefly review Wu's method in section 2. In section 3, we present the details of our solution method for the construction of solitary wave solutions of nonlinear wave and evolution equations. In section 4 we exemplify our technique by first applying it to the nonintegrable KdV-Burgers equation and we obtain the exact solutions in the form of superposition of a soliton and a shock. We then treat Kawachara equation which is the KdV equation with an additional fifth-order dispersive term. This example shows we can get more general solution than that using Hereman's method. We consider a variant of Boussinesq equations lastly and show how the extension of the technique leads to the solitary solutions of the coupled wave equations. Finally, we dicuss further possible generalization of the present method, and draw some conclusions.

2. Wu's method of polynomial equations-solving

Let K be a field of characteristic 0 and $X = (x_1, \cdots, x_n)$ be a set of indeterminates to be fixed in what follows. For a polynomial set consisting of non-zero polynomials $P_1, \cdots, P_m$ in the ring $K[X]$ we shall write simply $PS = 0$ for the set of equations $P_1 = 0, \cdots, P_m = 0$.

By polynomial equations-solving is meant the determination of the totality of all possible solutions of $PS = 0$ or zero of PS which will be denoted by $Zero(PS)$. As varity of problems, either from mathematics proper or from other domains, will lead naturally to such polynomial equations $PS = 0$, an effective method of such a polynomial equations-solving will have naturally an immense field of possible applications to problems from mathematics, sinence, and technology. In recent years Wu had developed a mechanization method of solving arbitrary set of polynomial equations in the form of determination of $Zero(PS)$ for an arbitrary polynomial set PS. In this section we briefly review Wu' method.

To begin with, let us first introduce the following all the more important concept of an ascending set .

For a non-constant polynomial P let c be the greatest subscript for which x_c appears actually in P. Let d be then the degree of P considered as a polynomial in x_c. Such a polynomial can then be written in the cannonical form

$$P = I \times x_c^d + \text{ lower degree terms in } x_c$$

with coefficients themselves polynomials in the ring $K[x_1, \cdots, x_{c-1}]$. We shall call c the class and d the degree of P, to be denoted by $cls(P)$ and $deg(P)$ respectively. For a non-zero constant polynomial the class will be defined to be 0 while the degree will be left undefined.

Consider now a polynomial set AS for which the polynomials in AS can be arranged as $A_1, \cdots, A_r$ in canonical forms below:

$$A_i = I_i \times x_{c_i}^{d_i} + \text{ lower degree terms in } x_{c_i}, \quad i = 1, \cdots, r. \qquad (1)$$

where $0 < c_1 < \cdots < c_r \leq n$. In (1) for each i the coefficients of x_{c_i}-powers in A_i are polynomials in $K[x_1, \cdots, x_{c_i-1}]$. We suppose also each polynomial I_i has a degree in

x_j less than $deg(A_j) = d_j$ for any pair (i, j) with $j < i$. In that case we say that the polynomials AS is an ascending set and the coefficient polynomials I_i are initials of the corresponding polynomials A_i.

Given an ascending set defined by (1), a polynomial R for which the degree in x_{c_i} is less than d_i wll be said to be reduced w.r.t. the polynomial A_i, and reduced w.r.t. the ascending set AS if this is so for each i.

Given an arbitrary polynomial G let us divide G successively by the polynomial in AS in the reverse order. This will give rise to an identity of the form

$$I_1^{s_1} \times \cdots \times I_r^{s_r} \times G = \sum_{i=1}^{r} Q_i \times A_i + R. \tag{2}$$

In the identity Q_i, R are polynomials with R reduced w.r.t. AS, and s_i are non-negative integers which may be chosen to be as small as possible. The polynomial R thus obtained from G will be called the remainder of G w.r.t. AS, to be denoted by $R = Remd(G/AS)$. The identity (2) is then called the remainder formula of G w.r.t. AS.

A polynomial set consisting of a single non-zero constant will be called a trivial ascending set. An ascending set of form (1) will then be said to be a non-trivial one. For non-trivial ascending sets the notion of irreducible ones will be defined in a natural way.

For a non-trivial ascending set of the form (1) we may solve $A_1 = 0$ for x_{c_1} in numerical values when K is the rational, the real, or the complex field and x_i, $i < c_1$, are given in numerical value. In general we may simply consider x_{c_1} as a well-determined algebraic function of x_i, $i < c_1$, as defined by $A_1 = 0$. with x_{c_1} thus determined we may then determine x_{c_2} from $A_2 = 0$ either in numerical or as well-determined algebraic function in x_i,$i < c_2$, etc. In this sence $Zero(AS)$ may thus be considered as well-determined. In case AS is a trivial ascending set, then it is clear that $Zero(AS)$ is an empty set.

From the above we see that $Zero(AS)$ and more generally $Zero(AS/G)$ (i.e. the totality of zeros of PS which are not zeros of G) for any ascending set AS and any polynomial G may be consider as well determined. A principle for the solving of the polynomial equations $PS = 0$ or equivalently the determination of $Zero(PS)$ is to determin from the given polynomial set PS a set of ascending sets so that $Zero(PS)$ is closely connected or eventually completely determined from the zero-sets of these ascending sets.

Following the principle we get methods of polynomial equations-solving in the form of following theorems:

Theorem 1 There is an algorithm which permits to determin for any polynomial set PS an ascending set CS in finite number of steps such that

$$Zero(PS) \subset Zero(CS) \tag{3}$$

and

$$Zero(PS) = Zero(CS/J) \cup \bigcup_i Zero(PS \cup I_i). \tag{4}$$

In (4) I_i are the initials of polynomials in CS and J is their product. The ascending set CS is to be called a characteristic set of the polynomial set PS.

Theorem 2 There is an algorithm which permits to determine for any polynomial set PS a finite set of ascending sets AS_i in finite number of steps such that

$$Zero(PS) = \bigcup_i Zero(AS_i/J_i), \tag{5}$$

in which each J_i is the product of all initials of polynomials in the corresponding AS_i.

The formulas (4) and (5) will be called decomposition formulas for $Zero(PS)$. They form the corner-stones on which are based Wu's method of polynomial equation-solving as well as all the applications.

3. The solution method

In this section we discuss the details of the general solution method for exactly solving nonlinear wave and evolution equations leading to mainly hyperbolic-type solitary waves.

The starting point is a nonlinear partial differential equation in 1+1 dimensions with x and t as the space and time coordinates respectively, which describes the dynamical evolution of the wave form $u(x,t)$. The method, which is straightforwardly applicable, goes in the following five steps.

Step 1: A travelling wave solution requires one coordinate: $\xi = k(x - ct)$ and thus $u(x,t) = U(\xi)$, where $U(\xi)$ represents the wave solution, which travels with speed c. It exemplifies a stational wave with a characteristic width k^{-1}. As a consequence, the derivatives are changed into

$$\frac{\partial}{\partial t} \longrightarrow -k\,c\,\frac{d}{d\xi}, \quad \frac{\partial}{\partial x} \longrightarrow k\,\frac{d}{d\xi}. \tag{6}$$

This transforms the given nonlinear PDE in $u(x,t)$ into an ODE in $U(\xi)$. For convenience, we usually take k and $w = k\,c$ as independent parameters.

Step 2: The crucial step now is the introduction of $\eta = \tanh(\xi)$ as a new independent variable. The corresponding derivatives are then changed as follows:

$$\begin{aligned} &\tfrac{d}{d\xi} \longrightarrow (1-\eta^2)\tfrac{d}{d\eta}, \\ &\tfrac{d^2}{d\xi^2} \longrightarrow (1-\eta^2)(-2\eta\tfrac{d}{deta} + (1-\eta^2)\tfrac{d^2}{d\eta^2}), \\ &\tfrac{d^3}{d\xi^3} \longrightarrow -2\eta(1-\eta^2)(-2\eta\tfrac{d}{d\eta} + (1-\eta^2)\tfrac{d^2}{d\eta^2}) \\ &\qquad +(1-\eta^2)(-2\tfrac{d}{d\eta} - 2\eta\tfrac{d^2}{d\eta^2} + (1-\eta^2)\tfrac{d^3}{d\eta^3}). \end{aligned} \tag{7}$$

Higher-order derivatives can be found accordingly. After performing this step, we get, setting $S(\eta) = U(\xi)$, another nonlinear ODE for S.

Step 3: We now have to solve the nonlinear equation in S. There is no general procedure at this stage. Often, the following series expansion is most favorable:

$$S(\eta) = \sum_{n=0}^{m} a_n \eta^n. \tag{8}$$

The parameter m will be found by balancing the linear term(s) of highest order with the nonlinear term(s). The linear term of highest oreder is contained in the highest derivative when applied to $S(\eta)$. This is readily seen using the set of relations(7). The η^2 component of the first relation in (7) leads to the order $(2+m-1) = m+1$, the η^3 component of the first and the η^4 component of the second derivative of the second relation in (7) both lead to the order $m+2$, etc. On the other hand, the nonlinear terms yield multiples of m. The highest-order contribution of the latter must be equal to the former result. In this way the value of m will be determined.

Step 4: Substituting the series expansion (8) into the relevant equation yields recursion relations because the coefficients of η^n $(n = 0, 1, \cdots, m)$ have to vanish. The problem is now compltely reduced to an algebraic one. The remaining unknowns, being the coefficients a_i, the value (nonzero) for k and c will follow from solving a nonlinear system *i.e.* a polynomial equations.

Step 5: A characteristic set of the polynomial equations can be staightforwardly determined by a symbolic manipulation package based on Wu's method, and the unknonws are then calculated. Returning to the original depentent variable u and the independent variables x and t, the exact solitary wave solution(s) of the noninear wave and evolution equation is/are obtained.

Remark: Normally m will be a positive integer so that a closed analytical solution can be obtained. However, we have found some examples where m takes a non-integer value. In this case, we can transfom the original nonlinear wave and evolution equation in u into equation in v by introducing substitution $u = v^{1/m}$ and proceed then as before.

The process described above may not always work, of course, but it certainly works well for a very large class of very interesting nonlinear wave and evolution equations. In the next section, we shall study some examples of various wave equations, to illustrate in detail the method.

4. Examples

4.1 The Korteweg de-Vries-Burgers equation

As a first example to demomstrate the method outlined in section 3, we have chosen a non-integrable nonlinear wave equation, the Korteweg de-Vries-Burgers (KdVB) equation

$$u_t + u_x + u\,u_x - p\,u_{xx} - q\,u_{xxx} = 0. \tag{9}$$

where p and q are dissipative and dispersive coefficients, respectively. The equation (9), quoted as the most simple nonlinear dissapative and dispersive system, models fluid turbulence [11]. Numerical evaluation and qualitative investigation to (9) have been carried out by Canosa *et al* [12] and Bona *et al* [13].

To look for travelling waves of (9), the parameter m is first determined by balancing the order of the linear term of highest order with the nonlinear term: $2m+1 = m+3$ or $m = 2$. Next, the polynomial equations is derived using a simple program in a symbolic algebraic system such as *Mathematica*:

$$\left\{\begin{array}{l} P_1=-wa_1+ka_1+2qk^3a_1+ka_0a-1-2pk^2a_2=0,\\ P_2=2pk^2a_1+ka_1^2-2wa_2+2ka_2+16qk^3a_2+2ka_0a_2=0,\\ P_3=-wa_1+ka_1+8qk^3a_1+ka_0a_1-8pk^2a_2-3ka_1a_2=0,\\ P_4=2pk^2a_1+ka_1^2-2wa_2+2ka_2+40qk^3a_2+2ka_0a_2-2ka_2^2=0,\\ P_5=2qk^2a_1-2pka_2-a_1a_2=0,\\ P_6=12qk^2-a_2=0. \end{array}\right. \tag{10}$$

A characteristic set CS of the polynomial set PS is then easily computed with a package WSolve:

$$\left\{\begin{array}{l} C_1=p^2-100q^2k^2,\\ C_2=-w+k+12qk^3+ka_0,\\ C_3=12pk+5a_1,\\ C_4=12qk^2-a_2 \end{array}\right. \tag{11}$$

In view of the fact that we deal with four equations and five unknowns (a_0, a_1, a_2, k, w) we choose w as a free parameter, and obtain

$$a_0=-1\pm\frac{10qw}{p}-\frac{3p^2}{25q},\quad a_1=\mp\frac{6p^2}{25q},\quad a_2=\frac{3p^2}{25q},\quad k=\pm\frac{p}{10q}.$$

Hence the exact solitary wave solutions of the KdVB equation (9) are

$$\begin{array}{c} u(x,t)=-1\pm\frac{10qw}{p}-\frac{3p^2}{25q}\mp\frac{6p^2}{25q}\tanh(\pm\frac{p}{10q}x-wt)\\ +\frac{3p^2}{25q}\tanh^2(\pm\frac{p}{10q}x-wt) \end{array} \tag{12}$$

with w an arbitrary constant. Taking $w=p(5q-6p^2)/(50pq)$ yields from (12)

$$u^*(x,t)=-\frac{3p^2}{25q}(\mathrm{sech}^2 z\mp 2\tanh z\mp 2) \tag{13}$$

where $z=\pm p[x+(\frac{6p^2}{25q}-1)t]/(10q)$. We can see easily that u^* consist of the kink-type solutions of a Burgers equation and the pulse-type solutions of a KdV equation.

4.2 The Kawachara equation

As an interesting example of a nonlinear equation let us consider the Kawachara equation [14]:

$$u_t+uu_x+pu_{xxx}+qu_{xxxxx}=0. \tag{14}$$

This equation has been also derived and studied by Hunter *et al* [15]. They have given some analysis of the existence of solitary wave solutions and numerical calculations. Using different way, Yamamoto *et al* [16], Hereman *et al* [17] and Huang *et al* [18] have showed that (14) allows a sech^4-type solitary wave solution:

$$u(x,t)=-\frac{105}{169}\frac{p^2}{q}\mathrm{sech}^4\{\sqrt{-\frac{p}{52q}}[x+\frac{36p^2}{169q}t]\} \tag{15}$$

It is our aim to solve (14) by our method. We obtain here a more general solution. In fact, it is easy to judge that $m=4$. Appling the same program as before, we get the polynomial equations:

$$\left\{\begin{array}{l}
P_1=-wa_1-2pk^3a_1+16qk^5a_1+ka_0a_1+6pk^3a_3-120qk^5a_3=0,\\
P_2=ka_1^2-2wa_2-16pk^3a_2+272qk^5a_2+2ka_0a_2+24pk^3a_4\\
\qquad -960qk^5a_4=0,\\
P_3=-wa_1-8pk^3a_1+136qk^5a_1+ka_0a_1-3ka_1a_2+3wa_3\\
\qquad +60pk^3a_3-1848qk^5a_3-3ka_0a_3=0,\\
P_4=ka_1^2-2wa_2-40pk^3a_2+1232qk^5a_2+2ka_0a_2-2ka_2^2\\
\qquad -4ka_1a_3+4wa_4+152pk^3a_4-7744qk^5a_4-4ka_0a_4=0,\\
P_5=-6pk^3a_1+240qk^5a_1-3ka_1a_2+3wa_3+114pk^3a_3-5808qk^5a_3\\
\qquad -3ka_0a_3+5ka_2a_3+5ka_1a_4=0,
\end{array}\right.$$

$$\begin{array}{l}
P_5=-6pk^3a_1+240qk^5a_1-3ka_1a_2+3wa_3+114pk^3a_3-5808qk^5a_3\\
\qquad -3ka_0a_3+5ka_2a_3+5ka_1a_4=0,
\end{array}$$

$$\begin{array}{l}
P_6=-24pk^3a_2+1680qk^5a_2-2ka_2^2-4ka_0a_3+3ka_3^2+4wa_4\\
\qquad +248pk^3a_4-19264qk^5a_4-4ka_0a_4+6ka_2a_4=0,\\
P_7=-k(120qk^4a_1+60pk^2a_3-6600qk^4a_3+5a_2a_3+5a_1a_4\\
\qquad -7a_3a_4)=0,\\
P_8=k(-720qk^4a_2-3a_3^2-120pk^2a_4+19200qk^4a_4-6a_2a_4+4a_4^2)=0,\\
P_9=-7ka_3(360qk^4+a_4)=0,\\
P_{10}=1680k^4q+a_4=0
\end{array} \qquad (16)$$

Computing characteristic set CS of the polynomial set PS with the package WSolve, we obtain:

$$\left\{\begin{array}{l}
C_1=31p^3-56784pq^2k^4+1406080q^3k^6,\\
C_2=pw-104qwk^2+8p^2k^3-1904pqk^5+51584q^2k^7-pka_0\\
\qquad +104qk^3a_0,\\
C_3=a_1,\\
C_4=-280pk^2+29120qk^4-13a_2,\\
C_5=a_3,\\
C_6=1680k^4q+a_4=0
\end{array}\right. \qquad (17)$$

We have seven unknowns and six equations and so may choose w as a free parameter. The other variables are then found. Finally, the solitary wave solution of the Kawachara equation (14) reads:

$$\begin{array}{l}
u(x,t)=-\frac{69p^2}{169q}\mp 2\sqrt{13}\sqrt{-\frac{p}{q}}w+\frac{210p^2}{169q}\tanh^2(\pm\sqrt{-\frac{p}{52q}}x-wt)\\
\qquad -\frac{105p^2}{169q}\tanh^4(\pm\sqrt{-\frac{p}{52q}}x-wt)
\end{array} \qquad (18)$$

with w arbitrary constant.

For $w=-\sqrt{-\frac{p}{52q}\frac{36p^2}{169q}}$ we retrieve the solution (15) from (18).

4.3 The variant of Boussinesq equations

It is generally known that coupled nonlinear equations are difficult to handle. Nevertheless, the following set of equations, can be solved by our solution technique without any difficulty.

We deal with

$$\begin{cases} u_t + v_x + uu_x + s\,u_{xx} + q\,u_{xxt} = 0, \\ v_t + (uv)_x + 2s\,v_{xx} + p\,u_{xxx} = 0. \end{cases} \tag{19}$$

Notice that this set of equation can be seen as the extended Boussinesq's equations which arise in the study of shallow water theory [19]. As a model for water waves, u is a velocity and v is the total depth, so physical solutions ought to have $v > 0$. Here formal properties are considered and the physical restriction $v > 0$ will be ignored.

Now, we show how our method applies to coupled system (19). By using the same steps, we can also make the ansatz:

$$u = \sum_{i=0}^{m} a_i \tanh^i \xi, \qquad v = \sum_{j=0}^{n} b_j \tanh^j \xi \tag{20}$$

with $\xi = kx - wt$, where $a_i(i = 1, 2, \ldots, m), b_j(j = 1, 2, \ldots, n)$, k, w and the integers m and n are parameters to be determined.

The requirment that the highest power of the function $\tanh z$ for the nonlinear terms and that for the linear terms of highest order in each equation of (19) must be equal gives the following relations

$$\begin{cases} 2m + 1 = m + 3, \\ m + n + 1 = m + 3. \end{cases}$$

Thus we know $m = n = 2$. The polynomial equations are then easily obtained.

$$\begin{cases} P_1 = 6wk^2qa_1 + 6k^2sa_2 - 3ka_1a_2 = 0, \\ P_2 = 2ka_2(12wkq - a_2) = 0, \\ P_3 = wa_1 - 8wk^2qa_1 - ka_0a_1 - 8k^2sa_1 + 3ka_1a_2 - kb_1 = 0, \\ P_4 = -wa_1 + 2wk^2qa_1 + ka_0a_1 + 2k^2sa_2 + kb_1 = 0, \\ P_5 = 2k^2sa_1 - ka_1^2 + 2wa_2 - 40wk^2qa_2 - 2ka_0a_2 + 2ka_0a_2 \\ \qquad +2ka_2^2 - 2kb_2 = 0, \\ P_6 = -2k^2sa_1 + ka_1^2 - 2wa_2 + 16wk^2qa_2 + 2ka_0a_2 + 2kb_2 = 0, \\ P_7 = -2k^3pa_1 + ka_1b_0 - wb_1 + ka_0b_1 + 4k^2sb_2 = 0, \\ P_8 = -16k^3pa_2 + 2ka_2b_0 - 4k^2sb_1 + 2ka_1b_1 - 2wb_2 + 2ka_0b_2 = 0, \\ P_9 = -2k^2pa_1 - a_2b_1 + 4ksb_2 - a_1a_2 = 0, \\ P_{10} = 8k^3pa_1 - ka_1b_0 + wb_1 - ka_0b_1 + 3ka_2b_1 - 16k^2sb_2 + 3ka_1b_2 = 0, \\ P_{11} = -24k^3pa_2 - 4ka_2b_2 = 0, \\ P_{12} = 40k^3pa_2 - 2ka_2b_0 + 4k^2sb_1 - 2ka_1b_1 + 2wb_2 - 2ka_0b_2 + 4ka_2b_2 = 0 \end{cases} \tag{21}$$

We find a characteristic set CS of PS in the order $w \prec b_0 \prec a_0 \prec b_1 \prec a_1 \prec a_2 \prec b_2$:

$$\begin{cases} C_1 = s(s^2 - 100w^2q^2), \\ C_2 = 75k^2p^2 + 1200w^2k^2pq^2 + 16k^2ps^2 - 300w^2q^2b_0, \\ C_3 = -25k^2p - 50w^2q + 400w^2k^2q^2 + 2k^2s^2 + 50wkqa_0, \\ C_4 = 6k^2ps + 5wqb_1, \\ C_5 = 12ks - 5a_1, \\ C_6 = 12wkq - a_2, \\ C_7 = 6k^2p + b_2. \end{cases} \quad (22)$$

Determining $Zero(CS)$, we get the solitary wave solutions of the variant of Boussinesq equations (19):

$$\begin{cases} u(x,t) = \pm(\frac{5kp}{s} + \frac{s}{10kq} - \frac{12ks}{5}) \pm \frac{6ks}{5}[1 \pm \tanh(kx \mp \frac{s}{10q}t)]^2, \\ v(x,t) = \frac{25k^2p}{s^2} + 12k^2p - 6k^2p[1 \pm \tanh(kx \mp \frac{s}{10q}t)]^2. \end{cases} \quad (23)$$

where k is an arbitrary constant.

If we choose $p = 0$ in (23), we see that the shocklike wave solutions of Benjamin-Bona-Mahony-Burgers equation

$$u_t + uu_x + su_{xx} + qu_{xxt} = 0 \quad (24)$$

found by Li *et al* [20] is recovered from (23):

$$u(x,t) = \pm\frac{s}{10kq} \mp \frac{12ks}{5} \pm \frac{6ks}{5}[1 \pm \tanh(kx \mp \frac{s}{10q}t)]^2. \quad (25)$$

Taking $s = 0$, the set of equation (19) becomes

$$\begin{cases} u_t + v_x + uu_x + qu_{xxt} = 0, \\ v_t + (uv)_x + pu_{xxx} = 0. \end{cases} \quad (26)$$

Sachs [21] has constructed the rational solutions and the associated higher-order flows of this equations. The sech^2-type solution is readily found from (22):

$$\begin{cases} u(x,t) = \frac{w}{k} + \frac{kp}{2wq} - 8wkq + 12wkq\tanh^2(kx - wt), \\ v(x,t) = \frac{k^2p^2}{4w^2q^2} + 4k^2p - 6k^2p\tanh^2(kx - wt), \end{cases} \quad (27)$$

where k, w are arbitrary constants. Obviously, choice $k = \frac{1}{2\sqrt{-2q}}, w = \pm\frac{\sqrt{p}}{2\sqrt{2q}}$ in (27), produces:

$$\begin{cases} u(x,t) = \pm\frac{3\sqrt{p}}{2\sqrt{-q}}\text{sech}^2[\frac{1}{2\sqrt{-2q}}(x \mp \frac{\sqrt{p}}{\sqrt{-q}}t)], \\ v(x,t) = -\frac{3p}{4q}\text{sech}^2[\frac{1}{2\sqrt{-2q}}(x \mp \frac{\sqrt{p}}{\sqrt{-q}}t)]. \end{cases} \quad (28)$$

Other nonlinear wave and evolution equations such as generalized KdV, combined KdV-mKdV, KS, Klein-Gorden, generalized Fisher-Burgers equations, and so on, and coupled systems of evolution equations or coupeld wave equations, for example, coupled system of KdV equations and nonlinear B-Z reaction system may be found solvable by our method. Detailed calculations will be presented elsewhere.

5. Discussion and conclusion

We have derived travelling wave solutions of solitary wave form using relatively simple mechanization method. We believe that it can be used in a wider context. Of course we are aware of the fact that not all fundamental wave equations can be treated with this method. For example, the nonlinear Schrödinger and the sine-Gordon equations are notable exceptions, although their solutions contain hyperbolic functions as well. Nevertheless, we are convinced that we have achieved our goal of mechanization-mathematics. Finally, we should mention that the method is readily applicable to (2+1)-dimensional equations such as K-P equation. We are also investigating how our method must be modified to treat complex, higher-dimensional or vector nonlinear wave and evolution equations or PDEs of a different kind.

References

[1] Gardner C S, Greene J M, Kruskal M D and Miura R M, Phys.Rev.Lett., **19**(1967):1095-1097

[2] Hiroda R, *Solitons (Topics in Current Physics 17)*, ed P K Bullough and P J Caudry (Berlin: Springer)(1980):157-175

[3] Hereman W Korpel A and Banerjee P P, Wave Motion, **7**(1985):283-290

[4] Hereman W, *Partial Integrable Evolution Equations in Physics*, Proc. of SSTP'90, Les Houches, ed P Conte (Dordrecheht: Kluwer)(1990):585-586

[5] Ince E L, *Ordinary Differential Equations*, (New York: Dover)(1956)

[6] Takaoka M, J.Phys.Soc.Japan, **58**(1989):73-81

[7] Lan H B and Wang K L, J.Phys.A:Math.Gen.**23**(1990):4097-4105

[8] Wu W J, Kexue Tongbao, **31**(1986):1-5

[9] Wu W J, Adv.Comp.Res., **6**(1992):103-138

[10] Wu W J, *Algorithms and Computation*, Proc. of ISAAC'94, Beijing, ed D Z Du and X S Zhang, Springer-Verlag (1994):1-9

[11] Gao G, Sci.Sinica A 30(1985) 457-468

[12] Canosa J and Gazdag J, J.Comp.Phys., **23**(1977):393-403

[13] Bona J L and Schonbck M E, Proc.Roy.Soc.Edinburgh, **101A**(1985):207-226

[14] Kawachara T, J.Phys.Soc.Japan, **33**(1972):260-269

[15] Hunter J K and Scheurle J, Physica D, **32**(1988) 253

[16] Yamamoto Y and Takizawa E I, J.Phys.Soc.Japan, **50**(1981):L1421-1422

[17] Hereman W, Banerjee P P and Korpel A, J.Phys.A:Math.Gen., **19**(1986):607-628

[18] Huang G X, Luo S and Dai X X, Phys.Lett.A **139**(1989):37-39

[19] Whitham G B, *Linear and Nonlinear Waves*,(New York)(1973)

[20] Li Z B and Wang M L, J.Phys.A:Math.Gen., **26**(1993):6027-6029

[21] Sachs R L, Physica D, **3091988)**:1-27

THE IMMERSED INTERFACE METHOD FOR 3D ELLIPTIC EQUATIONS WITH DISCONTINUOUS COEFFICIENTS AND SINGULAR SOURCES

Zhilin Li
Department of Mathematics
University of California at L. A.
Los Angeles, CA 90024, USA
e-mail: zhilin@math.ucla.edu

Abstract: The *Immersed Interface Method* proposed by LeVeque and Li [SIAM J. Numer. Anal. 31, No.4 (1994)] is extended to three dimensional elliptic equations of the form:

$$\nabla \cdot (\beta(x)\nabla u(x)) + \kappa(x)u(x) = f(x).$$

We study the situation in which there is an irregular interface (surface) S contained in the solution domain across which β, κ and f may be discontinuous or even singular. As a result, the solution u will usually be non-smooth or even discontinuous. A finite difference approach with a uniform Cartesian grid is used in the discretization. Local truncation error analysis is performed to estimate the accuracy of the numerical solution.

1 Introduction

Solving elliptic equations with discontinuous coefficients is a fundamental problem in various important applications, for example at the interface between two materials with different diffusion parameters in steady state heat diffusion or electrostatic problems. Such problems also arise in multicomponent flow problems, *e.g.*, the porous media equations used to model the interface between oil and injected fluid in simulations of secondary recovery in oil reservoirs [1], [8] *etc.*

Consider the general three-dimensional problem

$$(\beta u_x)_x + (\beta u_y)_y + (\beta u_z)_z + \kappa(x,y,z)\, u = f(x,y,z), \quad (x,y,z) \in \Omega, \quad (1)$$

in some region Ω, where all the coefficients β, κ, and f may be discontinuous across an interface, which is usually a surface S: $x = x(\mu,\nu)$, $y = y(\mu,\nu)$, $z = z(\mu,\nu)$. The interface S divides the solution domain in two parts which we denote as the + side and − side respectively. By convention the normal direction $\vec{n}$ points toward the + side. When there is a sources or a sink on the interface, a delta function or its derivative singularity would appear in the expression of $f(x,y,z)$, for example

$$f(x,y,z) = \iint_S C(\mu,\nu)\, \delta_3\left(\vec{x} - \vec{X}(\mu,\nu)\right)\, d\mu\, d\nu, \quad (2)$$

where $\vec{x} = (x,y,z)$, $\vec{X}(\mu,\nu) = (x(\mu,\nu), y(\mu,\nu), z(\mu,\nu))$, δ_3 is a three dimensional Dirac function and $C(\mu,\nu)$ is the source strength. Consequently there would be a jump either in the normal derivative or in the solution itself or both. To make the problem (1) well-posed, we assume we know the following jump conditions defined as the difference of the limiting values between the + side and − side at any point of the interface

$$[u] = w(\mu,\nu), \qquad [\beta \frac{\partial u}{\partial n}] = q(\mu,\nu). \quad (3)$$

Such jump conditions can usually be obtained either from physical reasoning or the partial differential equation itself.

In the *immersed interface method* developed in [2], [3] and here, we are able to handle discontinuities and singularities simultaneously. The computational frame is finite difference with a uniform Cartesian grid. One obvious advantage of such a grid is that there is almost no cost for grid generation, and the conventional difference schemes can be used at most grid points (regular). Only those near the interface, which are usually much fewer than the regular grid points, need special attention.

Among other approaches in dealing with discontinuous coefficients problems are *harmonic averaging* [1], [8], *the smoothing technique*, [9], and *flux correction* [5] *etc.* Some of these methods work well in one dimensional problems but are hard to implement in two or three dimensions. Almost all of them are only first order accurate in two or more dimensions.

Another more complicated application with singular sources arises in using Peskin's *immersed boundary method* to solve PDEs with complicated geometry [7]. In Peskin's approach the boundary is immersed in a uniform Cartesian grid and the boundary condition is converted as a singular force term which usually has the form of (2). The immersed boundary method has being widely used for many problems in recent years. In Peskin's immersed boundary method, the discretized delta function is used to deal with the singularity in the source term. However this approach would smooth the solution and hence is only first order accurate. In a different approach, Mayo [6] uses an integral equation to get the necessary information to modify the difference scheme for Poisson and biharmonic equations on irregular regions.

In our approach we make use of the jump conditions to modify the difference scheme at irregular points near the interface. The local truncation error is controlled at these points so that the solution of the difference equations maintains second order accuracy on the entire uniform grid.

Although we only consider elliptic equations here, the immersed interface method has been used for time dependent problems as well such as heat equations [4], the Stokes flow with a moving interface [3] and wave propagations *etc.* Some other problems are currently being investigated.

This paper is an extension of the work by LeVeque and Li [2] to three dimensional problems. Readers are referred to [2] and [3] for more background on the problem, and analysis in one and two dimensions.

2 Interface relations.

At a point (x^*, y^*, z^*) on the interface, we need to use local coordinates to simplify the derivation of the interface relations. The local coordinates transformation from (x, y, z) to (ξ, η, ζ) are chosen so that ξ is parallel to the normal direction of the interface pointing toward the $+$ side. The η- and ζ- axes are in the tangent plane passing through (x^*, y^*, z^*). In the neighborhood of this point, the interface can be expressed as

$$\xi = \chi(\eta, \zeta), \qquad \text{with} \qquad \chi(0,0) = 0, \;\; \chi_\eta(0,0) = 0, \;\; \chi_\zeta(0,0) = 0. \tag{4}$$

Similar to the proof given for two dimensional case in [3], we can prove that in the local coordinates the equation (1) is unchanged, so we will use the same notation for u, w, q, β, κ and f. The jumps w and q in (3) are only defined on the interface and they are functions of η and ζ in the local coordinates.

As we did in [2] and [3], we use the jump conditions and their derivatives as well as the differential equation itself to get the interface relations between the quantities on the two sides of the interface surface. Due to the page limitation we just list the results below:

$$
\begin{aligned}
u^+ &= u^- + w, \\
u_\xi^+ &= \frac{\beta^-}{\beta^+} u_\xi^- + \frac{q}{\beta^+}, \\
u_\eta^+ &= u_\eta^- + w_\eta, \\
u_\zeta^+ &= u_\zeta^- + w_\zeta, \\
u_{\eta\zeta}^+ &= u_{\eta\zeta}^- + u_\xi^- \chi_{\eta\zeta} - u_\xi^+ \chi_{\eta\zeta} + w_{\eta\zeta}, \\
u_{\eta\eta}^+ &= u_{\eta\eta}^- + (u_\xi^- - u_\xi^+)\chi_{\eta\eta} + w_{\eta\eta}, \\
u_{\zeta\zeta}^+ &= u_{\zeta\zeta}^- + (u_\xi^- - u_\xi^+)\chi_{\zeta\zeta} + w_{\zeta\zeta}, \\
u_{\xi\eta}^+ &= \frac{\beta^-}{\beta^+} u_{\xi\eta}^- + \left(u_\eta^+ - \frac{\beta^-}{\beta^+} u_\eta^- \right) \chi_{\eta\eta} + \left(u_\zeta^+ - \frac{\beta^-}{\beta^+} u_\zeta^- \right) \chi_{\eta\zeta} \\
&\quad + \frac{\beta_\eta^-}{\beta^+} u_\xi^- - \frac{\beta_\eta^+}{\beta^+} u_\xi^+ + \frac{q_\eta}{\beta^+}, \\
u_{\xi\zeta}^+ &= \frac{\beta^-}{\beta^+} u_{\xi\zeta}^- + \left(u_\eta^+ - \frac{\beta^-}{\beta^+} u_\eta^- \right) \chi_{\eta\zeta} + \left(u_\zeta^+ - \frac{\beta^-}{\beta^+} u_\zeta^- \right) \chi_{\zeta\zeta} \\
&\quad + \frac{\beta_\zeta^-}{\beta^+} u_\xi^- - \frac{\beta_\zeta^+}{\beta^+} u_\xi^+ + \frac{q_\zeta}{\beta^+}. \\
u_{\xi\xi}^+ &= \frac{\beta^-}{\beta^+} u_{\xi\xi}^- + \left(\frac{\beta^-}{\beta^+} - 1 \right) u_{\eta\eta}^- + \left(\frac{\beta^-}{\beta^+} - 1 \right) u_{\zeta\zeta}^- + \\
&\quad u_\xi^+ \left(\chi_{\eta\eta} + \chi_{\zeta\zeta} - \frac{\beta_\xi^+}{\beta^+} \right) - u_\xi^- \left(\chi_{\eta\eta} + \chi_{\zeta\zeta} - \frac{\beta_\xi^-}{\beta^+} \right) \\
&\quad + \frac{1}{\beta^+} \left(\beta_\eta^- u_\eta^- - \beta_\eta^+ u_\eta^+ \right) + \frac{1}{\beta^+} \left(\beta_\zeta^- u_\zeta^- - \beta_\zeta^+ u_\zeta^+ \right) \\
&\quad - \frac{1}{\beta^+} \left([\kappa] u^- + \kappa^+ [u] \right) + \frac{[f]}{\beta^+} - w_{\eta\eta} - w_{\zeta\zeta}.
\end{aligned}
\tag{5}
$$

3 The difference scheme

At regular grid points, we still use the classical central difference scheme which has a seven-point stencil. The local truncation errors at these grid points are $O(h^2)$. So we will concentrate below on developing difference formulas for the irregular grid points for which the interface cuts through the classical seven-point stencil. Taking a typical irregular grid point, say (x_i, y_j, z_k), we try to develop the modified difference scheme at this point. Because the interface is one dimension lower than the solution domain, the number of irregular grid points will be $O(n^2)$ compared to the total number of grid points, which is $O(n^3)$. We can require the local truncation error for the difference scheme at irregular grid points to be $O(h)$ without affecting the second order accuracy globally. Let us write the difference scheme as follows:

$$\sum_m \gamma_m \, u_{i+i_m, j+j_m, k+k_m} + \kappa_{ijk} \, u_{ijk} = f_{ijk} + C_{ijk}, \tag{6}$$

where i_m, j_m, k_m may be $0, \pm 1, \pm 2, \cdots$. Of course we want the number of grid points involved to be as few as possible. So first we need to determine the stencil, and then find the coefficients γ_m and the correction term C_{ijk} for the given stencil.

The analysis is similar to the two dimensional case as presented in [2] and [3] but more complicated in three dimensions. We take a point (x^*, y^*, z^*) on the interface surface near (x_i, y_j, z_k) and use local coordinates (ξ, η, ζ) at (x^*, y^*, z^*). For the elliptic equation the coefficients γ_m should be of order $O(1/h^2)$. So if we expand $u_{i+i_m, j+j_m, k+k_m}$ in the difference scheme about the origin of the local coordinates from each side of the surface S, we need to match up to second derivatives to guarantee that the local truncation error is $O(h)$. Using the ten interface relations (5) to eliminate quantities at the $(+)$ side of the interface, the Taylor expansion of (6) will then contain u^-, u^-_ξ, u^-_η, u^-_ζ, $u^-_{\xi\xi}$, $u^-_{\eta\eta}$, $u^-_{\zeta\zeta}$, $u^-_{\xi\eta}$, $u^-_{\xi\zeta}$, and $u^-_{\eta\zeta}$. To match them we need altogether ten grid points to get ten equations for the γ_ms. Thus we need to find three additional points besides the standard seven-point stencil. The three additional grid points can be taken from any of the twenty grid points $(i \pm 1, j \pm 1, k \pm 1)$, $(i \pm 1, j \pm 1, k)$, $(i \pm 1, j, k \pm 1)$, $(i, j \pm 1, k \pm 1)$.

Once we have determined the stencil we need to find the coefficients of the difference scheme. The local truncation error of the difference scheme is

$$T_{ijk} = \sum_m \gamma_m \, u_{i+i_m, j+j_m, k+k_m} + \kappa_{ijk} \, u_{ijk} - f_{ijk} - C_{ijk}. \tag{7}$$

We will force T_{ijk} to be $O(h)$ by choosing the coefficients γ_ms. To get the equations for those coefficients we use Taylor expansion of (6) about (x^*, y^*, z^*), the origin of the local coordinates. If the grid point (x_i, y_j, z_k) is on the $(-)$ side, we will get

$$\begin{aligned} T_{ijk} \;=\; & a_1 u^- + a_2 u^+ + a_3 u^-_\xi + a_4 u^+_\xi + a_5 u^-_\eta + a_6 u^+_\eta + a_7 u^-_\zeta + a_8 u^+_\zeta \\ & + a_9 u^-_{\xi\xi} + a_{10} u^+_{\xi\xi} + a_{11} u^-_{\eta\eta} + a_{12} u^+_{\eta\eta} + a_{13} u^-_{\zeta\zeta} \\ & + a_{14} u^+_{\zeta\zeta} + a_{15} u^-_{\xi\eta} + a_{16} u^+_{\xi\eta} + a_{17} u^-_{\xi\zeta} + a_{18} u^+_{\xi\zeta} \\ & + a_{19} u^-_{\eta\zeta} + a_{20} u^+_{\eta\zeta} + \kappa^- u^- - f^- - C_{ijk} + O(h). \end{aligned} \tag{8}$$

The coefficients a_j are linear combination of the γ's as in the equation (3.26) of [2]. They depend only on the position of the stencil relative to the interface but independent of the functions u, κ and f.

Using the interface relations (5), and rearranging (8) we can express T_{ijk} in terms of the quantities from $-$ side. To make T_{ijk} to be $O(h)$ we should set

$$a_1 - a_{10} \frac{[\kappa]}{\beta+} + a_2 = 0, \tag{9}$$

$$\begin{aligned} & a_3 - a_{10}\left(\chi_{\eta\eta} + \chi_{\zeta\zeta} - \frac{\beta^-_\xi}{bm}\right) + a_{12}\,\chi_{\eta\eta} + a_{14}\,\chi_{\zeta\zeta} + a_{16}\frac{\beta^-_\eta}{\beta+} \\ & + a_{18}\frac{\beta^-_\zeta}{\beta+} + a_{20}\,\chi_{\eta\zeta} + \frac{\beta-}{\beta+}\left\{ a_4 + a_{10}\left(\chi_{\eta\eta} + \chi_{\zeta\zeta} - \frac{\beta^+_\xi}{\beta+}\right)\right. \\ & \left. -a_{12}\,\chi_{\eta\eta} - a_{14}\,\chi_{\zeta\zeta} - a_{16}\frac{\beta^+_\eta}{\beta+} - a_{18}\frac{\beta^+_\zeta}{\beta+} - a_{20}\,\chi_{\eta\zeta}\right\} = \beta^-_\xi, \end{aligned} \tag{10}$$

$$\begin{aligned} & a_5 + a_{10}\frac{\beta^-_\eta}{\beta+} - a_{16}\frac{\beta-}{\beta+}\chi_{\eta\eta} - a_{18}\frac{\beta-}{\beta+}\chi_{\eta\zeta} \\ & + a_6 - a_{10}\frac{\beta^+_\eta}{\beta+} + a_{16}\,\chi_{\eta\eta} + a_{18}\,\chi_{\eta\zeta} = \beta^-_\eta, \end{aligned} \tag{11}$$

$$a_7 + a_{10}\frac{\beta^-_\zeta}{\beta+} - a_{16}\frac{\beta-}{\beta+}\chi_{\eta\zeta} - a_{18}\frac{\beta-}{\beta+}\chi_{\zeta\zeta}$$

$$
\begin{aligned}
+ a_8 - a_{10} \frac{\beta_\zeta^+}{\beta^+} + a_{16} \chi_{\eta\zeta} + a_{18} \chi_{\zeta\zeta} &= \beta_\zeta^-, & (12) \\
a_9 + a_{10} \frac{\beta^-}{\beta^+} &= \beta^-, & (13) \\
a_{11} + a_{12} + a_{10} \left(\frac{\beta^-}{\beta^+} - 1 \right) &= \beta^-, & (14) \\
a_{13} + a_{14} + a_{10} \left(\frac{\beta^-}{\beta^+} - 1 \right) &= \beta^-, & (15) \\
a_{15} + a_{16} \frac{\beta^-}{\beta^+} &= 0, & (16) \\
a_{17} + a_{18} \frac{\beta^-}{\beta^+} &= 0, & (17) \\
a_{19} + a_{20} &= 0. & (18)
\end{aligned}
$$

This is a system of ten equations with ten variables. We can solve this system to get the coefficients γ_m of the difference scheme at this particular irregular grid point. Once we know the γ_m, we know the a_i as well, so we can calculate the correction term from the following:

$$
\begin{aligned}
C_{ijk} = \; & a_{10} \left(\frac{[f]}{\beta^+} - \frac{\kappa^+ [u]}{\beta^+} - w_{\eta\eta} - w_{\zeta\zeta} \right) + a_{12} w_{\eta\eta} + a_{14} w_{\zeta\zeta} \\
& + a_{16} \frac{q_\eta}{\beta^+} + a_{18} \frac{q_\zeta}{\beta^+} + a_{20} w_{\eta\zeta} + a_2 [u] \\
& + \frac{1}{\beta^+} \left\{ a_4 + a_{10} \left(\chi_{\eta\eta} + \chi_{\zeta\zeta} - \frac{\beta_\xi^+}{\beta^+} \right) - a_{12} \chi_{\eta\eta} \right. \\
& \left. - a_{14} \chi_{\zeta\zeta} - a_{16} \frac{\beta_\eta^+}{\beta^+} - a_{18} \frac{\beta_\zeta^+}{\beta^+} - a_{20} \chi_{\eta\zeta} \right\} q \qquad (19) \\
& + \left(a_6 - a_{10} \frac{\beta_\eta^+}{\beta^+} + a_{16} \chi_{\eta\eta} + a_{18} \chi_{\eta\zeta} \right) w_\eta \\
& + \left(a_8 - a_{10} \frac{\beta_\zeta^+}{\beta^+} + a_{16} \chi_{\eta\zeta} + a_{18} \chi_{\zeta\zeta} \right) w_\zeta .
\end{aligned}
$$

If the grid point is on $(+)$ side, there are two ways to deal with it. The first one is to modify the correction term C_{ijk} and the linear system (9)–(18)

slightly. Use the following relation

$$\begin{aligned} \kappa^+ u^+ &= \kappa^- u^- + \kappa^+ u^+ - \kappa^- u^- \\ &= \kappa^- u^- + \kappa^+ [u] + [\kappa] u^-, \end{aligned} \tag{20}$$

and let the difference scheme at this irregular (x_i, y_j, z_k) be:

$$\sum_m \hat{\gamma}_m u_{i+i_m, j+j_m, k+k_m} + \kappa_{ijk} u_{ijk} = f_{ijk} + \hat{C}_{ijk}. \tag{21}$$

Then $\hat{\gamma}_m$ still satisfy equations (10)–(18). Now the first equation becomes

$$a_1 - a_{10} \frac{[\kappa]}{\beta +} + a_2 = -[\kappa], \tag{22}$$

and the correction term $\hat{C}_{ijk}$ is

$$\hat{C_{ijk}} = C_{ijk} + \kappa^+ [u] - [f]. \tag{23}$$

The other way is simply to reverse the roles of the two sides $(+)$ and $(-)$ in the discussion above.

The immersed interface method discussed above has several nice properties. It appears to be second order accurate and can capture the sharpness if the solution is non-smooth or discontinuous. It can handle very general problems without much difficulty as we can see from an example below. Furthermore, if no interface is present then we will revert to the classical central difference scheme.

We have tested several examples. Although we can not take very fine grids to test the second order convergence due to the size limitation in three-dimensions, we do observe good numerical results. Below we give one test example.

Example 1 *We consider a problem in three dimensions with discontinuous coefficients as well as the singular sources. The equation is defined on the cube:* $-1 \le x, y, z, \le 1$ *and has the form*

$$(\beta u_x)_x + (\beta u_y)_y + (\beta u_z)_z + \kappa u = f,$$

where

$$\beta(x,y,z) = \begin{cases} 1 + x^2 + y^2 + z^2 & \textit{if } x^2 + y^2 + z^2 \leq \frac{1}{4} \\ 1 & \textit{if } x^2 + y^2 + z^2 > \frac{1}{4}, \end{cases}$$

$$f(x,y,z) = \begin{cases} 6 + 11\,(x^2 + y^2 + z^2) \\ \dfrac{1}{x^2 + y^2 + z^2} - \dfrac{1}{\sqrt{x^2 + y^2 + z^2}} - \log\left(2\sqrt{x^2 + y^2 + z^2}\right), \end{cases}$$

$$\kappa(x,y,z) = \begin{cases} 1 & \textit{if } x^2 + y^2 + z^2 \leq \frac{1}{4} \\ -1 & \textit{if } x^2 + y^2 + z^2 > \frac{1}{4}. \end{cases}$$

Dirichlet boundary conditions are determined from the exact solution

$$u(x,y,z) = \begin{cases} x^2 + y^2 + z^2 & \textit{if } x^2 + y^2 + z^2 \leq \frac{1}{4} \\ \dfrac{1}{\sqrt{x^2 + y^2 + z^2}} + \log\left(2\sqrt{x^2 + y^2 + z^2}\right) & \textit{if } x^2 + y^2 + z^2 > \frac{1}{4}. \end{cases}$$

Table 1 lists the local truncation and global errors defined as

$$\| E_n \|_\infty = \max_{i,j,k} | u(x_i, y_j, z_k) - u_{ijk} |, \qquad \| T_n \|_\infty = \max_{i,j,k} | T_{ijk} |,$$

where n is the mesh size, $u(x_i, y_j, z_k)$ is the exact solution and u_{ijk} is the computed solution. The results show that our numerical method is about second order accurate as we refine the mesh. The ratio in column 3 approaches 4 meaning that $\| E_n \|$ is $O(h^2)$. And the ratio in column 5 approaches 2 meaning that $\| T_n \|$ is $O(h)$ as we have predicted.

Table 1: Numerical results for three dimensional Example 1.

n	$\| E_n \|_\infty$	ratio	$\| T_n \|_\infty$	ratio
20	9.2824×10^{-3}		1.1675	
40	2.8176×10^{-3}	3.2945	0.6587	1.7524
80	7.1043×10^{-4}	3.9656	0.3757	1.7728

Acknowledgments: The author is indebted to Professor R. J. LeVeque for his continuous support and helpful discussions during the course of this work. This work was supported by US NFS Grants DMS-9204329, DMS-9303404, DOE Grant DE-FG06-93ER25181, and URI grant #N00014092-J-1890 from ARPA.

References

[1] J. B. Bell, C. N. Dawson, and G. R. Shubin, *An unsplit, higher order Godunov method for scalar conservation laws in multiple dimensions*, J. Comput. Phys., 74 (1988), pp. 1–24.

[2] R. J. LeVeque and Z. Li, *The immersed interface method for elliptic equations with discontinuous coefficients and singular sources*, SIAM J. Num. Anal., 31 (1994), pp. 1019-1044.

[3] Z. Li, *The Immersed Interface Method – A Numerical Approach for Partial Differential Equations with Interfaces*, Ph. D. Thesis, University of Washington, 1994.

[4] Z. Li and A. Mayo, *ADI Methods for Heat Equations with Discontinuities along an Arbitrary Interface*, Proceedings of Symposia in Applied Mathematics, 48 (1994), pp. 311-315.

[5] R. J. MacKinnon and G. F. Carey, *Analysis of material interface discontinuities and superconvergent fluxes in finite difference theory*, J. Comput. Phys., 75 (1988), pp. 151–167.

[6] A. Mayo, *The fast solution of Poisson's and the biharmonic equations on irregular regions*, SIAM J. Num. Anal., 21 (1984), pp. 285–299.

[7] C. S. Peskin, *Numerical analysis of blood flow in the heart*, J. Comput. Phys., 25 (1977), pp. 220–252.

[8] G. R. Shubin and J. B. Bell, *An analysis of the grid orientation effect in numerical simulation of miscible displacement*, Comp. Meth. Appl. Mech. Eng., 47 (1984), pp. 47–71.

[9] M. Sussman and P. Smereka and S. Osher, *A level set approach for computing solutions to incompressible two-phase flow*, UCLA Computational ans Applied Mathematics Report 93-18.

COMPUTER ALGEBRA APPLICATION TO THE DISTRIBUTION THEORY OF MULTIVARIATE STATISTICS

Naoto NIKI[1)], Shigekazu NAKAGAWA[2)] and Hiroki HASHIGUCHI[1)]

[1)]Department of Management Science, Science University of Tokyo,
Shinjuku-ku, Tokyo 162, JAPAN
niki@ms.kagu.sut.ac.jp, hiro@ms.kagu.sut.ac.jp

[2)]Department of Computer Science and Mathematics,
Kurashiki University of Science and the Arts, Kurashiki 712, JAPAN
nakagawa@soft.kusa.ac.jp

Summary. The authors have been using computer algebra to obtain the asymptotic expansions for probability integrals and percentiles of the distributions of multivariate statistics. The utility of higher order asymptotic approximation in distribution theory of statistics is shown by discussing the validity of asymptotic expansions, a class of normalizing transformations to suppress the spurious oscillations in parts of the tails (known as "tail difficulty") and the symbolic-algebraic procedures in derivation of the expansions. Change of bases of the ring of symmetric functions are applied on the expressions of statistics for obtaining the moments (or their asymptotic expansions) of the distributions of the statistics.

1. Introduction

Intensive works have been made, concerning the distributions of statistics in multivariate analysis, on the derivation of both exact and approximate distributions. See, *e.g.*, Muirhead [Mui], Anderson [And] and Siotani, Hayakawa and Fujikoshi [SHF] for an extensive bibliography in this area of distribution theory. It is to be noticed that we can rarely obtain an *explicit*, *exact* and *computable* formula to the probability integrals and percentiles for the distribution of a statistic, no matter whether its probability density is known and seems tractable or not. The above fact gives us a strong motivation to work at the tasks originated in approximate theory of sample distributions.

Many important statistics have symmetric property, which means that they remains unchanged however observations may be exchanged or permuted; besides, most of them are defined as appropriately smooth functions of sample moments. These are so-called *symmetric statistics*. Multivariate symmetric statistics appear in a variety of contexts; see for instance, von Mises [voM] and Denker [Den]. However, most of them are concerned only with the asymptotic behavior of the statistics, although the work of Bhattacharya and Ghosh [BhG] and Niki and Konishi [NK1] have made the use of asymptotic expansions attractive for obtaining accurate approximation to the sample distributions of practical size. The chief obstacle seems to lie in heaps of tedious algebraic calculations; see, *e.g.*, Withers [Wit].

The authors have worked on the development of a software toolkit to transfer the load on human heads and hands to computers. A subset of the tools which consists of a collection of formulas in REDUCE language (Hearn [Hea]), which give asymptotic expansions, is actually used and shows its ability in this research field; see, *e.g.*, Niki and Konishi [NK1][NK2], Konishi, Niki and Gupta [KNG] and Nakagawa and Niki [NN2]~[NN6].

2. Preliminaries — Multidimensionally symmetric polynomials

Let $x_{11}, \ldots, x_{m1}, \ldots, x_{1n}, \ldots, x_{mn}$ be $N = mn$ independent variables and consider the matrix

$$\boldsymbol{X} = (\,\boldsymbol{x}_1 \quad \ldots \quad \boldsymbol{x}_n\,) = \begin{pmatrix} x_{11} & \ldots & x_{1n} \\ \vdots & & \vdots \\ x_{m1} & \ldots & x_{mn} \end{pmatrix}.$$

The symmetric group S_n acts on the ring $\mathbf{Z}[\boldsymbol{X}] = \mathbf{Z}[x_{11}, \ldots, x_{m1}, \ldots, x_{1n}, \ldots, x_{mn}]$ of polynomials in N independent variables with integer coefficients by permuting the column vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ of $\boldsymbol{X}$. We call a polynomial m(-dimensionally)-symmetric if it is invariant under this action.

The family of m-symmetric polynomials form a subring $\Lambda = \Lambda(m, n)$ of the ring of ordinary symmetric functions. The ring Λ is a graded ring

$$\Lambda = \bigoplus_{\boldsymbol{d} \succ \boldsymbol{0}} {}^{\boldsymbol{d}}\Lambda,$$

where each $\boldsymbol{d}$ is a column vector $\boldsymbol{d} = (d_1 \ldots d_m)' \in \mathrm{N}_0^m$ of non-negative integers; and the free module ${}^{\boldsymbol{d}}\Lambda$ consists of the homogeneous m-symmetric polynomials of degree d_1 with respect to $x_{11}, \ldots, x_{1n}$, of degree d_2 with respect to $x_{21}, \ldots, x_{2n}$, and so on, including the zero polynomial. Any linear ordering $\succ$ of vectors is acceptable, if it satisfies the following conditions for any $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathrm{N}_0^m$:

$$\boldsymbol{a} \succ \boldsymbol{b} \Rightarrow \boldsymbol{a} + \boldsymbol{c} \succ \boldsymbol{b} + \boldsymbol{c} \quad \text{and} \quad \boldsymbol{a} \neq \boldsymbol{0} \Rightarrow \boldsymbol{a} \succ \boldsymbol{0}.$$

We shall also use the symbol $\succeq$ for denoting "$\succ$ or $=$".

Corresponding to the ordinary partitions, we define m-partitions $\boldsymbol{K} = (\boldsymbol{k}_1, \boldsymbol{k}_2, \ldots)$ to be the sequences of column vectors $\boldsymbol{k}_1, \boldsymbol{k}_2, \ldots \in \mathrm{N}_0^m$ in descending order, say $\boldsymbol{k}_1 \succeq \boldsymbol{k}_2 \succeq \ldots$, provided that any m-partition contains only finitely many, say $\ell(\boldsymbol{K})$, non-zero vectors. For the sake of convenience, we shall not distinguish between two such sequences of vectors which differ only by a string of zero vectors at the tails; and, if need be, we shall also use the notation $\boldsymbol{k} = (\boldsymbol{k}_1, \boldsymbol{k}_2, \ldots) = (\tilde{\boldsymbol{k}}_1^{\pi_1}, \ldots, \tilde{\boldsymbol{k}}_s^{\pi_s})$, where $\pi_1, \ldots \pi_s$ are the multiplicities of the distinct non-zero parts in $\boldsymbol{K}$, $\tilde{\boldsymbol{k}}_1 \succ \ldots \succ \tilde{\boldsymbol{k}}_s \succ \mathbf{0}$, respectively, and which satisfy $\pi_1 + \cdots + \pi_s = \ell(\boldsymbol{K})$. For $n \geq \ell = \ell(\boldsymbol{K})$, we can regard $\boldsymbol{K}$ as an $m \times n$ matrix and use the same symbol $\boldsymbol{K}$ for denoting

$$\boldsymbol{K} = (\boldsymbol{k}_1 \ldots \boldsymbol{k}_\ell \overbrace{\mathbf{0}_m \ldots \mathbf{0}_m}^{n-\ell}).$$

The weight vector $w(\boldsymbol{K})$ of the m-partition $\boldsymbol{K}$ is defined as $w(\boldsymbol{K}) = \boldsymbol{K}\mathbf{1}_n$.

For any $m \times n$ matrix $\boldsymbol{C} = (c_{ij})$ of non-negative integers, we write as $\boldsymbol{X}^{\boldsymbol{C}}$ the monomial

$$\boldsymbol{X}^{\boldsymbol{C}} = x_{11}^{c_{11}} \cdots x_{m1}^{c_{m1}} \cdots x_{1n}^{c_{1n}} \cdots x_{mn}^{c_{mn}}.$$

Let $\boldsymbol{K}$ be any m-partition of length $\ell = \ell(\boldsymbol{K}) \leq n$. The polynomial $A_{\boldsymbol{K}}$ in $\boldsymbol{X}$

$$A_{\boldsymbol{K}} = \frac{1}{(n-\ell)!} \sum_{\omega \in S_n} (\omega \boldsymbol{X})^{\boldsymbol{K}}$$

is clearly m-symmetric and is called the augmented (monomial) m-symmetric polynomial with argument $\boldsymbol{K}$. If $\boldsymbol{K} = (\tilde{\boldsymbol{k}}_1^{\pi_1}, \ldots, \tilde{\boldsymbol{k}}_s^{\pi_s})$, then $A_{\boldsymbol{K}}$ has the integral factor $\pi_1! \ldots, \pi_s!$ and can be written as

$$A_{\boldsymbol{K}} = \pi_1! \ldots, \pi_s!\, M_{\boldsymbol{K}}.$$

The m-symmetric polynomial $M_{\boldsymbol{K}}$ consists of distinct monic monomials each other in $\boldsymbol{x}$ and is called a monomial m-symmetric polynomial.

The product

$$P_{\boldsymbol{K}} = M_{(\boldsymbol{k}_1)} \cdots M_{(\boldsymbol{k}_n)} = A_{(\boldsymbol{k}_1)} \cdots A_{(\boldsymbol{k}_n)}$$

is also m-symmetric, where each $M_{(\boldsymbol{r})}$ has a one-part m-partition as its argument and is given by the summation

$$M_{(\boldsymbol{r})} = \sum_{i=1}^{n} \boldsymbol{x}_i^{\boldsymbol{r}} = \sum_{i=1}^{n} x_{1i}^{r_1} \cdots x_{mi}^{r_m}$$

which corresponds the *power sum* in case of $m = 1$. It is not so difficult to prove the following proposition;

PROPOSITION 1

(a) $\{M_{\boldsymbol{K}} \mid \ell(\boldsymbol{K}) \leq n,\ w(\boldsymbol{K}) = \boldsymbol{d}\}$ *forms a* $\mathbf{Z}$*-basis of* ${}^{\boldsymbol{d}}\Lambda$, *and*

(b) $\{P_{\boldsymbol{K}} \mid \ell(\boldsymbol{K}) \leq n,\ w(\boldsymbol{K}) = \boldsymbol{d}\}$ *forms a* $\mathbf{Q}$*-basis of the module* ${}^{\boldsymbol{d}}\Lambda_{\mathbf{Q}} = {}^{\boldsymbol{d}}\Lambda \otimes_{\mathbf{Z}} \mathbf{Q}$.

Other bases of ${}^{d}\Lambda$ or ${}^{d}\Lambda_{\mathbf{Q}}$ may be worth to be considered; for example, the elementary m-symmetric polynomials, the complete m-symmetric polynomials, and so forth. We do not, however, make any further discussion here.

3. Deriving the moments of symmetric statistics

3.1. A family $\mathcal{T}_s$ of statistics with symmetric property

Let $\{ \boldsymbol{X}_i \}$ $(i = 1, 2, \cdots, n)$ be a finite sequence of independent and identically distributed m-dimensional random vectors, and let $f_1, f_2, \cdots, f_k$ be polynomials of the elements of $\boldsymbol{X}_i$. Consider the family

$$\mathcal{T}_s = \{ H(\boldsymbol{Z}) \} \tag{1}$$

of symmetric statistics, where

$$\begin{aligned} \boldsymbol{Z} &= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Y}_i, \qquad \boldsymbol{Y}_i = (f_1(\boldsymbol{X}_i), \cdots, f_k(\boldsymbol{X}_i))', \\ \boldsymbol{\mu} &= \mathrm{E}\,\boldsymbol{Y}_i, \qquad \mathrm{E}\,|\boldsymbol{Y}_i|^s < \infty \quad (s \geq 3) \end{aligned} \tag{2}$$

and a real valued function H on $\mathbf{R}^k$ is s-times continuously differentiable in a neighborhood of $\boldsymbol{\mu}$. Note that this family $\mathcal{T}_s$ contains a wide variety of statistics given as "smooth" functions of sample moments including all the moment statistics.

Let W_n be the semi-standardized random variable

$$W_n = \sqrt{n}\,[H(\boldsymbol{Z}) - H(\boldsymbol{\mu})] \tag{3}$$

derived from $H(\boldsymbol{Z})$ then W_n has a limiting normal distribution with mean zero and variance

$$\sigma^2 = \sum_{i,j=1}^{k} v_{i,j} l_i l_j \tag{4}$$

where $V = (v_{i,j})_{i,j}$ is the dispersion matrix of $\boldsymbol{Y}_i$ and

$$l_{j_1,\cdots,j_p} = (D_{j_1} \cdots D_{j_p} H)(\boldsymbol{\mu}) = \left.\frac{\partial H(\boldsymbol{Z})}{\partial z^{(j)}}\right|_{\boldsymbol{Z}=\boldsymbol{\mu}} \tag{5}$$

$$1 \leq j_1, \cdots, j_p \leq k; \;\; \boldsymbol{Z} = (z^{(1)}, \cdots, z^{(k)}).$$

It is common practice among applied statisticians to calculate *approximate moments* of W_n by expanding around $\boldsymbol{\mu}$, keeping a certain number of terms, raising to appropriate power and taking expectations term by term. Those approximate moments are placed into the formal asymptotic expansions for probability integrals and for percentiles of the distribution of W_n. This is the so-called *delta method.* Bhattacharya and Ghosh [BhG] has discussed the asymptotic behavior of W_n and proved the validity of the resulting formulas given through the delta method.

4. Normalizing transformation

Niki and Konishi [NK3] have made discussions on the phenomenna so-called "tail difficulty" [Wal] that spurious oscillations appear in parts of the tails of higher order asymptotic expansions and shown the effects of a class of normalizing transformations in suppressing such oscillations.

The key to unlock the difficulty lies in finding a transformation of variate with which the lowest order term (of order $\frac{1}{\sqrt{n}}$, say) of the third cumulant vanishes. This is nothing but an asymptotic symmetrization of density or, in other words, correction of skewness which has been already discussed by several authors (see, *e.g.*, Konishi [KNG]) from some different points of view.

Niki and Konishi [NK3] shows that

1. With asymptotically "corrected" skewness, a remarkable number of highest order Hermite polynomials which are highly oscillatory do not appear in the asymptotic expansions.

2. Correction can be realized by finding a function which satisfies a simple differential equation of the second order at least for the asymptotic mean.

Fisher's z-transformation for sample correlation coefficients and the Wilson-Hilferty transformation for ξ^2 variates may be most popular among such functions.

4.1. Algorithm for taking expectation

When we assign $\boldsymbol{x}_i = \boldsymbol{X}_i$ $(i = 1, \ldots, n)$, one may notice that the formal Taylor series expansion of $H(\mathbf{Z})$ with respect to n^{-1} can be given in terms of P's, so that the approximate moments of W_n can be obtained by evaluating the expectations $E(P)$'s.

For this purpose, the following lemmas have much importance:

LEMMA 1 *Let* $\mu_{\boldsymbol{r}}$ *denote the population (product) moment* $\mu_{\boldsymbol{r}} = E(x_{1i}^{r_1} \cdots x_{mi}^{r_m})$ *for* $\boldsymbol{r} = (r_1, \ldots, r_m) \in \mathrm{N}_0^m$*, then*

$$E(A_{\boldsymbol{K}}) = n(n-1)\cdots(n-\ell)\,\mu_{\boldsymbol{k}_1}\cdots\mu_{\boldsymbol{k}_\ell},$$

where $\ell = \ell(\boldsymbol{K})$.

LEMMA 2 *Disregarding the order of parts in m-partitions, it holds, for* $\ell = \ell(\boldsymbol{K}) < n$*, that*

$$P_{(\boldsymbol{r})}A_{(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_\ell)} = A_{(\boldsymbol{r},\boldsymbol{k}_1,\ldots,\boldsymbol{k}_\ell)} + A_{(\boldsymbol{r}+\boldsymbol{k}_1,\boldsymbol{k}_2,\ldots,\boldsymbol{k}_\ell)} + \cdots + A_{(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_{\ell-1},\boldsymbol{r}+\boldsymbol{k}_\ell)}.$$

From Lemma 2, noting $P_{(\boldsymbol{r})} = A_{(\boldsymbol{r})}$ for any one-part m-partition $(\boldsymbol{r})$, we can compose an algorithm to get the expression of a P in terms of A's with integer coefficients. The sketch of the algorithm is as follows:

ALGORITHM 1 (PtoA)

Input: $\boldsymbol{K}$: *m-partition designating a P.*

Output: *weighted sum of A's.*

Step 1. **If** $\ell(\boldsymbol{K}) = 1$ **then** *return* $A_{\boldsymbol{K}}$ *as output.*

Step 2. $U \leftarrow P_{(\boldsymbol{k}_1)}$; $i \leftarrow 2$.

Step 3. $U \leftarrow P_{(\boldsymbol{k}_i)} \times U$ *by using Lemma 2.*

Step 4. $i \leftarrow i + 1$; **if** $i > \ell(\boldsymbol{K})$ **then** *return U as output* **else** *go back to Step 3.*

It is worth to note that Algorithm PtoA for transformation from P to A's yields the moments (and also the cumulants) of W_n represented in the population moments (or cumulants); which are to be used to bring forth the asymptotic expansions for the distribution and percentiles of W_n.

5. Deriving the Edgeworth and Cornish-Fisher inverse expansions

The Edgeworth expansion for the probability integral

$$F(x) = \Pr[W_n < x]$$

and the Cornish-Fisher inverse expansion [HiD] for the percentiles

$$x_\alpha = G(u_\alpha) \quad \text{s.t.} \quad \Pr[W_n < x_\alpha] = \Pr[U < u_\alpha] = \alpha, \; U \sim N(0,1)$$

up to order $O(n^{-\frac{s-2}{2}})$ can be derived by following the procedure:

Step 1. Expand W_n to get the truncated Taylor series ${}^sW_n$ up to order $O(n^{-\frac{s-2}{2}})$.

Step 2. Compute the Taylor series of the powers ${}^sW_n^j$ $(j = 2, \cdots, s)$.

Step 3. Rewrite ${}^sW_n^j$ $(j = 1, \cdots, s)$ by means of the base $\{ P_{\boldsymbol{K}} \}$.

Step 4. Transform each $P_{\boldsymbol{K}}$ into the corresponding expression with the base $\{ A_{\boldsymbol{K}} \}$ by applying Algorithm PtoA.

Step 5. Take expectations ${}^sW_n^j$ $(j = 2, \cdots, s)$ to get the approximate moments of ${}^sW_n$ term by term.

Step 6. Calculate the approximate cumulants of ${}^sW_n$ from the relations between moments and cumulants.

Step 7. Find a normalizing transformation, if need be, and compute the moments and cumulants for the transformed variate.

Step 8. Put the above results into the formal Edgeworth expansion and the formal Cornish-Fisher inverse expansion. Then we obtain the asymptotic expansions for $F(x)$ and $G(u_\alpha)$ or those to the distribution of the transformed variate.

The main part of the procedure, say, Steps 4 and 5, have been implemented on computer as a package of LISP programs. Interface between the LISP package and the REDUCE system has made the above series of operations much easier and executable in one piece. For details, see Nakagawa and Niki [NN1] [NN4] [NN5] and Nakagawa [Nak].

References

[And] T. W. Anderson. *An Introduction to Multivariate Analysis (2nd ed.)*. Wiley, New York, 1984.

[BhG] R. N. Bhattacharya and J. K. Ghosh. On the validity of the formal Edgeworth expansion. *Ann. Statist.*, **6**(2):434–451, 1978.

[Den] M. Denker. *Asymptotic Distribution Theory in Nonparametric Statistics*, Vieweg, 1985.

[Hea] A. C. Hearn. *REDUCE User's Manual*. Santa Monica, version 3.3 edition, 1987.

[HiD] G. W. Hill and A. W. Davis. Generalized asymptotic expansions of Cornish-Fisher type. *Ann. Math. Statist.*, **39**:1264–1273, 1968.

[KNG] S. Konishi. Normalizing transformations of some statistics in multivariate analysis. *Biometrika*, **68**:647–651, 1981.

[KNG] S. Konishi, N. Niki and A. K. Gupta. Asymptotic expansions for the distributions of quadratic forms in normal variables. *Ann. Inst. Statist. Math.*, **40**:279–296, 1988.

[Mui] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, 1982.

[Nak] S. Nakagawa. *Symmetric Polynomial Calculus in Distribution Theory of Multivariate Statistics*. PhD thesis, Kyushu Univ., Fukuoka, 1994.

[NHN] S. Nakagawa, H. Hashiguchi, and N. Niki. Deriving the sampling cumulants of k-statistics with the help of computer algebra (in Japanese). *Sūshiki-shori*, **1**(2):44–46, 1992.

[NN1] S. Nakagawa and N. Niki. Multi-system symmetric polynomials and its applications to multivariate statistics (in Japanese). *Commun. Symb. Algeb. Manipula.*, **5**(4), 1988.

[NN2] S. Nakagawa and N. Niki. Derivation of the distribution of sample regression coefficient by using computer algebra. *Engineering Sciences Reports, Kyushu Univ.*, **11**(3):357–365, 1989.

[NN3] S. Nakagawa and N. Niki. Asymptotic expansions for the distribution of the sample correlation coefficient from bivariate elliptical populations. *Proc. 4-th China-Japan Symp. on Statist.*, 220–223, 1991.

[NN4] S. Nakagawa and N. Niki. Changes of bases of multi-system symmetric polynomials and their applications in distribution theory of multi-variate statistics (in Japanese). *Bull. Comp. Statist. of Jpn.*, 4(2):35–43, 1991.

[NN5] S. Nakagawa and N. Niki. Changes of bases of multi-system symmetric polynomials of several systems of quantities and their applications in distribution theory of multivariate statistics. *Theory and applications in computational statistics*, 166–182, 1991. (Asano, C., Niki, N. and Okazaki, T. eds.).

[NN6] S. Nakagawa and N. Niki. Distribution of the sample correlation coefficient for non-normal populations. *J. Japan. Soc. Comp. Statist.*, **5**:1–19, 1992.

[Ni1] N. Niki. Program packages for symmetric functions (in Japanese). *Commun. Symb. Algeb. Manipula.*, 4(2):14–17, 1986.

[Ni2] N. Niki. *Symbolic Computations in Distribution Theory of Multivariate Statistics.* DSc thesis, Kyushu Univ., 1986.

[Ni3] N. Niki. Algorithms on the representation of multi-system symmetric polynomials and their applications in statistics. *Proc. 11th Int. Symp. Computer at the University*, 5.1.1–6, 1989. (V. Čerić, V. Laužar, V. Mildner, eds.).

[NK1] N. Niki and S. Konishi. Higher order asymptotic expansions for the distribution of the sample correlation coefficient. *Commun. Statist.-Simula. Computa.*, **13**:169–182, 1984.

[NK2] N. Niki and S. Konishi. Distribution of the coefficient of variation for non-normal populations. *Proc. 2nd Japan-China Symp. Statist.*, 185–188, 1986. (Asano, C. *ed.*).

[NK3] N. Niki and S. Konishi. Effects of transformations in higher order asymptotic expansions. *Ann. Inst. Statist. Math.*, **38**:371–383, 1986.

[SHF] M. Siotani, T. Hayakawa and Y. Fujikoshi. *Modern Multivariate Statistical Analysis*, American Sciences Press, Ohio, 1985.

[voM] R. von Mises. On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.*, **18**, 1947.

[Wal] D. L. Wallace. Asymptotic approximations to the distributions. *Ann. Math. Statist.*, **29**:635–654, 1958.

[Wit] C. S. Withers. Expansions for the distribution and quantiles of the empirical distribution with applications to nonparametric confidence intervals, *Ann. Statist.*, **11**: 577–587, 1983.

CHAOS IN ITERATED CUBIC MAPS: Topological Entropy and Bifurcation.

Kiyoko NISHIZAWA
DEP. MATH. FAC.SCI. JOSAI UNIVERAITY
1-1, Keyakidai, Sakado, Saitama 350-02 JAPAN
e-mail: kiyoko@euclides.josai.ac.jp

Keywords: algebraic geometry - chaotic dynamical systems - topological entropy.
AMS(MOS)subject classification: 14Q05, 58F03, 58F20

1 Introduction.

System of iterated maps of the interval, viewed as dynamical systems, is considered as an important model for the chaotic behavior in certain physical, chemical and biological systems. Since there are many notions of chaos, in this paper we consider **topological chaos**, meaning positivity of the **topological entropy**.

For a parametrized family of functions, we have a vague general question; "how does the **complexity** of a dynamical system vary with parameters ?" One measure of complexity would be the numbers of periodic point of various periods. And the topological entropy of a map is also one of the particular useful indicator of the complexity of the system. In [11], Milnor and Thurston considered the topological entropy $h(f)$ and growth number $s(f) = \exp h(f)$ of continuous maps f. For a piecewise monotone map f the topological entropy is the exponential growth rate of the number of monotone pieces of the graph of the k-th iterate of f, denoted by $f^{\circ k}$, and also the exponential growth rate of the total variation of k-th iterate of f.

Definition. For a piecewise monotone map f considered as a map from the compact interval $[-\infty, \infty]$ to itself, the topological entropy $h(f)$ of f is defined as follows:

$$h(f) = \lim_{k\to\infty} \frac{1}{k} \log(l(f^{\circ k}))$$

where $l(f^{\circ k}) - 1$ is the number of **laps** of k-th iterate, that is the number of maximal intervals of monotonicity. For polynomial maps f with degree ≥ 2, $h(f)$ can be identified with the number

$$h_{per}(f) \doteq \limsup_{k\to\infty} \frac{1}{k} \log(\sharp fix(f^{\circ k})),$$

where $\sharp fix$ is the number of fix points ([11]).

In the special case of a piecewise linear map with |slope| =constant(≥ 1), the topological entropy is precisely equal to this constant |slope|. Except in very special cases as the above example, the entropy cannot be computed using these definitions and results. Thus the problem of finding an **algorithm** to compute topological entropy to any accuracy is discussed in many papers; for an example, let $\{f_\lambda(x) = \lambda x(1-x);\ \ \lambda \in [1,4]\ \}$ be the quadratic family of maps defined on the unit interval, then by using an algorithm in [2] the entropy $h(f_\lambda)$ is calculated and graphed in Figure 1 as a continuous function of the parameter λ . The quadratic(logistic) family of maps is important as a population

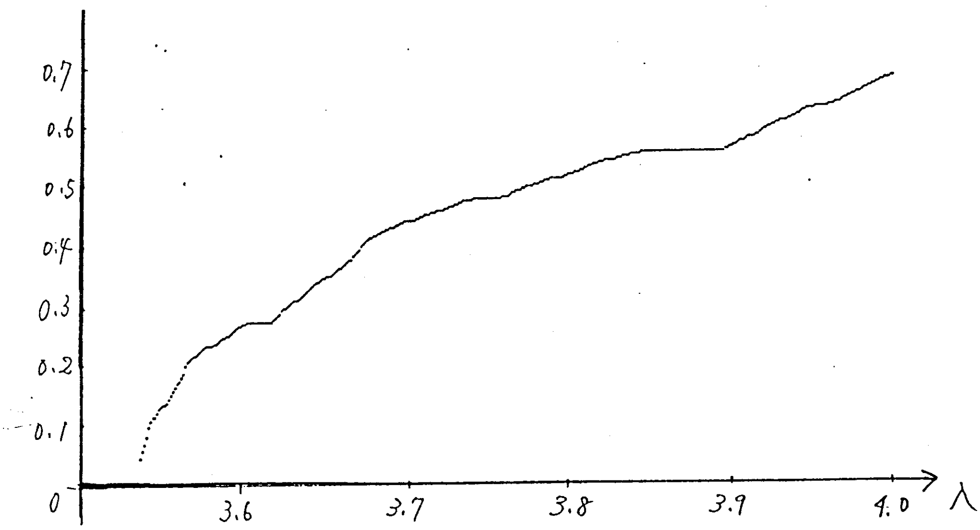


Figure 1: Topological entropy versus the parameter λ for the logistic family with $3.5 \leq \lambda \leq 4$.

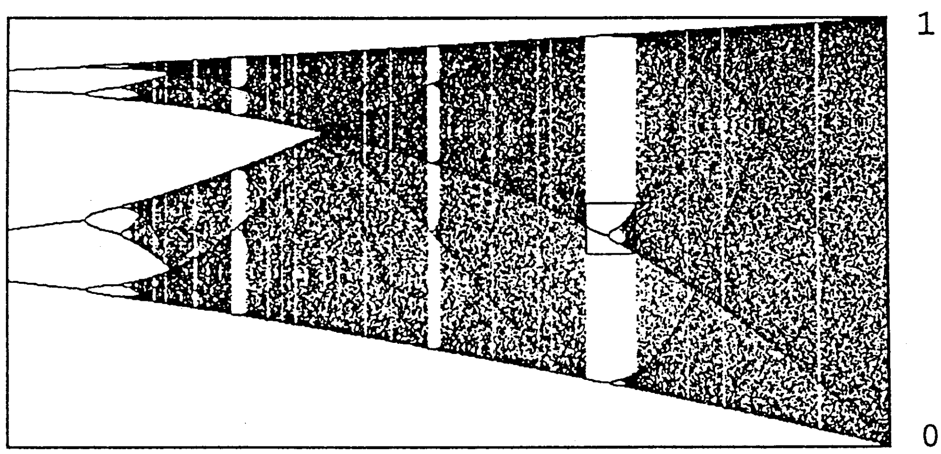


Figure 2: Bifurcation diagram for the logistic family for the parameter $3.5 \leq \lambda \leq 4$

growth model in theoretical population dynamics and an example of a family of simple maps with extremely complicated dynamics. Milnor and Thurston([11]), based on the Douady-Hubbard-Sullivan argument(unpublished) proved that this family has only orbit-creation values and no orbit-annihilation values and that the topological entropy $h(f_\lambda)$ is monotone increasing as a function of λ. This monotone bifurcation behavior of the family is illustrated by the uniform orientation of the pichforks depicted in Figure 2.

Recently an algorithm for computing the entropy in the cubic case has been given in [1]. In Figure 3 (fig.12 in [1]), a graph of the topological entropy of the parameterized cubic family $f(x) = x^3 + ax - b$ is shown. Unlike the monotone bifurcation

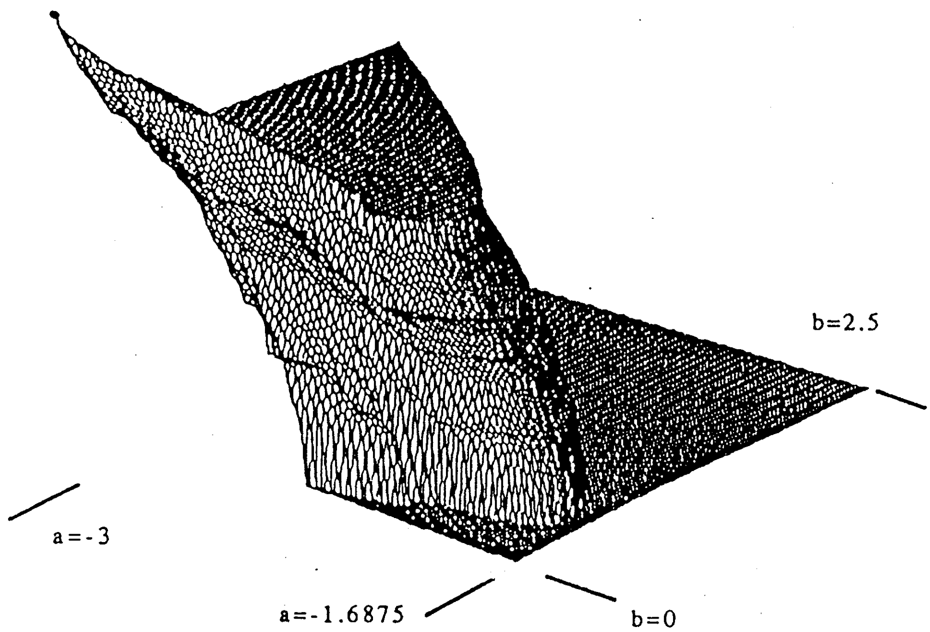


Figure 3: Topological entropy of the cubic maps $f(x) = x^3 + ax + b$ for $(a, b) \in [-3, -1.6875] \times [0, 2.5]$. The z-axis is $[0, \log 3]$. (see Figure 12, page 26 in [1])

structure of the logistic family, there exist many one-parameter families with negative Schwarizian derivative, exhibiting an antimonotone orbit-bifurcation structure and the corresponding topological entropy functions have local maxima as the parameter varies monotonely (see [5], [15]).

Example: one-parameter unimodal family with period-halving bifurcation.([15]) Let $F(x) = 1.1\, y\,(0.974 - y)$ with $y = \exp(11x)/(\exp(11x) + 19) - 0.05$, $x \in [0, 1]$. The one-parameter family of maps $F(\ ; m)$ is given by $F(x, m) = mF(x)$ with the parameter $m : 1 \leq m \leq 30$. See Figure 4for the graph of $F(x, 2)$ and for bifurcation diagram of this family. It is well known that if there is a period orbit of period three then the topological entropy of the map in question is at least $\log[(1 + \sqrt{5})/2]$. The topological entropy of $F(\ , m)$ is continuous as a function of m, and it has at least one local maximum and the maximaum value at this point is at least $\log[(1 + \sqrt{5})/2]$. Furthermore , the topological entropy is equal to zero for $m = 1$ and for $m = 30$.

The bulk of the present paper is devoted to the study of the transition to chaos for the cubic polynomials. Some conjectures concerning the entropy in case of cubic maps were enunciated by Milnor in [9]. In section 3 ,we shall discuss these problems. Our

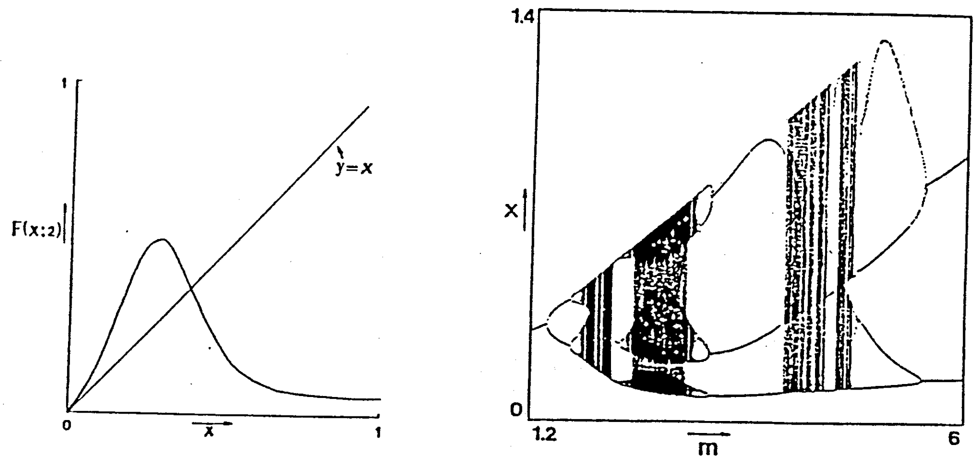


Figure 4: The graph of $F(x,2)$ and period halving bifurcation diagram for the family $F(x,m)$ for the parameter m : $(1.2 \leq m \leq 6)$ (see Figure 2 and 4, page 331 in [15])

main result is a classification of the routes to topological chaos along an algebraic curve defined in the moduli space of the real cubic polynomials.

2 Moduli space of the complex cubic polynomials.

We consider the family of cubic maps $x \mapsto g(x) = c_3x^3+c_2x^2+c_1x+c_0$ $(c_3 \neq 0, c_i \in \mathbf{C})$. For such a cubic map g , we have two normal forms ; $x^3 - 3Ax \pm \sqrt{B}$, $A, B \in \mathbf{C}$. Therefore, the complex affine conjugacy class of g can be represented by (A, B). The **moduli space**, denoted by $\mathcal{M}$, consisting of all affine conjugacy classes of cubic maps, can be identified with the coordinate space $\mathbf{C}^2 = \{(A, B)\}$ ([9]).

Moduli space of the real cubic polynomials. By a suitable real affine transformation, any **real** cubic map $g(x)$ is transformed to a unique map $f(x) = \sigma x^3 - 3Ax + \sqrt{|B|}$, where $\sigma := sgn(g''')$, $A, B \in \mathbf{R}$. The real affine conjugacy class of g or f can be represented by $(A, B) \in \mathbf{R}^2$ if $B \neq 0$. But if $B = 0$, σ should be added as an essential class invariant, as $x \mapsto x^3 - 3Ax$ and $x \mapsto -x^3 - 3Ax$ belong to different classes. Thus ,due to J.Milnor([9]), the real Moduli space, or real cut of the moduli space $\mathcal{M}$, of real affine conjugacy classes of real cubic maps can be described as the disjoint union of the upper half-plane $\mathbf{H}^+ = \{(A, B) | B \geq 0\}$ and the lower half-plane $\mathbf{H}^- = \{(A, B) | B \leq 0\}$ We denote this space by $\mathcal{M}_R$.

A complex cubic map f, or the corresponding point $(A, B) \in \mathcal{M}$, belongs to the **connectedness locus** if the orbits of both critical points p_i such that $f'(p_i) = 0$, $i = 1, 2$, are bounded. And f is **hyperbolic** if both of these critical orbits converge towards attracting periodic orbits. The set of all hyperbolic points in the moduli space $\mathcal{M}$ forms an open set. The famous hyperbolic conjecture is that this open set is precisely equal to the interior of the connectedness locus and is everywhere dense in the connectedness locus. Each connected component of this open set is called a **hyperbolic component**. By M.Rees([16]), each hyperbolic component contains a unique post-critically finite complex cubic map. So following A. Douady and J. Hubbard ([3]), this map is called a **center map** or **Thurston map** and the coordinates (A, B) of f will be called a **center** in the moduli space. Following M.Rees([16]) and J.Milnor([9]), the centers are roughly classified into four diffeent types, as follows. In the following t, p, q denote integers. A center is of the type $\mathcal{A}_p$ if two critical points p_1, p_2. of the center map coinside and has the period p : $f^p(p_1) = p_1$. In fact, only possible values for p in this case are $1, 2$. A center is of the type $\mathcal{B}_{p+q}$ if $f^p(p_1) = p_2$ and $f^q(p_2) = p_1$; of the type $\mathcal{C}_{(t)q}$ if $f^t(p_1) = p_2$ and $f^q(p_2) = p_2$; of the type $\mathcal{D}_{p,q}$ if $f^p(p_1) = p_1$ and $f^q(p_2) = p_2$. These exhaust all types of centers. It is clear that there are only a finite number of centers of a given type. See Figure 5 ([9]).

Example: There exist three centers of type $\mathcal{C}_{(3)1} \in \mathcal{M}_R$. The corresponding parameters are $(A, B) = (-.75040, -.18820)$, $(-.74949, -.18679)$, $(-.0924912, -.0614376)$.

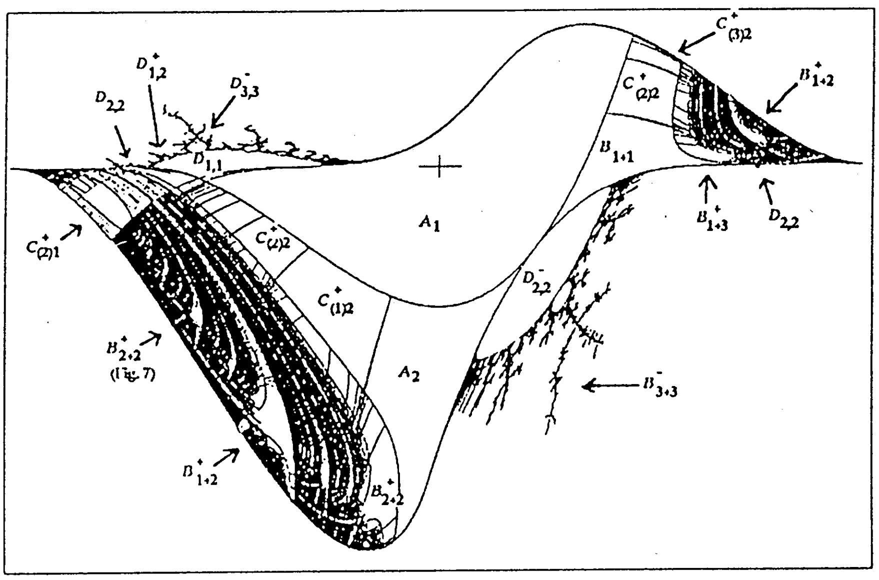


Figure 5: Complex connectedness locus with the four distinct classes of hyperbolic components. Region: $(A, B) \in [-1, 1] \times [-1.7, 0.65]$, (see Figure 18 in [9]).

Center curves in the moduli space.

The **center curves** CDp, BCp, which are algebraic curves, can be defined according to the above four renormalization-type. We show how the equations of these curves are obtained by induction on p ([12], [13] and [14]).

Theorem 2.1 : Defining equation of a center curve *For a given p, there exist an algebraic curve CDp containing all centers of the type $\mathcal{C}_{(k)p}$ and $\mathcal{D}_{k,p}$, and another algebraic curve BCp containing all centers of the type $\mathcal{B}_{p+k}$ and $\mathcal{C}_{(p)k}$.*

For example, we obtain precisely the following curves;

$$
\begin{aligned}
CD1 &: \ B = 4A(A+\tfrac{1}{2})^2, \\
BC1 &: \ B = 4A(A-\tfrac{1}{2})^2, \\
CD2 &: \ B^2 - 8A^3B + 4A^2B - 5AB + 2B + 16A^6 - 16A^5 \\
& \qquad -12A^4 + 16A^3 - 4A + 1 = 0, \\
BC2 &: \ B^3 - 12A^3B^2 - 6AB^2 + 2B^2 + 48A^6B + 24A^3B + 21A^2B \\
& \qquad -6AB + B - 64A^9 + 96A^7 - 20A^5 - 12A^3 - A = 0,
\end{aligned}
$$

We can embed $\mathbf{C}^2$ canonically in $\mathbf{P}^2(\mathbf{C}) : (A, B) \to (1 : A : B)$. Then an affine algebraic curve $V_0 = \{(A, B) \in \mathbf{C}^2 : h(A, B) = 0\}$ uniquely determines a projective algebraic curve $V = \{(C : A : B) \in \mathbf{P}^2(\mathbf{C}) : H(C : A : B) = 0\}$ in $\mathbf{P}^2(\mathbf{C})$ such that $h(A, B) = H(1 : A : B)$ and $V \cap \mathbf{C}^2 = V_0$.

Definition. For a center curve V_0, the corresponding projective algebraic curve V is called the **projective center curve**. We denote by PBCp and PCDp, these curves corresponding to BCp and CDp respectively.

We give some algebraic-geometric properties of these curves.

Theorem 2.2 : The interseciton with the line at infinity ([12], [14]). *Each projective center curve and the line at infinity, $L_\infty : C = 0$, intersect at the point $P_\infty = (0 : 0 : 1)$ only. P_∞ is singular and its multiplicity can be calculated explicitly.*

The irreducibility of each projective center curve is determined based on Kaltofen's algorithms on *risa-asir* (computer algebra system by FUJITSU CO.LTD.) ([17] , [8]).

Theorem 2.3 : Irreducibility and Singurarity ([14]). *For projective center curves PCDi and PBCi for only $i = 1$ and 2,*

- *PCD1 and PBC1 are irreducible curves of degree 3. PCD2 is an irreducible curve of degree 6. P_∞ is 4-fold and $(0.25, -0.4375)$ is an ordinary double point .*
- *PBC2 is an irreducible curve of degree 9. P_∞ is 6-fold and four ordinary double points are as follows :*

$$
\begin{aligned}
&(-0.1341351918179714, -1.373444849102 64), \\
&(-0.5531033117555605, -0.6288238268413773), \\
&(0.3041906503790061 * i + 0.3436192517867655, \\
&\qquad 0.6886343379400248 - 0.04267412324347224 * i), \\
&(0.3436192517867655 - 0.3041906503790061 * i, \\
&\qquad 0.04267412329900053 * i + 0.6886343379735695),
\end{aligned}
$$

To calculate genus g of each projective center curve Γ, we determine the principal part at P_∞ of the curves by using Newton Polygons and apply the Plücker's formula. I am grateful to Y. Komori([6]) for helpful suggestions on the genus.

Theorem 2.4 : Principal part of the center curves and Genus. ([12]) *The principal part at P_∞ of PCD1 and of PBC1 is $(C^2 - 4A^3)^1$, of PCD2 is $(C^2 - 4A^3)^2$, and* *of PBC2 is* $(C^2 - 4A^3)^3$. *The curves PCD1 and PBC1 are rational. Hence the genus is 0. The genus of PCD2 is* 1. *The genus of PBC2 is* 3.

We would like to state the following conjectures for the projective center curves:

Conjectures

- All projective center curves are irreducible.
- All singular points except P_∞ are ordinary double points.
- Especially, for real graph of center curves, the sigular point exists only in $\mathcal{R}_1$.
- The pricipale part at P_∞ of every projective center curve has a form $(C^2 - 4A^3)^k$.

3 Monotonicity of topological entropy along center curves.

The conjecture that the toplogical entropy $h(f)$ of a real cubic map f depends **monotonely** on its parameters was enunciated by Milnor in [9] and [10]. Namely, each locus of constant entropy in parameter space is connected. Another conjecture due to Milnor([9], [10]) is a **maximum and minimum principle** for entropy: the maximum and minimum values for the entropy function on any closed region in the moduli space must occur the boundary. Real graphs of BC1 and CD2-2 are shown in Figure 6 and

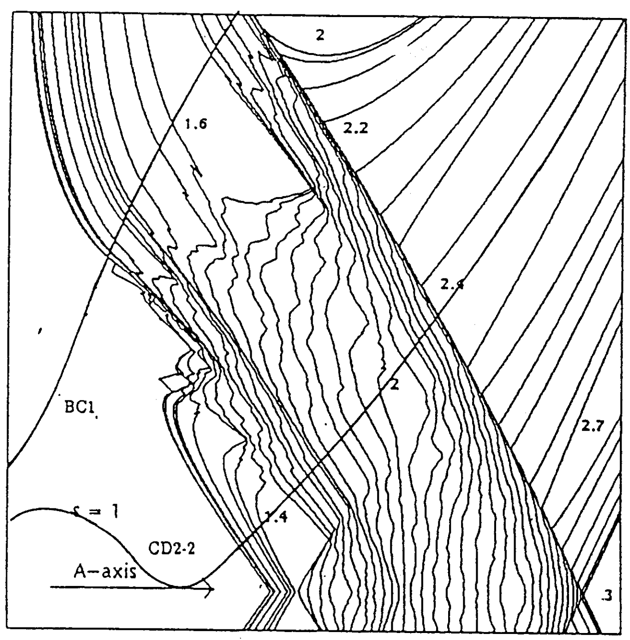


Figure 6: Real graphs of BC1 , CD2-2 with the equi-growth number lines. The region is $[.57, 1.03] \times [0, .43]$ in (A, b)− plane

of CD1, BC1, and CD2-2 in Figure 7 together with the equi-growth number lines in the figures due to Block and Keesling ([1]).

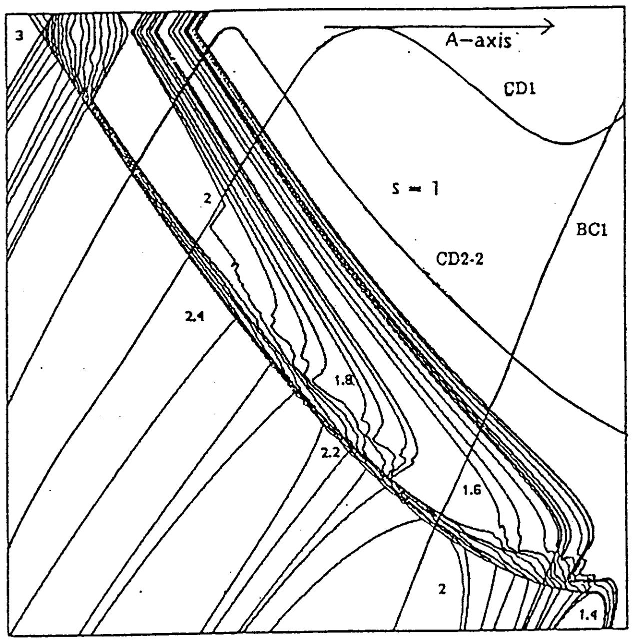


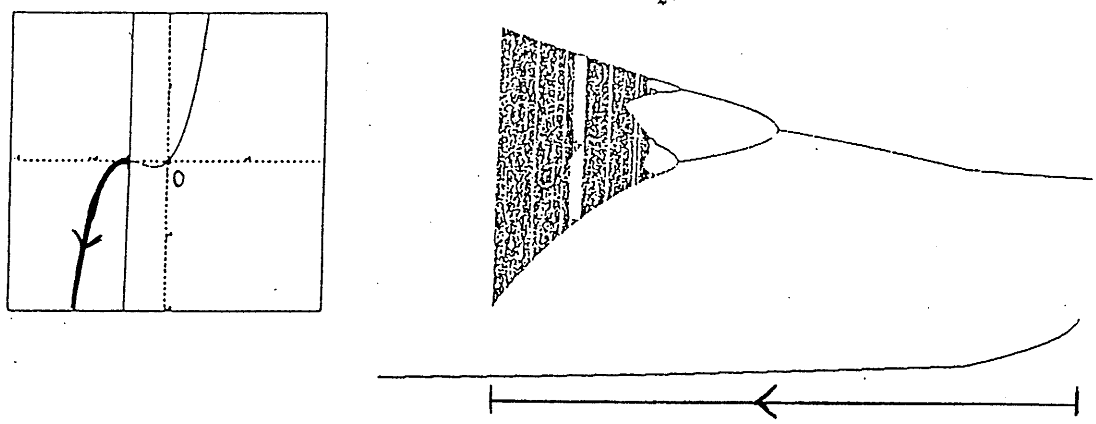
Figure 7: Real graphs of CD1, BC1, CD2-2 with the equi-growth number lines. The region is $[-1.05, -.09] \times [0, -1.35]$ in $(A, b')-$)plane

A glance at these figures suggests that the growth number and the topological entropy vary monotonously along a part of real graph of center curve. We proposed in [12],[14] this entropy-monotonicity conjecture alog a center curve.. And furthermore Figure 8shows that bifurcation along a part of center curve is monotone. This monotone bifurcations as in the logistic case seems to lend strong support to our conjecture. In [10], J.Minor gives a result for this monotonicity conjecture as follows : *If the hyperbolic conjecture is true, then each isentrope* $\{(A, B) : h(f_{(A,B)}) = constant\}$ *is connected.* The famous hyperbolic conjecture is that the open set of all hyperbolic points is precisely equal to the interior of the connectedness locus and is everywhere dense in the connectedness locus. Even for the qudratic polynomial maps, this conjectur is unsolved.

Recentely we can prove that our monotonicity conjecture is true on CD1. I am grateful to Y. Komori([7]) for helpful suggestions on this conjecture. He can prove that our monotonicity conjecture is true for any center curves and his idea can be applied to suitable family of polynomials with higher degree. A fundamental tool is the following Thurston's Theorem for real Polynomial maps.

Thurston's rigidity theorem: *Given any post-critically finite m-modal map, there exists one and up to positive affine conjugation only one polynomial of degree* $m + 1$ *with the same kneading data.*

Theorem 3.1 : entropy-monotonicity *The topological entropy of a cubic mapf is monotonely continuous if f varies along a part(called a bone) of center curve 'CD1,* *CD1* : $B = 4A(A + \frac{1}{2})^2$,



BC1 : $B = 4A(A - \frac{1}{2})^2$,

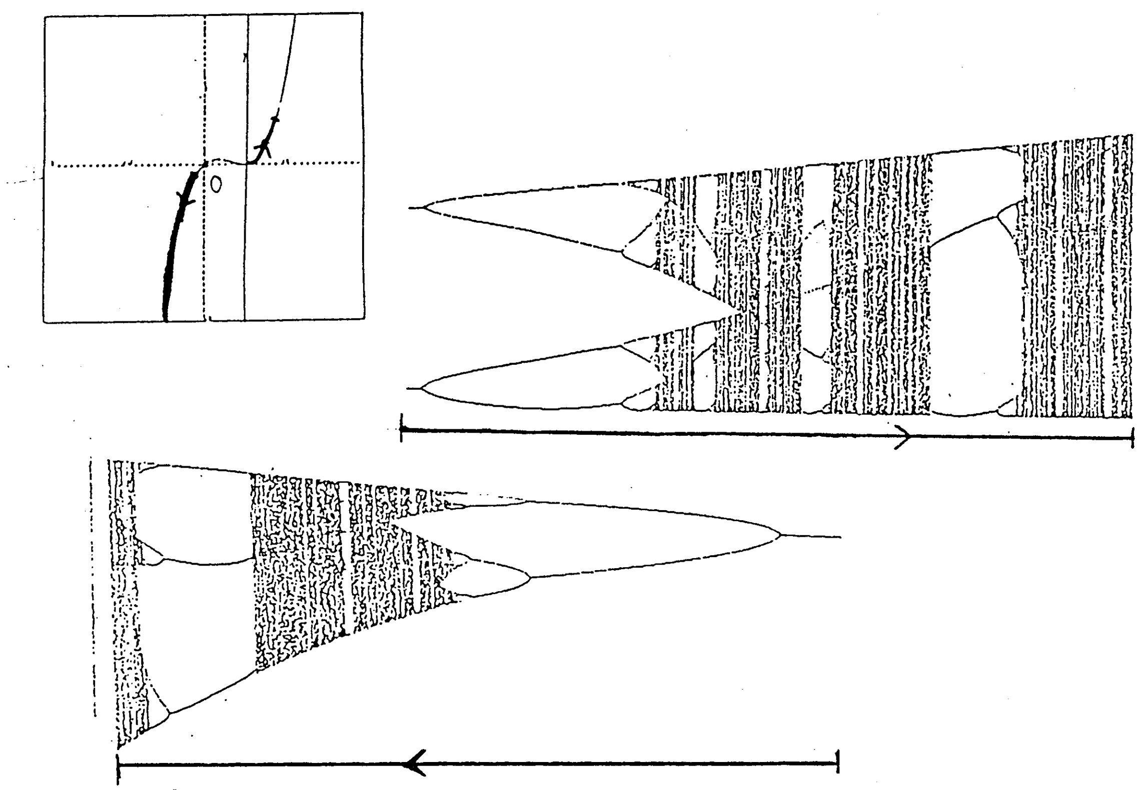


Figure 8: Bifurcation diagrams along CD1, BC1

References

[1] L. Block and J. Keesling, Computing the topological entropy of maps of the interval with three monotone pieces. J. Statist. Phys.,**66**(1992)755-774.

[2] L. Block, J. Keesling, Shihai Li, K. Peterson, An Improved Algorithm for Computing Topological Entropy. J. Statist. Phys.,**55**(1989) 929-939.

[3] A. Douady & J. Hubbard, Etude dynamique des polynomes complexes I(1984) & II(1985), Publ. Math. Orsay.

[4] P. Griffiths and J. Harris, Principles of Algebraic Geometry, John Wiley & Sons, (1978).

[5] Ittai Kan, H. Koçak, J. Yorke, Antimonotonicity:Concurrent creation and annihilation of periodic orbits, Ann. Math., **130**(1978) 219-252.

[6] Y. Komori, Principal part and blowing-up, private communication , 1994.3.22.

[7] Y. Komori, Monotonicity of topological entropy in the cubic maps, private communication, 1994.10.

[8] S. Landau, Factoring polynomials over algebraic number fields, SIAM J. Comput., **14**(1985) 184–195.

[9] J. Milnor, Remarks on iterated cubic maps. Preprint # 1990/6, SUNY StonyBrook.

[10] J. Milnor, Entropy for Bimodal map, resume of international conference of dynamical systems, Denmark, June, 1993.

[11] J. Milnor & W. Thurston, On iterated maps of the interval. L.N.Math. 1342, *Springer-Verlag*, (1988) 465–563 .

[12] K. Nishizawa, COMPLEX DYNAMICAL SYSTEMS:ALGEBRAIC CURVES IN THE MODULI SPACE OF THE CUBIC MAPS, to appear,

[13] K. Nishizawa & A. Nojiri, Center curves in the moduli space of the real cubic maps, Proc. Japan Acad. Ser.A, **69**(1993) 179–184.

[14] K. Nishizawa & A. Nojiri, Algebraic geomerty of center curves in the moduli space of the cubic maps, Proc. Japan Acad. Ser.A, **70**(1994) 99–103.

[15] H.E. Nusse & J. Yorke, Period halving for $x_{n+1} = mF(x_n)$ where F has negative Schwarzian Derivative,, Phys. Letters A, **127**(1988) 328-334.

[16] M. Rees, Components of degree two hyperbolic rational maps, Invent. Math., **100**(1990) 357–382.

[17] K. Yokoyama, M. Noro, & T. Takeshima, On factoring multi-variate polynomials over algebraically closed fields, RISC-Linz Series, no.90-26.0 (1990) 1-8.

COMPUTING AVERAGES WITH GROUP THEORY

Edgar N. Reyes
Department of Mathematics, Southeastern Louisiana University
Hammond, Louisiana USA 70402
E-mail: ereyes@selu.edu

1 Introduction

Group theory is a rich subject which is made more interesting with advancing technology. To illustrate this point, we consider an elementary example where we compute averages of numbers labeled on the 12 faces of a dodecahedron. This process of taking averages is described as a linear operator. A natural representation of the group of rotations of the dodecahedron is shown to intertwine with the linear operator. This way, the operator can be analyzed through the characters of the group. Many of our calculations will make use of GAP which is a system for computational discrete algebra [1].

To describe the dodecahedron, a few notations are needed, denote the faces by $f(1)$, $f(2)$, $f(3)$,..., $f(12)$. Arrange them in a manner so the following pairs of faces are opposite each other

$$\begin{array}{ll} f(1) \textit{ and } f(2) & f(3) \textit{ and } f(10) \\ f(4) \textit{ and } f(11) & f(5) \textit{ and } f(12) \\ f(6) \textit{ and } f(8) & f(7) \textit{ and } f(9) \end{array} \qquad (1.1)$$

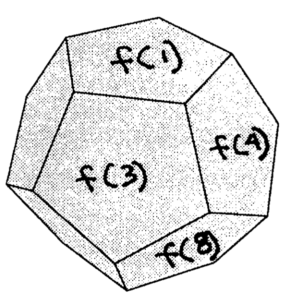


Figure 1

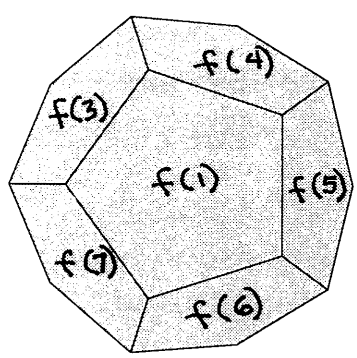


Figure 2

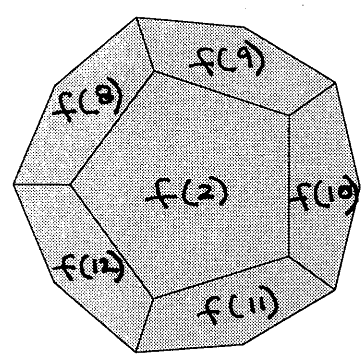


Figure 3

The views from the top and bottom of the dodecahedron in Figure 1 are shown in Figures 2 and 3, respectively.

Initially, assume for each $i = 1, ..., 12$, that face $f(i)$ is painted with the number i. On the second day, erase the label on face $f(i)$ and replace it by the average of yesterday's labels except those that were on $f(i)$ and on the face opposite $f(i)$; do this for $i = 1, ..., 12$. Repeat this process on a daily basis. For the first four days, the labels are given by

Day	$f(1)$	$f(2)$	$f(3)$	$f(4)$	$f(5)$	$f(6)$	$f(7)$	$f(8)$	$f(9)$	$f(10)$	$f(11)$	$f(12)$
1	1	2	3	4	5	6	7	8	9	10	11	12
2	$\frac{15}{2}$	$\frac{15}{2}$	$\frac{13}{2}$	$\frac{63}{10}$	$\frac{61}{10}$	$\frac{32}{5}$	$\frac{31}{5}$	$\frac{32}{5}$	$\frac{31}{5}$	$\frac{13}{2}$	$\frac{63}{10}$	$\frac{61}{10}$
3	$\frac{63}{10}$	$\frac{63}{10}$	$\frac{13}{2}$	$\frac{327}{50}$	$\frac{329}{50}$	$\frac{163}{25}$	$\frac{164}{25}$	$\frac{163}{25}$	$\frac{164}{25}$	$\frac{13}{2}$	$\frac{327}{50}$	$\frac{329}{50}$
4	$\frac{327}{50}$	$\frac{327}{50}$	$\frac{13}{2}$	$\frac{1623}{250}$	$\frac{1621}{250}$	$\frac{812}{125}$	$\frac{811}{125}$	$\frac{812}{125}$	$\frac{811}{125}$	$\frac{13}{2}$	$\frac{1623}{250}$	$\frac{1621}{250}$

(1.2)

Note, each label on the fourth day is approximately 6.5 which is the average of the integers $1, ..., 12$. This leads us to speculate whether the sequence of labels on each face will converge to this average. We will show that, indeed, this is the case and if we choose a different labeling initially then the sequence of labels of each face will converge to the average of the initial labeling.

2 Group Theory at Work

Let V be the vector space of complex-valued functions defined on the faces of the dodecahedron. One can identify a function $A \in V$ by a 12-tuple $(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}, a_{12})$ where $A(f(i)) = a_i$. The 12-tuple $(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)$ is the initial labeling in (1.2). The process of changing the labels from one day to the next is given by the linear operator

$$(LA)(f(i)) = \frac{1}{10} \sum A(f(j)) \tag{2.1}$$

where the sum is taken over all faces $f(j)$ except $f(i)$ and the face opposite $f(i)$.

If α permutes $1, 2, 3, ..., 12$ then α permutes the faces of the dodecahedron according to $f(i)^\alpha = f(i^\alpha)$ where i^α is the image of i under the α. Let G be the group generated by the rotations about the centers of the faces. The generators of G are given by the permutations

$$\begin{aligned} g1 &= (3,4,5,6,7)(8,9,10,11,12) \\ g2 &= (1,7,12,8,4)(5,6,11,2,9) \\ g3 &= (1,3,8,9,5)(2,10,6,7,12) \\ g4 &= (1,4,9,10,6)(2,11,7,3,8) \\ g5 &= (1,5,10,11,7)(3,4,9,2,12) \\ g6 &= (1,6,11,12,3)(2,8,4,5,10). \end{aligned} \tag{2.2}$$

Our convention in writing permutations, follow GAP i.e. actions act from the right and $i^{\alpha\beta} = (i^\alpha)^\beta$. According to GAP, the group G is isomorphic to the alternating group A_5 (see 3.1).

Denote the group of linear isomorphisms of V by $GL(V)$ and let $\pi : G \to GL(V)$ be the representation of G given by

$$(\pi_g(B))(f(i)) = B(f(i)^g). \tag{2.3}$$

It is easy to show L commutes with the representation π. To see this, denote the face opposite $f(j)$ by $-f(j)$. Since each generator in (2.2) of G is a rotation about a center of a face then $\pm f(j)^g = (\pm f(j))^g$ for all $g \in G$. So

$$\begin{aligned}
(L \circ \pi_g)B(f(i)) &= \frac{1}{10} \sum_{f(j) \neq \pm f(i)} (\pi_g B)(f(j)) \\
&= \frac{1}{10} \sum_{f(j) \neq \pm f(i)} B(f(j)^g) \\
&= \frac{1}{10} \sum_{f(j)^{g^{-1}} \neq \pm f(i)} B(f(j)) \\
&= \frac{1}{10} \sum_{f(j) \neq (\pm f(i))^g} B(f(j)) \\
&= \frac{1}{10} \sum_{f(j) \neq \pm f(i)^g} B(f(j)) \\
&= LB(f(i)^g) \\
&= (\pi_g \circ L)B(f(i))
\end{aligned}$$

and $L \circ \pi_g = \pi_g \circ L$ for all $g \in G$. A decomposition of π into invariant subspaces will prove useful in understanding the operator L.

Since G acts transitively on the faces and if we let $H = \{g \in G : f(1)^g = f(1)\}$ then the mapping $\psi : Hg \mapsto f(1)^g$ defines a bijection from the coset space $H\backslash G$ onto the set of faces of the dodecahedron. By GAP, H is the cyclic subgroup generated by $g1$ (see 3.2).

Let Φ be the representation of G induced by the identity one dimensional representation of H. The representation Φ acts on the vector space $L(H\backslash G)$ of complex valued functions on $H\backslash G$ such that $\Phi(x)F(Hg) = F(Hgx)$ where $F \in L(H\backslash G)$ (see [3, page 182]).

Note $\psi_* : f \mapsto f \circ \psi$ defines a vector space isomorphism from V onto $L(H\backslash G)$ and

$$\begin{aligned}
(\psi_*^{-1} \circ \Phi(x) \circ \psi_* F)(f(1)^g) &= (\Phi(x) \circ \psi_* F)(Hg) \\
&= (\psi_* F)(Hgx) \\
&= F(f(1)^{gx}) \\
&= F((f(1)^g)^x) \\
&= (\pi_x F)(f(1)^g).
\end{aligned}$$

So π is equivalent to Φ. In decomposing π, it will be helpful to find the number of double cosets $H\backslash G/H$. By using GAP (see 3.3), there are 4 double cosets and they are represented by

(i) e or ( ),

(ii) $(1,2)(3,8)(4,12)(5,11)(6,10)(7,9)$,

(iii) $(1,3)(2,10)(4,7)(5,12)(6,8)(9,11)$, and

(iv) $(1,8,10)(2,6,3)(4,9,5)(7,12,11)$.

This suggest V can be written as a direct sum of four invariant subspaces [3, page 205]. To identify these subspaces, knowing the irreducible characters of $G = A_5$ will be helpful. The group $G = A_5$ has five conjugacy classes and five irreducible characters. By using GAP, we find elements of A_5 belonging to the five conjugacy classes $cl[1], cl[2], cl[3], cl[4], cl[5]$. Namely, (see 3.4)

(i) $e \in cl[1]$,

(ii) $(3,4,5,6,7)(8,9,10,11,12) \in cl[2]$,

(iii) $(3,5,7,4,6)(8,10,12,9,11) \in cl[3]$,

(iv) $(1,2)(3,8)(4,12)(5,11)(6,10)(7,9) \in cl[4]$, and

(v) $(1,3,4)(2,10,11)(5,7,8)(6,12,9) \in cl[5]$.

Let $\chi_1, \chi_2, \chi_3, \chi_4$, and χ_5 be the irreducible characters of $G = A_5$. The values of these characters are constant on conjugacy classes. From GAP, we find the following where $\alpha = e^{2\pi i/5}$ (see 3.5)

(i) $\chi_1(cl[i]) = 1$ for $i = 1, \ldots, 5$,

(ii) $\chi_2(cl[1]) = 3$, $\chi_2(cl[2]) = -\alpha^2 - \alpha^3$, $\chi_2(cl[3]) = -\alpha - \alpha^4$, $\chi_2(cl[4]) = -1$, $\chi_2(cl[5]) = 0$,

(iii) $\chi_3(cl[1]) = 3$, $\chi_3(cl[2]) = -\alpha - \alpha^4$, $\chi_3(cl[3]) = -\alpha^2 - \alpha^3$, $\chi_3(cl[4]) = -1$, $\chi_3(cl[5]) = 0$,

(iv) $\chi_4(cl[1]) = 4$, $\chi_4(cl[2]) = -1$, $\chi_4(cl[3]) = -1$, $\chi_4(cl[4]) = 0$, $\chi_4(cl[5]) = 1$, and

(v) $\chi_5(cl[1]) = 5$, $\chi_5(cl[2]) = 0$, $\chi_5(cl[3]) = 0$, $\chi_5(cl[4]) = 1$, $\chi_5(cl[5]) = -1$.

If χ_i and χ_π are the characters of representations π_i (which is irreducible) and π, respectively, then the multiplicity of π_i in π is $\langle \chi_i, \chi_\pi \rangle$ where

$$\langle \chi_i, \chi_\pi \rangle = \frac{1}{|G|} \sum_{g \in G} \chi_i(g) \overline{\chi_\pi(g)}.$$

Let $\chi_i|H$ be the restriction of χ_i to the subgroup H. Since π is the representation induced by the trivial representation of H, then by the Frobenious Reciprocity Theorem [2, page 56]

$$\begin{aligned} \langle \chi_i, \chi_\pi \rangle &= \langle \chi_i | H, 1 \rangle \\ &= \frac{1}{|H|} \sum_{g \in H} \chi_i(h). \end{aligned} \tag{2.4}$$

To calculate the multiplicities, it is will be easier to use (2.4). By using GAP, we find $e \in cl[1]$, $g1 \in cl[2]$, $(g1)^2 \in cl[3]$, $(g1)^3 \in cl[3]$, $(g1)^4 \in cl[2]$ (see 3.6). Since $1+\alpha+\alpha^2+\alpha^3+\alpha^4=0$, then

$$\begin{aligned}
\langle \chi_2, \chi_\pi \rangle &= \frac{1}{5}\left[\chi_2(cl[1]) + \chi_2(cl[2]) + \chi_2(cl[3]) + \chi_2(cl[3]) + \chi_2(cl[2])\right] \\
&= \frac{1}{5}\left[3 + (-\alpha^2-\alpha^3) + (-\alpha-\alpha^4) + (-\alpha-\alpha^4) + (-\alpha^2-\alpha^3)\right] \\
&= \frac{1}{5}\left[3 - 2(\alpha+\alpha^2+\alpha^3+\alpha^4)\right] \\
&= \frac{1}{5}\left[3-2(-1)\right] \\
\langle \chi_2, \chi_\pi \rangle &= 1
\end{aligned}$$

Similarly, one can show $\langle \chi_1, \chi_\pi \rangle = 1$, $\langle \chi_3, \chi_\pi \rangle = 1$, $\langle \chi_4, \chi_\pi \rangle = 0$, and $\langle \chi_5, \chi_\pi \rangle = 1$.
So $\chi_\pi = \chi_1 + \chi_2 + \chi_3 + \chi_5$ [2, page 16].

Next, we will find the isotypic summands V_1, V_2, V_3, and V_5 of V such that the restriction of π to V_i is equivalent to π_i for $i = 1, 2, 3, 5$. This is a classical result. The image of V under the projection

$$P_i = \frac{n_i}{\#G} \sum_{g \in G} \chi_i(g^{-1})\pi_g$$

is V_i where n_i is dimension of V_i [2, page 21]. Clearly, the space of constant functions is invariant under the representation π and, thus, this space must be V_1. Define the following functions in V,

$$\begin{aligned}
e1 &= [1,1,1,1,1,1,1,1,1,1,1,1] \\
e2 &= [1,-1,0,0,0,0,0,0,0,0,0,0] \\
e3 &= [0,0,1,0,0,0,0,0,0,-1,0,0] \\
e4 &= [0,0,0,1,0,0,0,0,0,0,-1,0] \\
e5 &= [0,0,0,0,1,0,0,0,0,0,0,-1] \\
e6 &= [0,0,0,0,0,1,0,-1,0,0,0,0] \\
e7 &= [0,0,0,0,0,0,1,0,-1,0,0,0] \\
e8 &= [1,1,0,0,0,0,-1,0,-1,0,0,0] \\
e9 &= [0,0,1,0,0,0,-1,0,-1,1,0,0] \\
e10 &= [0,0,0,1,0,0,-1,0,-1,0,1,0] \\
e11 &= [0,0,0,0,1,0,-1,0,-1,0,0,1] \\
e12 &= [0,0,0,0,0,1,-1,1,-1,0,0,0]
\end{aligned}$$

If LS denotes the linear span then $V_1 = LS\{e1\}$. By using GAP, it can be shown P_5 leaves the vectors $e8, e9, e10, e11, e12$ invariant. In particular, in (3.7)-(3.8) we have
$P_5(e8) = Sum(z[8])/12, \ldots, P_5(e12) = Sum(z[12])/12$. Since V_5 is 5-dimensional,

$$V_5 = LS\{e8, e9, e10, e11, e12\}. \tag{2.5}$$

When L is restricted to V_5, it is acts by multiplication by $-1/5$. Similarly, from (3.9) we see $P_2(e2) = Sum(z[2])/20$ and from a few more calculations which we leave out, we obtain

$$\begin{aligned}
P_2(e2) &= [1/2, -1/2, \beta, \beta, \beta, \beta, \beta, -\beta, -\beta, -\beta, -\beta, -\beta] \\
P_2(e3) &= [\beta, -\beta, 1/2, \beta, -\beta, -\beta, \beta, \beta, -\beta, -1/2, -\beta, \beta] \\
P_2(e4) &= [\beta, -\beta, \beta, 1/2, \beta, -\beta, -\beta, \beta, \beta, -\beta, -1/2, -\beta]
\end{aligned}$$

$$\begin{aligned}
P_3(e5) &= [-\beta,\beta,\beta,-\beta,1/2,-\beta,\beta,\beta,-\beta,-\beta,\beta,-1/2]\\
P_3(e6) &= [-\beta,\beta,\beta,\beta,-\beta,1/2,-\beta,-1/2,\beta,-\beta,-\beta,\beta]\\
P_3(e7) &= [-\beta,\beta,-\beta,\beta,\beta,-\beta,1/2,\beta,-1/2,\beta,-\beta,-\beta]
\end{aligned}$$

where $\beta = \frac{1}{10}\left(e^{2\pi i/5} - e^{4\pi i/5} - e^{6\pi i/5} + e^{8\pi i/5}\right) = \sqrt{5}/10$. Since V_2 and V_3 are both 3-dimensional,

$$V_2 = LS\{P_2(e2), P_2(e3), P_2(e4)\} \quad \textit{and} \quad V_3 = LS\{P_3(e5), P_3(e6), P_3(e7)\}.$$

But L sends each of $P_2(e2), P_2(e3), P_2(e4), P_3(e5), P_3(e6), P_3(e7)$ to the zero vector, and so L is the zero operator when restricted to V_2 and V_3.

With respect to the inner-product

$$\langle (a_1, a_2, , a_3, ..., a_{12}), (b_1, b_2, b_3, ..., b_{12})\rangle = \sum_{z=1}^{12} a_z \overline{b_z} \tag{2.6}$$

one can normalize the vector $e1$ i.e. divide $e1$ by it's magnitude to get $n_1 = \frac{1}{2\sqrt{3}}[1,1,1,1,1,1,1,1,1,1,1,1]$. By the Gram-Schmidt orthogonalization method, one can obtain an orthonormal basis $\{n_1, n_2, ..., n_{12}\}$ for V in such a way that $V_2 = LS\{n_2, n_3, n_4\}$, $V_3 = LS\{n_5, n_6, n_7\}$, and $V_5 = LS\{n_8, n_9, n_{10}, n_{11}, n_{12}\}$. For any $B = [B_1, ..., B_{12}] \in V$,

$$\begin{aligned}
B &= \langle B, n_1\rangle n_1 + \sum_{z=2}^{12}\langle B, n_z\rangle n_z\\
&= \frac{1}{2\sqrt{3}}\left(\sum_{i=1}^{12} B_i\right) n_1 + \sum_{z=2}^{12}\langle B, n_z\rangle n_z\\
&= \left(\frac{1}{2\sqrt{3}}\right)^2 \left(\sum_{i=1}^{12} B_i\right)[1,1,1,1,1,1,1,1,1,1,1,1] + \sum_{z=2}^{12}\langle B, n_z\rangle n_z\\
&= \frac{1}{12}\left(\sum_{i=1}^{12} B_i\right)[1,1,1,1,1,1,1,1,1,1,1,1] + \sum_{z=2}^{12}\langle B, n_z\rangle n_z.
\end{aligned}$$

Iterating the operator L, we have

$$L^k(B) = \frac{1}{12}\left(\sum_{i=1}^{12} B_i\right)[1,1,1,1,1,1,1,1,1,1,1,1] + \left(-\frac{1}{5}\right)^k \sum_{z=8}^{12}\langle B, n_z\rangle n_z.$$

for $k \geq 1$. Thus, as $k \to \infty$, $L^k(B)$ approaches $\frac{1}{12}\left(\sum_{i=1}^{12} B_i\right)[1,1,1,1,1,1,1,1,1,1,1,1]$. In other words, the sequences of labels on the faces of the dodecahedron all converge to the average of the labels of any particular day.

3 Computations From GAP

In this section, we show calculations obtained through GAP version 3.4. The following are the syntax and results obtained except that the output from input (3.9) is abbreviated because the exact output is long.

$$gap > \quad g1 = (3,4,5,6,7)*(8,9,10,11,12);;$$

$$gap > \quad g2 = (1,7,12,8,4)*(5,6,11,2,9);;$$

$$gap > \quad g3 = (1,3,8,9,5)*(2,10,6,7,12);;$$

$$gap > \quad g4 = (1,4,9,10,6)*(2,11,7,3,8);;$$

$$gap > \quad g5 = (1,5,10,11,7)*(3,4,9,2,12);;$$

$$gap > \quad g6 = (1,6,11,12,3)*(2,8,4,5,10);;$$

$$gap > \quad g := Group(g1,g2,g3,g4,g5,g6);;$$

$$gap > \quad g.name := \text{"}G\text{"};$$

$$\text{"}G\text{"}$$

$$gap > \quad GroupId(g); \tag{3.1}$$

$$rec($$

$$catalogue := [60,13],$$

$$names := [\text{"}A5\text{"}, \text{"}PSL(2,4)\text{"}, \text{"}PSL(2,5)\text{"}],$$

$$size := 60)$$

$$gap > \quad h := Stabilizer(g,1); \tag{3.2}$$

$$Subgroup(G,[(3,4,5,6,7)(8,9,10,11,12)])$$

$$gap > \quad h.name := \text{"}H\text{"}$$

$$\text{"}H\text{"}$$

$$gap > \quad DoubleCosets(g,h,h); \tag{3.3}$$

$$[DoubleCoset(H,(\ ),H),$$

$$DoubleCoset(H,(1,2)(3,8)(4,12)(5,11)(6,10)(7,9),H),$$

$$DoubleCoset(H,(1,3)(2,10)(4,7)(5,12)(6,8)(9,11),H),$$

$$DoubleCoset(H,(1,8,10)(2,6,3)(4,9,5)(7,12,11),H)]$$

$$gap > \quad cl := ConjugacyClasses(g); \tag{3.4}$$

$$[ConjugacyClass(G,(\ )),$$

$$ConjugacyClass(G,(3,4,5,6,7)(8,9,10,11,12)),$$

$$ConjugacyClass(G,(3,5,7,4,6)(8,10,12,9,11)),$$

```
        ConjugacyClass(G,(1,2)(3,8)(4,12)(5,11)(6,10)(7,9)),
        ConjugacyClass(G,(1,3,4)(2,10,11)(5,7,8)(6,12,9))]
gap > gl := Elements(g);;
gap > ct := CharTable(g);;
gap > PrintCharTable(ct);                                              (3.5)
    rec(size := 60, centralizers := [60,5,5,4,3],
      orders := [1,5,5,2,3], classes := [1,12,12,15,20],
      irreducibles := [[1,1,1,1,1],
        [3,-E(5)^2 - E(5)^3, -E(5) - E(5)^4, -1, 0],
        [3,-E(5) - E(5)^4, -E(5)^2 - E(5)^3, -1, 0],
        [4,-1,-1,0,1],[5,0,0,1,-1]],
      operations := CharTableOps, identifier := "G",
      order := 60, name := "G",
      powermap := [,[1,3,2,1,5],[1,3,2,4,1],,[1,1,1,4,5]],
      galomorphisms := Group((2,3)),
      text := "origin : Dixon'sAlgorithm", group := G)
gap > g1 in cl[2]; g1^2 in cl[3]; g1^3 in cl[3]; g1^4 in cl[2];      (3.6)
        true
        true
        true
        true
gap > chi5 := function(n)
    > if n^(-1) in cl[1] then
    > return ct.irreducibles[5][1];
    > elif n^(-1) in cl[2] then
    > return ct.irreducibles[5][2];
    > elif n^(-1) in cl[3] then
    > return ct.irreducibles[5][3];
    > elif n^(-1) in cl[4] then
    > return ct.irreducibles[5][4];
    > elif n^(-1) in cl[5] then
```

```
>       return ct.irreducibles[5][5];
>       fi; end;
                function(n) ... end
gap >   e1 := [1,1,1,1,1,1,1,1,1,1,1,1];;
gap >   e2 := [1,-1,0,0,0,0,0,0,0,0,0,0];;
gap >   e3 := [0,0,1,0,0,0,0,0,0,-1,0,0];;
gap >   e4 := [0,0,0,1,0,0,0,0,0,0,-1,0];;
gap >   e5 := [0,0,0,0,1,0,0,0,0,0,0,-1];;
gap >   e6 := [0,0,0,0,0,1,0,-1,0,0,0,0];;
gap >   e7 := [0,0,0,0,0,0,1,0,-1,0,0,0];;
gap >   e8 := [1,1,0,0,0,0,-1,0,-1,0,0,0];;
gap >   e9 := [0,0,1,0,0,0,-1,0,-1,1,0,0];;
gap >   e10 := [0,0,0,1,0,0,-1,0,-1,0,1,0];;
gap >   e11 := [0,0,0,0,1,0,-1,0,-1,0,0,1];;
gap >   e12 := [0,0,0,0,0,1,-1,1,-1,0,0,0];;
gap >   for i in [1..12] do
>       z[i] := [ ]; for j in [1..60] do
>       z[i][j] := [ ]; for k in [1..12] do
>       z[i][j][k] := [ ]; od; od; od;
gap >   for i in [8..12] do
>       for j in [1..60] do
>       for k in [1..12] do
>       z[i][j][k] := chi[5][gl[j]] * e[i][k^gl[j]];
>       od; od; od;
gap >   Sum(z[8])/12 = e8; Sum(z[9])/12 = e9; Sum(z[10])/12 = e10;        (3.7)
            true
            true
            true
gap >   Sum(z[11])/12 = e11; Sum(z[12])/12 = e12;                         (3.8)
            true
            true
```

```
gap >   chi2 := function(n)
    >   if n^(-1) in cl[1] then
    >   return ct.irreducibles[2][1];
    >   elif n^(-1) in cl[2] then
    >   return ct.irreducibles[2][2];
    >   elif n^(-1) in cl[3] then
    >   return ct.irreducibles[2][3];
    >   elif n^(-1) in cl[4] then
    >   return ct.irreducibles[2][4];
    >   elif n^(-1) in cl[5] then
    >   return ct.irreducibles[2][5];
    >   fi; end;
            function(n) ... end
gap >   for i in [2..4] do
    >   for j in [1..60] do
    >   for k in [1..12] do
    >   z[i][j][k] := chi[2][gl[j]] * e[i][k^gl[j]];
    >   od; od; od;
gap >   Sum(z[2])/20; Sum(z[3])/20; Sum(z[4])/20;                    (3.9)
```

$$[1/2, -1/2, \beta, \beta, \beta, \beta, \beta, -\beta, -\beta, -\beta, -\beta, -\beta]$$
$$[\beta, -\beta, 1/2, \beta, -\beta, -\beta, \beta, \beta, -\beta, -1/2, -\beta, \beta]$$
$$[\beta, -\beta, \beta, 1/2, \beta, -\beta, -\beta, \beta, \beta, -\beta, -1/2, -\beta]$$

Finally, similar discussions on other groups of rotations can be found in [3, page 260].

References

[1] M. Schönert et al., GAP-*Groups, Algorithms, and Programming*, Version 3 Release 4, Lehrstuhl D für Mathematik, RWTH, Aachen, Germany, 1994.

[2] Serre, J.-P., *Linear Representations of Finite Groups*, Graduate Texts in Mathematics, Vol. 42, Springer-Verlag 1977. preprint.

[3] Kirillov, A.A., *Elements of the Theory of Representations*, Springer-Verlag 1976.

COMPUTING DIOPHANTINE APPROXIMATE GRÖBNER BASES

Hiroshi Sekigawa
Kiyoshi Shirayanagi
NTT Communication Science Laboratories
Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan
sekigawa@progn.kecl.ntt.jp, shirayan@progn.kecl.ntt.jp

1 Introduction

In computational algebra, as is well known, if we rely on exact computation to execute some algorithms, it can be very expensive. It thus would be very useful if we could apply some approximate approach. However, the problem with this is that we could not obtain a "reasonably" approximate result if we *simply* evaluated an original algorithm on an approximate input. That is, even if a sequence converges to a given input, the sequence of the outputs for the initial sequence does not necessarily converge to the true output. Let us call algorithms which have such instability, *unstable* algorithms. A typical example is Buchberger's algorithm ([Buc85]) for computing Gröbner bases.

In [Shi93], we proposed techniques for stabilizing Buchberger's algorithm with floating-point approximation. I.e. we provided an algorithm which computes Gröbner bases with floating-point coefficients. The motivation of this algorithm was to cope with the immense growth of intermediate coefficients from exact approach.

Later the idea behind these techniques has been extended by Shirayanagi and Sweedler ([SS]) in two directions. One direction aims at dealing with not only floating-point approximation but other approximations which meet some axioms. The other aims at dealing with not only Buchberger's algorithm but other algorithms which satisfy desired properties as well. These two generalizations have been systematized to a theory of stabilization.

The goal of our research is to apply the theory to various approximations and various algorithms, to verify the usefulness of the theory. As an example, this paper reports on some experimental results of applying the theory to the Diophantine approximation and Buchberger's algorithm. The importance of this experiment lies in avoiding the coefficient swell from exact approach when input coefficients are *not* rational, and verifying whether we really can obtain stability in this specified case by use of the theory of stabilization. Moreover, the Diophantine approximation approach to computing Gröbner bases is totally new.

In Section 2 we outline the algorithm which computes Diophantine approximate Gröbner bases, based upon the original Buchberger algorithm and our stabilization techniques. In Section 3 we show some experimental results of running this algorithm for some examples of polynomial systems with *irrational* coefficients. Finally in Section 4, we give concluding remarks with an open question.

Throughout this paper, $\mathbf{Q}$ and $\mathbf{R}$ denote the fields of the rationals and the reals, respectively.

2 Diophantine approximate Gröbner bases

In [Shi93] we proposed an algorithm called **FP-GB** which computes floating-point Gröbner bases. In this paper we propose an algorithm **DA-GB** which uses continued fraction (or the Diophantine) approximation instead of floating-point approximation.

We introduce *bracket coefficients* (BC's). A BC is based on an approximate value and its error. For an irrational number $\omega \in \mathbf{R}$, let $\frac{p_\mu}{q_\mu} \in \mathbf{Q}$ $((p_\mu, q_\mu) = 1, q_\mu > 0)$ be the result from the μ-th truncation of the simple continued fraction expansion of ω, where

$$|\omega - \frac{p_\mu}{q_\mu}| < \frac{1}{q_\mu^2}.$$

For example, the first, second, third truncations for π lead to $\frac{22}{7}, \frac{333}{106}, \frac{355}{113}$, respectively. We refer the reader to [HW79] or [Hua82] for the theory of continued fractions. We then define the BC for $\omega \in \mathbf{R}$ by $[\frac{p_\mu}{q_\mu}, \frac{1}{q_\mu^2}]$ if $\omega \notin \mathbf{Q}$, and $[\omega, 0]$ if $\omega \in \mathbf{Q}$. Let us call this μ which indicates the place of truncation, *precision* of the Diophantine approximation. Of course, as precision approaches infinity, the approximation part of the BC converges to the true value ω. BC is somewhat like an extension of *circular interval* in interval analysis (see [AH83]).

We write **DA-GB** with a specified precision μ, as $\mathbf{DA\text{-}GB}_\mu$. This algorithm mimics the traditional Buchberger algorithm (see [Buc85]). As is well known, it has many variations from the crude version to improved versions (see [Buc85, CLO92, BWK93] etc.). We can select any one of them as a template for $\mathbf{DA\text{-}GB}_\mu$'s to mimic. However, whatever is selected, for a practical reason, we will use the *non-division form* of it, as explained later.

Now $\mathbf{DA\text{-}GB}_\mu$ is outlined as follows. Input polynomials are in $\mathbf{R}[x_1, \ldots, x_n]$, that is, polynomials with real coefficients in $x_1, \ldots, x_n$.

R-to-BC Conversion. Convert each real coefficient E of each input polynomial into the bracket coefficient with precision μ, as defined above, to make the input BC polynomials.

BC Arithmetic. Perform arithmetic between BC's in a natural fashion:

$$[A,\alpha]+[B,\beta]=[A+B,\alpha+\beta],$$
$$[A,\alpha]-[B,\beta]=[A-B,\alpha+\beta],$$
$$[A,\alpha]\times[B,\beta]=[AB,\alpha\beta+|A|\beta+\alpha|B|].$$

Zero-Rewriting. For any bracket coefficient $[E,\epsilon]$ of any intermediate polynomial,

$$\text{rewrite } [E,\epsilon] \text{ into } \mathbf{0} \text{ if } |E|\leq\epsilon$$

in the course of the algorithm, where $\mathbf{0}=[0,0]$ such that $\mathbf{0}t=0$ for any power product t.

BC-to-Q Conversion. Finally pick up the approximation part ($\in \mathbf{Q}$) of each BC from the resulting set of BC polynomials to make a set of $\mathbf{Q}$ polynomials which is expected to be approximate to the true $\mathbf{R}$ Gröbner basis.

Roughly speaking, $\mathbf{DA\text{-}GB}_\mu$ is the original Buchberger algorithm with input BC polynomials with precision μ, BC arithmetic, and the zero-rewriting.

In [Shi93] a BC consisted of a pair of floating-point numbers and BC arithmetic was based on floating-point arithmetic. Thus it was necessary to compute rounding errors from floating-point arithmetic. In contrast, a BC in the present paper is simply made of a pair of rational numbers. Hence, BC arithmetic here just uses exact arithmetic in $\mathbf{Q}$ and need *not* compute rounding errors. That is, approximation is made only for an initial input but later computation simply follows exact arithmetic.

We avoid division between BC's since the error analysis for division is complicated. We then define S-polynomials and polynomial reductions without division as when working with polynomials having coefficients in a ring that is not a field. Given an admissible term ordering, $LP(f)$ denotes the leading power product of a BC polynomial f. If we write $f=f'+rest(f)$, then f' is a term of f and $rest(f)$ denotes the other terms of f.

S-BC polynomial. Let f and g be BC polynomials, $f=[A,\alpha]LP(f)+rest(f)$, $g=[B,\beta]LP(g)+rest(g)$. And let LCM be the least common multiple of $LP(f)$ and $LP(g)$. Then the S-BC polynomial of f and g is the BC polynomial

$$[B,\beta]\cdot\frac{LCM}{LP(f)}\cdot f-[A,\alpha]\cdot\frac{LCM}{LP(g)}\cdot g.$$

BC reduction. Let f and g be BC polynomials, $f=[A,\alpha]\cdot u\cdot LP(g)+rest(f)$ ($[A,\alpha]\neq\mathbf{0}$), $g=[B,\beta]LP(g)+rest(g)$, where u is a power product. Then, we say that f is BC reducible using g, and by setting

$$h=[B,\beta]\cdot f-[A,\alpha]u\cdot g,$$

we say that f BC reduces to h (using g).

To describe the property of **DA-GB** we shall define *supportwise convergence.* The *support* of a polynomial f is the set of power products of f with nonzero coefficients, denoted $Supp(f)$. The support of a finite set $F = \{f_1, \ldots, f_r\}$ of polynomials is $\{Supp(f_1), \ldots, Supp(f_r)\}$, denoted $Supp(F)$. A sequence $\{F_\mu\}_\mu$ of finite sets of polynomials supportwise converges to a finite set F of polynomials as μ approaches infinity iff $\{F_\mu\}_\mu$ coefficientwise converges to F, and there is an M such that $Supp(F_\mu) = Supp(F)$ for all $\mu \geq M$.

Let us call a pre-selected (non-division form of) Buchberger's algorithm which is a template for **DA-GB**$_\mu$'s to simulate, **R-GB**.

Theorem 1 *Given a finite subset F of $\mathbf{R}[x_1, \ldots, x_n]$. For each precision μ let G_μ be the output of* **DA-GB**$_\mu$ *for F. Let G be the output of* **R-GB** *for F, which is a true Gröbner basis of F. Then G_μ supportwise converges to G as μ approaches infinity.*

The strict proof is the same as the proof in [Shi93, Shi] except that it uses the Diophantine approximation instead of floating-point approximation. Let us give a sketch of the proof.

Sketch of the Proof. First of all we should remark the two fundamental properties of BC arithmetic: error propagation and error convergence. That is, let E be the true value computed by a finite number of steps of addition, subtraction, multiplication by exact arithmetic, from input coefficients in $\mathbf{R}$. Let $[E_\mu, \epsilon_\mu]$ be the associated BC computed by BC arithmetic from the associated input BC's. Then, (1) $|E_\mu - E| \leq \epsilon_\mu$ and (2) $\epsilon_\mu \to 0$ as $\mu \to \infty$.

Second, let us compare the behavior of **R-GB** with that of **DA-GB**. More specifically, let us compare an arbitrary coefficient E (of an intermediate polynomial) in the course of **R-GB** with the associated BC $[E_\mu, \epsilon_\mu]$ in **DA-GB**$_\mu$. If $E = 0$, then from (1) we have $|E_\mu| \leq \epsilon_\mu$, and so by the zero-rewriting $[E_\mu, \epsilon_\mu]$ is rewritten to **0**. The problem is that we have $E \neq 0$ but mistakenly may rewrite $[E_\mu, \epsilon_\mu]$ to **0** from $|E_\mu| \leq \epsilon_\mu$ at some precision μ. But by (2) we can prove that there is a precision M' such that $\mu \geq M' \Rightarrow |E_\mu| > \epsilon_\mu$, and so eventually we will *not* rewrite $[E_\mu, \epsilon_\mu]$ to **0**. Consequently, we will obtain the same support. Once we have the same support, after that by (1) and (2) again we will obtain coefficientwise convergence. ■

3 Experimental Results

As **R-GB** we selected an improved version of Buchberger's Algorithm ([Buc85], Algorithm 6.3, page 196), and we implemented **DA-GB** based on it in Maple V Release 3 ([Maple92]).

As examples, we took two kinds of polynomial systems. One is a small example (from [Shi93]) just for observing supportwise convergence as precision increases.

Below e is Napier's number (2.71828...), and *tdeg* (resp. *plex*) denotes total degree inverse lexicographic (resp. purely lexicographic) term ordering.

Example 1 $F = \{\sqrt{2}ex^3y + \sqrt{3}xy + \frac{\sqrt{7}}{e}, \frac{\sqrt{3}}{e}x^2y^2 - \sqrt{7}xy + \frac{e\sqrt{11}}{11}\} \subset \mathbf{R}[x,y]$, tdeg with $x > y$.

The other is examples which are various modifications of famous benchmark examples of Gröbner bases: Cyclic systems ([Sugar91]) and Katsura systems ([BGK86]). We modified these systems to have irrational coefficients since if input coefficients were all rational, then all the associated BC's would have 0 as the errors and so **DA-GB** would essentially be equivalent to **R-GB**. For reference let us display one example (which seems to be the most complicated) for each family of the modified systems.

Modified Cyclic3's: $C31, \ldots, C35 \subset \mathbf{R}[c_0, c_1, c_2]$
e.g. $C35 = \{\frac{\pi+\sqrt{7}}{e^2}c_0c_1c_2 - 1, c_0c_1 + (\pi + e)c_0c_2 + c_1c_2, c_0 + c_1 + \frac{e+\sqrt{11}}{\pi}c_2\}$,
plex with $c_0 > c_1 > c_2$.

Modified Cyclic4's: $C41, \ldots, C45 \subset \mathbf{R}[c_0, c_1, c_2, c_3]$
e.g. $C45 = \{\pi c_0c_1c_2c_3 - 1, c_0c_1c_2 + ec_0c_1c_3 + c_0c_2c_3 + c_1c_2c_3,$
$c_0c_1 + c_0c_3 + \sqrt{2}c_1c_2 + c_2c_3, c_0 + c_1 + c_2 + \sqrt{3}c_3\}$, tdeg with $c_0 > c_1 > c_2 > c_3$.

Modified Katsura2's: $K21, \ldots, K25 \subset \mathbf{R}[u_0, u_1, u_2]$
e.g. $K25 = \{(\sqrt{5} - \pi)eu_0^2 - u_0 + 2u_1^2 + 2u_2^2, 2u_0u_1 + (\pi^2e - \sqrt{11})u_1u_2 - u_1,$
$u_0 + 2u_1 + (e^3 - \pi^2)u_2 - 1\}$, plex with $u_0 > u_1 > u_2$.

Modified Katsura3's: $K31, \ldots, K35 \subset \mathbf{R}[u_0, u_1, u_2, u_3]$
e.g. $K35 = \{\frac{\pi+e^2}{\sqrt{11}}u_0^2 - u_0 + 2u_1^2 + 2u_2^2 + 2u_3^2, 2u_0u_1 + \frac{\sqrt{2}+\sqrt{3}}{\sqrt{5}}u_1u_2 + 2u_2u_3 - u_1,$
$2u_0u_2 + u_1^2 + 2u_1u_3 - u_2, u_0 + 2u_1 + 2u_2 + 2u_3 - 1\}$,
tdeg with $u_0 > u_1 > u_2 > u_3$.

In the exact method, i.e. **R-GB**, when evaluating expressions involving irrationals and judging them as zero, we used the "simplify" function of Maple, which applies some simplification rules to given expressions. In **DA-GB**, when converting input coefficients into BC's, we used the "numtheory" package of Maple for the Diophantine approximation. As a convention, we let **DA-GB** stop immediately when any product of two nonzero bracket coefficients is judged to be zero. (For example, $[1, 1/2]^2 = [1, 5/4]$ is judged as zero, since $|1| < 5/4$.) This is because such a product cannot be zero and the final result from the assumption that it is zero should be unreliable. Let us call this stop a *zero-stop*.

Experiments were carried out on HP 9000/735.

Table 1 shows how $G_\mu = \mathbf{DA\text{-}GB}_\mu(F)$ changes as μ increases from 1 to 10, with cpu times and zero-rewriting numbers. Here a zero-rewriting number means the number of cases where *actually* a BC $[E, \epsilon]$ was rewritten to zero by the fact of $|E| \leq \epsilon$.

To preserve space, each rational coefficient of G_μ is represented by floating-point with digits 10.

Tables 2 and 3 show the experimental results for the modified Cyclic systems and Katsura systems. To preserve space, each table only shows the supports. For $\mathbf{DA\text{-}GB}_\mu(F)$ we specified μ as 10 or 20. "object too large" is a Maple error message indicating that the number of "terms" used are too large and computation exits.

In this experiment, we had $Supp(G_\mu) = Supp(G)$ when both of the computation of $\mathbf{DA\text{-}GB}_\mu(F)$ and $\mathbf{R\text{-}GB}(F)$ terminated without an error or a zero-stop.

4 Conclusion

From the experimental results which we have obtained thus far, **DA-GB** gives the correct supports with relatively low precision. Of course, for more complex examples it will need higher precision. An important open question here is a theoretical analysis about *lucky* precisions where the correct support will be obtained.

As a conclusion, in particular when input coefficients involve complicated irrational numbers, the Diophantine approximation method can be useful, since the exact method cannot give results within reasonable time and memory, from the swell of intermediate coefficients in **R**. Furthermore when irrational coefficients include transcendental numbers, it is the more so. On the other hand, the Diophantine approximation method also uses exact arithmetic in **Q** for BC arithmetic, and thus as precision becomes too large, the growth of numerators and denominators in BC's can be huge. As is well known, such a deficiency is seen also in the exact computation of the original Buchberger algorithm over the rationals. Therefore we cannot expect efficiency of **DA-GB** for very large examples. In fact **FP-GB** in [Shi93] which uses the floating-point method is often much more efficient than **DA-GB**.

Rather than efficiency, however, we want to emphasize stability. Namely, the theory of stabilization has been verified to be effective not only for floating-point approximation but also the Diophantine approximation. We believe that this paper can be a step toward verification of the usefulness of the theory for various approximations and various algorithms in computational algebra.

Table 1: Example 1

μ	floating-point expression of G_μ with digits 10	N	T
1	**zero-stop**	0	.85
2	**zero-stop**	0	.33
3	**zero-stop**	1	.67
4	**zero-stop**	4	.62
5	**zero-stop**	10	1.3
6	$\{-.7945565578 \times 10^{125}x^2 - .1938503422 \times 10^{124}y^2 - .9563057511 \times 10^{125},$ $-.7649721545 \times 10^{78}xy - .9481460080 \times 10^{77}y^2 + .2490022374 \times 10^{78},$ $.8763273882 \times 10^{124}x + .1070414296 \times 10^{124}y^3 + .3131649828 \times 10^{125}y\}$	13	1.4
7	$\{-.6701490085 \times 10^{267}x^2 - .1634877470 \times 10^{266}y^2 - .8067715905 \times 10^{267},$ $-.1140790108 \times 10^{150}xy - .1413868354 \times 10^{149}y^2 + .3712324881 \times 10^{149},$ $-.7337872250 \times 10^{116}x - .8964695535 \times 10^{115}y^3 - .2623676704 \times 10^{117}y\}$	12	2.1
8	$\{-.8560635885 \times 10^{391}x^2 - .2088409551 \times 10^{390}y^2 - .1030547921 \times 10^{392},$ $-.5115412164 \times 10^{94}xy - .6339823994 \times 10^{93}y^2 + .1664704758 \times 10^{94},$ $.2298225662 \times 10^{153}x + .2807591996 \times 10^{152}y^3 + .8216720314 \times 10^{153}y\}$	14	2.2
9	$\{-.1213310764 \times 10^{356}x^2 - .2959950552 \times 10^{354}y^2 - .1460620029 \times 10^{356},$ $-.5520339030 \times 10^{200}xy - .6841726597 \times 10^{199}y^2 + .1796467755 \times 10^{200},$ $-.1753817134 \times 10^{160}x - .2142559371 \times 10^{159}y^3 - .6270413484 \times 10^{160}y\}$	14	3.1
10	$\{-.8346625917 \times 10^{304}x^2 - .2036217693 \times 10^{303}y^2 - .1004789630 \times 10^{305},$ $.2256307567 \times 10^{259}xy + .2796394645 \times 10^{258}y^2 - .7342658493 \times 10^{258},$ $-.1938462839 \times 10^{192}x - .2368129007 \times 10^{191}y^3 - .6930537252 \times 10^{192}y\}$	14	2.2

N: zero-rewriting number
T: cpu time (sec)

Notes.

1. floating-point expression of $G = \mathbf{R\text{-}GB}(F)$ with digits 10 (cpu time: 40.5sec)
   $\{ -3.076101960x^2 - .07504363785y^2 - 3.703099133,$
   $-13.44535658xy - 1.666373738y^2 + 4.375492414,$
   $-3.233915437x - .3950725311y^3 - 11.56215932y\}$
2. $Supp(G_\mu) = Supp(G)$ for $6 \leq \mu \leq 10$.

Table 2: Modified Cyclic systems

	DA-GB (approximate)				R-GB (exact)	
	μ	$Supp(G_\mu)$	N	T	$Supp(G)$	T
$C31$	10	$\{\{c_0, c_2, c_2^4\}, \{c_1, c_2, c_2^4\}, \{1, c_2^3, c_2^6\}\}$	4	.63	$\{\{c_0, c_2, c_2^4\}, \{c_1, c_2, c_2^4\}, \{1, c_2^3, c_2^6\}\}$	3.0
$C32$	10	$\{\{c_0, c_2, c_2^4\}, \{c_1, c_2, c_2^4\}, \{1, c_2^3, c_2^6\}\}$	6	.43	$\{\{c_0, c_2, c_2^4\}, \{c_1, c_2, c_2^4\}, \{1, c_2^3, c_2^6\}\}$	1.1
$C33$	10	$\{\{c_0, c_1, c_2\}, \{c_1c_2, c_1^2, c_2^2\}, \{1, c_2^3\}\}$	1	.32	$\{\{c_0, c_1, c_2\}, \{c_1c_2, c_1^2, c_2^2\}, \{1, c_2^3\}\}$	.20
$C34$	10	$\{\{c_0, c_2, c_2^4\}, \{c_1, c_2, c_2^4\}, \{1, c_2^3, c_2^6\}\}$	4	.67	$\{\{c_0, c_2, c_2^4\}, \{c_1, c_2, c_2^4\}, \{1, c_2^3, c_2^6\}\}$	2.1
$C35$	10	$\{\{c_0, c_2, c_2^4\}, \{c_1, c_2, c_2^4\}, \{1, c_2^3, c_2^6\}\}$	4	.92	object too large	1046.7
$C41$	10	$\{\{c_2^2c_3^4, c_1c_2, c_1c_3, c_2c_3, c_3^2\}, \{c_2^3c_3^2, c_2^2c_3^3, c_2, c_3\}, \{c_2^2c_3^6, c_2^2c_3^2, c_3^4, 1\} \{c_0, c_1, c_2, c_3\}, \{c_1^2, c_1c_3, c_3^2\}, \{c_1c_3^4, c_3^5, c_1, c_3\}\}$	100	5.8	$\{\{c_2^2c_3^4, c_1c_2, c_1c_3, c_2c_3, c_3^2\}, \{c_2^3c_3^2, c_2^2c_3^3, c_2, c_3\}, \{c_2^2c_3^6, c_2^2c_3^2, c_3^4, 1\} \{c_0, c_1, c_2, c_3\}, \{c_1^2, c_1c_3, c_3^2\}, \{c_1c_3^4, c_3^5, c_1, c_3\}\}$	4.3
$C42$	10	$\{\{c_1c_2c_3^2, c_1c_3^3, c_2^2c_3^2, c_2c_3^3, c_3^4, 1\}, \{c_2^2c_3^4, c_1c_2, c_1c_3, c_2c_3, c_3^2\}, \{c_1c_2^2, c_1c_3^2, c_2^2c_3, c_3^3\} \{c_2^3c_3^2, c_2^2c_3^3, c_2, c_3\}, \{c_0, c_1, c_2, c_3\} \{c_1c_3^4, c_3^5, c_1, c_3\}, \{c_1^2, c_1c_3, c_3^2\}\}$	24	.97	$\{\{c_1c_2c_3^2, c_1c_3^3, c_2^2c_3^2, c_2c_3^3, c_3^4, 1\}, \{c_2^2c_3^4, c_1c_2, c_1c_3, c_2c_3, c_3^2\}, \{c_1c_2^2, c_1c_3^2, c_2^2c_3, c_3^3\} \{c_2^3c_3^2, c_2^2c_3^3, c_2, c_3\}, \{c_0, c_1, c_2, c_3\} \{c_1c_3^4, c_3^5, c_1, c_3\}, \{c_1^2, c_1c_3, c_3^2\}\}$	1.4
$C43$	10	$\{1\}$	11	1.1	$\{1\}$	2.8
$C44$	20	$\{\{c_2^3, c_3c_2^2, c_3^3, c_3^2c_2, c_3^2c_1\}, \{1, c_2^2c_3^2, c_3^3c_1, c_3^3c_2, c_3^4\}, \{c_1c_3, c_3c_2, c_3^2, c_2^2, c_1^2\}, \{c_1c_3, c_3c_2, c_3^2, c_2^2, c_1c_2\}, \{c_1c_3, c_3c_2, c_3^2, c_2^2, c_3^6\}, \{c_3, c_1, c_2, c_1c_3^4, c_3^5\}, \{c_3, c_1, c_2, c_2c_3^4, c_3^5\}, \{c_3, c_1, c_0, c_2\}\}$	65	935.2	$\{\{c_2^3, c_3c_2^2, c_3^3, c_3^2c_2, c_3^2c_1\}, \{1, c_2^2c_3^2, c_3^3c_1, c_3^3c_2, c_3^4\}, \{c_1c_3, c_3c_2, c_3^2, c_2^2, c_1^2\}, \{c_1c_3, c_3c_2, c_3^2, c_2^2, c_1c_2\}, \{c_1c_3, c_3c_2, c_3^2, c_2^2, c_3^6\}, \{c_3, c_1, c_2, c_1c_3^4, c_3^5\}, \{c_3, c_1, c_2, c_2c_3^4, c_3^5\}, \{c_3, c_1, c_0, c_2\}\}$	1753.5
$C45$	20	$\{\{c_3c_2, c_3c_1, c_2^2, c_3^2, c_1^2\}, \{c_2c_3^4, c_3^5, c_1, c_3, c_2\}, \{c_3c_2, c_3c_1, c_2^2, c_3^2, c_3^6\}, \{c_1c_3^2, c_3c_2^2, c_3^3, c_2^3, c_3^2c_2\}, \{1, c_3^4, c_3^2c_2^2, c_3^3c_1, c_2c_3^3\}, \{c_3c_2, c_3c_1, c_2^2, c_1c_2, c_3^2\}, \{c_3^5, c_1, c_3, c_2, c_1c_3^4\}, \{c_1, c_0, c_3, c_2\}\}$	57	663.8	object too large	6323.9

N: zero-rewriting number
T: cpu time (sec)

Table 3: Modified Katsura systems

	DA-GB (approximate)			R-GB (exact)	
	$Supp(G_\mu)$	N	T	$Supp(G)$	T
$K21$	$\{\{1,u_0,u_2,u_2^2,u_2^3\}$, $\{1,u_1,u_2,u_2^2,u_2^3\}$, $\{1,u_2,u_2^2,u_2^3,u_2^4\}\}$	12	1.3	$\{\{1,u_0,u_2,u_2^2,u_2^3\}$, $\{1,u_1,u_2,u_2^2,u_2^3\}$, $\{1,u_2,u_2^2,u_2^3,u_2^4\}\}$	14.3
$K22$	$\{\{1,u_0,u_2,u_2^2,u_2^3\}$, $\{u_1,u_2,u_2^2,u_2^3\}$, $\{u_2^4,u_2,u_2^2,u_2^3\}\}$	7	1.9	$\{\{1,u_0,u_2,u_2^2,u_2^3\}$, $\{u_1,u_2,u_2^2,u_2^3\}$, $\{u_2^4,u_2,u_2^2,u_2^3\}\}$	12.1
$K23$	$\{\{1,u_2^4,u_2,u_2^2,u_2^3\}$, $\{1,u_0,u_2,u_2^2,u_2^3\}$, $\{1,u_1,u_2,u_2^2,u_2^3\}\}$	12	1.8	$\{\{1,u_2^4,u_2,u_2^2,u_2^3\}$, $\{1,u_0,u_2,u_2^2,u_2^3\}$, $\{1,u_1,u_2,u_2^2,u_2^3\}\}$	225.7
$K24$	$\{\{1,u_1,u_2,u_2^2,u_2^3\}$, $\{1,u_0,u_2,u_2^2,u_2^3\}$, $\{1,u_2,u_2^2,u_2^3,u_2^4\}\}$	12	6.1	object too large	13024.1
$K25$	$\{\{1,u_1,u_2,u_2^2,u_2^3\}$, $\{1,u_0,u_2,u_2^2,u_2^3\}$, $\{1,u_2,u_2^2,u_2^3,u_2^4\}\}$	8	6.4	object too large	7896.3
$K31$	$\{\{1,u_0,u_1,u_3,u_2\}$, $\{1,u_1,u_3,u_2,u_2^2,u_3^2,u_2u_3,u_1u_3\}$, $\{1,u_1,u_3,u_2,u_3^2,u_2u_3,u_1u_3,u_3^4,u_3^3\}$, $\{1,u_1,u_3,u_2,u_3^2,u_2u_3,u_1u_3,u_3^3,u_3^2u_1\}$, $\{1,u_3^2u_2,u_1,u_3,u_2,u_3^2,u_2u_3,u_1u_3,u_3^3\}$, $\{1,u_1,u_3,u_2,u_3^2,u_1u_2,u_2u_3,u_1u_3\}$, $\{1,u_1,u_3,u_2,u_1^2,u_3^2,u_2u_3,u_1u_3\}\}$	41	43.5	$\{\{1,u_0,u_1,u_3,u_2\}$, $\{1,u_1,u_3,u_2,u_2^2,u_3^2,u_2u_3,u_1u_3\}$, $\{1,u_1,u_3,u_2,u_3^2,u_2u_3,u_1u_3,u_3^4,u_3^3\}$, $\{1,u_1,u_3,u_2,u_3^2,u_2u_3,u_1u_3,u_3^3,u_3^2u_1\}$, $\{1,u_3^2u_2,u_1,u_3,u_2,u_3^2,u_2u_3,u_1u_3,u_3^3\}$, $\{1,u_1,u_3,u_2,u_3^2,u_1u_2,u_2u_3,u_1u_3\}$, $\{1,u_1,u_3,u_2,u_1^2,u_3^2,u_2u_3,u_1u_3\}\}$	141.3
$K32$	$\{\{1,u_2u_3,u_1u_3,u_1,u_3,u_2,u_3^2,u_3^3,u_3^2u_2\}$, $\{1,u_1u_2,u_2u_3,u_1u_3,u_1,u_3,u_2,u_3^2\}$, $\{1,u_2u_3,u_1u_3,u_1,u_3,u_2,u_3^2,u_3^3,u_3^4\}$, $\{1,u_2u_3,u_1u_3,u_1,u_3,u_2,u_3^2,u_3^3,u_3^2u_1\}$, $\{1,u_0,u_1,u_3,u_2\}$, $\{1,u_2u_3,u_1u_3,u_1,u_3,u_2,u_2^2,u_3^2\}$, $\{1,u_2u_3,u_1u_3,u_1,u_3,u_2,u_1^2,u_3^2\}\}$	40	130.3	$\{\{1,u_2u_3,u_1u_3,u_1,u_3,u_2,u_3^2,u_3^3,u_3^2u_2\}$, $\{1,u_1u_2,u_2u_3,u_1u_3,u_1,u_3,u_2,u_3^2\}$, $\{1,u_2u_3,u_1u_3,u_1,u_3,u_2,u_3^2,u_3^3,u_3^4\}$, $\{1,u_2u_3,u_1u_3,u_1,u_3,u_2,u_3^2,u_3^3,u_3^2u_1\}$, $\{1,u_0,u_1,u_3,u_2\}$, $\{1,u_2u_3,u_1u_3,u_1,u_3,u_2,u_2^2,u_3^2\}$, $\{1,u_2u_3,u_1u_3,u_1,u_3,u_2,u_1^2,u_3^2\}\}$	179.7
$K33$	$\{\{1,u_3^3,u_3^4,u_1,u_3,u_2,u_3^2,u_2u_3,u_1u_3\}$, $\{1,u_3^3,u_3^2u_2,u_1,u_3,u_2,u_3^2,u_2u_3,u_1u_3\}$, $\{1,u_0,u_1,u_3,u_2\}$, $\{1,u_3^3,u_1,u_3,u_2,u_3^2,u_2u_3,u_1u_3,u_3^2u_1\}$, $\{1,u_1,u_3,u_2,u_3^2,u_1u_2,u_2u_3,u_1u_3\}$, $\{1,u_1,u_3,u_2,u_2^2,u_3^2,u_2u_3,u_1u_3\}$, $\{1,u_1,u_3,u_2,u_1^2,u_3^2,u_2u_3,u_1u_3\}\}$	40	431.7	object too large	19323.8
$K34$	$\{\{1,u_3^3,u_2u_3,u_1u_3,u_1,u_3,u_2,u_3^2,u_3^2u_2\}$, $\{1,u_3^3,u_2u_3,u_1u_3,u_1,u_3,u_2,u_3^2,u_3^2u_1\}$, $\{1,u_3^3,u_2u_3,u_1u_3,u_1,u_3,u_2,u_3^2,u_3^4\}$, $\{1,u_0,u_1,u_3,u_2\}$, $\{1,u_2u_3,u_1u_3,u_1,u_3,u_2,u_2^2,u_3^2\}$, $\{1,u_2u_3,u_1u_3,u_1,u_3,u_2,u_1^2,u_3^2\}$, $\{1,u_2u_3,u_1u_3,u_1,u_3,u_2,u_3^2,u_2u_1\}\}$	40	661.5	object too large	5580.2
$K35$	$\{\{1,u_1,u_3,u_2,u_1^2,u_3^2,u_2u_3,u_1u_3\}$, $\{1,u_1,u_3,u_2,u_3^2,u_2u_3,u_1u_3,u_2u_1\}$, $\{1,u_1,u_3,u_2,u_3^2,u_2u_3,u_1u_3,u_3^2u_1,u_3^3\}$, $\{1,u_1,u_3,u_2,u_3^2,u_2u_3,u_1u_3,u_3^4,u_3^3\}$, $\{1,u_1,u_3,u_2,u_3^2,u_2u_3,u_1u_3,u_3^2u_2,u_3^3\}$, $\{1,u_1,u_3,u_2,u_2^2,u_3^2,u_2u_3,u_1u_3\}$, $\{1,u_0,u_1,u_3,u_2\}\}$	41	815.0	object too large	3535.0

N: zero-rewriting number
T: cpu time (sec)
$\mu = 10$ for $K21,\ldots,K25$.
$\mu = 20$ for $K31,\ldots,K35$.

References

[AH83] Alefeld, G. and Herzberger, J., *Introduction to Interval Computations, Computer Science and Applied Mathematics, Academic Press* (1983).

[BWK93] Becker, T., Weispfenning, V., and Kredel, H., *Gröbner Bases: A Computational Approach to Commutative Algebra, Springer-Verlag* (1993).

[BGK86] Boege, W., Gebauer, R., and Kredel, H., Some Examples for Solving Systems of Algebraic Equations by Calculating Groebner Bases, *J. Symb. Comp.* **1** (1986), 83-98.

[Buc85] Buchberger, B., Gröbner Bases: An Algorithmic Method in Polynomial Ideal Theory, *Multidimensional Systems Theory (ed. N. K. Bose), D. Reidel Publishing Company* (1985), 184-232.

[CLO92] Cox, D., Little, J., and O'Shea D., *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra, Springer-Verlag* (1992).

[Maple92] Char, B. W., Geddes, K. O., Gonnet, G. H., Leong, B. L., Monagan, M. B., and Watt, S. M., *First Leaves: A Tutorial Introduction to Maple V, Springer-Verlag* (1992).

[Sugar91] Giovini, A., Mora T., Niesi G., Robbiano L., and Traverso C., "One sugar cube, please" OR Selection strategies in the Buchberger algorithm, *Proc. of the 1991 Int. Symp. on Symbolic and Algebraic Computation (ISSAC '91), ACM Press* (1991), 49-54.

[HW79] Hardy, G. H., and Wright, E. M., *An introduction to the theory of numbers (Fifth edition), Oxford University Press* (1979).

[Hua82] Hua, L. K., *Introduction to Number Theory, Springer-Verlag* (1982).

[Shi93] Shirayanagi, K., An Algorithm to Compute Floating Point Gröbner Bases, *Mathematical Computation with Maple V: Ideas and Applications (ed. T. Lee), Birkhäuser* (1993), 95-106.

[Shi] Shirayanagi, K., Floating Point Gröbner Bases, *Mathematics and Computers in Simulation, North-Holland, to appear.*

[SS] Shirayanagi, K. and Sweedler, M., A Theory of Stabilizing Algebraic Algorithms, *preprint.*

ZALGALLER POLYHEDRA AND COMPUTER ALGEBRA

Jiro Sekiguchi
Department of Mathematics,
Himeji Institute of Technology
Himeji 671-22, JAPAN

A convex polyhedron M is called a *Zalgaller polyhedron* if every face of M is a regular polygon. In this note, we call a *Z-polyhedron* instead of a Zalgaller polyhedron for short. Needless to say, regular and semiregular polyhedra are Z-polyhedra.

It is shown in V.A.Zalgaller [15] that there are 92 kinds of Z-polyhedra in addition to regular polyhedra, semiregular polyhedra. You can also find all the 92 Z-polyhedra in [6].

We consider two Z-polyhedra M and M'. Assume that there are a face F of M and a face F' of M' mutually congruent. Then, gluing F and F' together, we obtain a new polyhedra L. The figure L is called a composition of M and M' (gluing F and F' together) and both M and M' are composants of L. Sometimes L is also a Z-polyhedron. A Z-polyhedron is indecomposable if it is not obtained as a composition of two other Z-polyhedra. On the other hand, a Z-polyhedron is decomposable if it is obtained as a composition of two other Z-polyhedra. In [15], V. A. Zalgaller classified all the indecomposable Z-polyhedra and determined composants of decomposable Z-polyhedra. It suffices to treat indecomposable ones for our purpose. There are 28 indecomposable Z-polyhedra in addition to prisms and anti-prisms. The 28 Z-polyhedra in question are denoted by M_1-M_{28} in [15]. In Table 1, we classify these Z-polyhedra into four kinds:

- (Z1) regular polyhedra,
- (Z2) semiregular polyhedra,
- (Z3) composants of (semi)regular polyhedra
- (Z4) the remaining Z-polyhedra

Table 1: Z-polyhedra

(Z1)	M_1, M_{15}
(Z2)	M_{10}, M_{11}, M_{12}, M_{16}, M_{17}, M_{18}, M_{19}, M_{26}, M_{27}
(Z3)	M_2, M_3, M_4, M_5, M_6, M_7, M_9, M_{13}, M_{14}
(Z4)	M_8, M_{20}, M_{21}, M_{22}, M_{23}, M_{24}, M_{25}, M_{28}

Our interest is to construct figures in $\mathbf{R}^3$ which are a kind of stellations of Z-polyhedra.

Definition 1 *Let M be a Z-polyhedron with m faces denoted by $F_1, \cdots, F_m$. A figure N consisting of m faces $F'_1, \cdots, F'_m$ is a quasi-Zalgaller polyhedron (belonging to M) if the following conditions hold for $F'_1, \cdots, F'_m$. (In the sequel, we call a QZ-polyhedron instead of a quasi-Zalgaller polyhedron for short.)*

(i) If F_j is a regular polygon with p vertices, then F'_j is a regular polygon with p vertices or a regular star polygon with p vertices ($j = 1, \cdots, m$).

(ii) The configuration relation of $F'_1, \cdots, F'_m$ is same as that of $F_1, \cdots, F_m$. Namely, if F_j and F_k have a common edge, so do F'_j and F'_k. In particular, all the edges of faces F'_j ($j = 1, \cdots, m$) have the same length.

It is then natural to ask the problem:

Problem 1 *Find and classify all QZ-polyhedra.*

I believe that if M is a Z-polyhedron contained in one of (Z1), (Z2), (Z3), all QZ-polyhedra belonging to M are uniform polyhedra. (For the definition of a uniform polyhedron, see [14].) Noting this, we focus our attention to eight Z-polyhedra contained in (Z4). I formulated an algebraic method finding QZ-polyhedra in [10] and actually constructed some QZ-polyhedra belonging to each of Z-polyhedra in (Z4) (cf. [8], [10], [11]). These seem not to be found in literatures.

We are now going to define Minkowski-Zalgaller polyhedra. For this purpose, we consider a 3-dimensional affine space $\mathbf{R}^3$ with coordinate system (x, y, z). For two points $\mathrm{P}(x, y, z)$, $\mathrm{Q}(x', y', z')$, we put

$$|PQ|_{\mathrm{M}} = \sqrt{(x - x')^2 + (y - y')^2 - (z - z')^2} \tag{1}$$

Then $|PQ|_{\mathrm{M}}$ is called a Minkowski metric. Sometimes $|PQ|_{\mathrm{M}}$ becomes 0 or pure imaginary. We call the affine space $\mathbf{R}^3$ with the "distance" $|\cdot|_{\mathrm{M}}$ *the Minkowski space* and denote it by $\mathbf{M}$. It is possible to define a regular p-gon with respect to $|\cdot|_{\mathrm{M}}$). For example, a convex polygon S with five vertices P_1, P_2, P_3, P_4, P_5 is a pentagon in the Minkowski space $\mathbf{M}$ if

$$\begin{aligned} |\mathrm{P}_1\mathrm{P}_2|_{\mathrm{M}} = |\mathrm{P}_2\mathrm{P}_3|_{\mathrm{M}} = |\mathrm{P}_3\mathrm{P}_4|_{\mathrm{M}} = |\mathrm{P}_4\mathrm{P}_5|_{\mathrm{M}} = |\mathrm{P}_5\mathrm{P}_1|_{\mathrm{M}} \\ |\mathrm{P}_1\mathrm{P}_3|_{\mathrm{M}} = |\mathrm{P}_3\mathrm{P}_5|_{\mathrm{M}} = |\mathrm{P}_5\mathrm{P}_2|_{\mathrm{M}} = |\mathrm{P}_2\mathrm{P}_4|_{\mathrm{M}} = |\mathrm{P}_4\mathrm{P}_1|_{\mathrm{M}} \end{aligned} \tag{2}$$

If the convexity condition is forgotten, we also obtain a pentagram in $\mathbf{M}$ from a solution of (2).

We are going to define a Minkowski-Zalgaller polyhedron modifying Definition 1.

Definition 2 *Let M be a Z-polyhedron with m faces denoted by $F_1, \cdots, F_m$. A figure N in the Minkowski space consisting of m faces $F'_1, \cdots, F'_m$ is a Minkowski-Zalgaller polyhedron (belonging to M) if the following conditions hold for $F'_1, \cdots, F'_m$. (In the sequel, we call a MZ-polyhedron instead of a Minkowski-Zalgaller polyhedron for short.)*

(i) If F_j is a regular polygon with p vertices, then F'_j is a regular polygon with p vertices or a regular star polygon with p vertices in the Minkowski space ($j = 1, \cdots, m$).

(ii) The configuration relation of $F'_1, \cdots, F'_m$ is same as that of $F_1, \cdots, F_m$. Namely, if F_j and F_k have a common edge, so do F'_j and F'_k. In particular, all the edges of faces F'_j ($j = 1, \cdots, m$) have the same length with respect to the distance $|\cdot|_{\mathrm{M}}$.

It is also natural to ask the problem below as in QZ-polyhedra case:

Problem 2 *Find and classify all MZ-polyhedra.*

Since it is easy to prove that there is no MZ-polyhedron belonging to Tetrahedron, it is not clear to construct an example of MZ-polyhedra. But, by applying the algebraic method, we safely constructed a lot of QZ-polyhedra and MZ-polyhedra belonging to Z-polyhedra in (Z4) (cf. [4], [8], [10], [11]). The number of these polyhedra are collected in Table 2.

Table 2: QZ-polyhedra and MZ-polyhedra

	M_8	M_{20}	M_{21}	M_{22}	M_{23}	M_{24}	M_{25}	M_{28}	Total
Number of QZ-polyhedra	2	2	16	4	8	4	3	4	43
Number of MZ-polyhedra	0	0	8	2	2	4	2	4	22

We count Z-polyhedra as QZ-polyhedra in Table 2.

The readers who are interested in the details of numerical data for coordinates of vertives of QZ-polyhedra and MZ-polyhedra also refer to [4], [8], [10], [11].

I am very indebted to symbolic computation soft ware REDUCE3.4, Asir and *Mathematica*. It is easy for *Mathematica* users to draw figures in three dimensions from the numerical data in question.

Closing this note, we remark that Problems 1, 2 are still open.

References

[1] M. Noro and T. Shimomura, *Asir* User's Manual, 1994.

[2] S. Hitotsumatsu, '*Solve Regular Polyhedra*', in Japanese, Tokai Daigaku Shuppankai.

[3] S. Hitotsumatsu, '*Regular Polyhedra in Higher Dimensions*', in Japanese, Nihon Hyoronsha.

[4] K. Hori, Minkowski Space Analogue of Zalgaller Polyhedra, Graduation Thesis, Uninversity of Electro-Com. (1995).

[5] S. Kitahara, Computer algebra, Quasi-Zalgaller polyhedra and Graphics, Graduation Thesis, Uninversity of Electro-Com. (1994).

[6] M. Kobayashi and T. Suzuki, 'Calculation of coordinates of vertices of all convex polyhedra with regular faces,' (in Japanese) Bulletin of Univ. of Electro-Comm., vol.5 (1992), 147-184.

[7] S. Wolfram, '*Mathematica*' (Second Edition) Addison-Wesley Publishing Company.

[8] M. Noro and J. Sekiguchi, 'Quasi-Zalgaller polyhedra belonging to M_{21}', preprint.

[9] A. C. Hearn, '*REDUCE User's Manual*', The Rand Corporation, Santa Monica (1987).

[10] J. Sekiguchi, 'Non-convex Analogues of Sphenocorona, Sphenomegacorona, Hebesphenomegacorona and Disphenocingulum,' *Sūshikishori*, **3** No.1 (1994), 17-36.

[11] J. Sekiguchi, 'Non-convex Analogues of Zalgaller polyhedra M_8, M_{20}, M_{25}, M_{28},' in preparation.

[12] S. Tayama, Polyhedra and Graphics, Graduation Thesis, Uninversity of Electro-Com. (1993).

[13] T. Uchitomi, Quasi-Polyhedra and computer algebra, Graduation Thesis, Uninversity of Electro-Com. (1993).

[14] M. J. Wenninger, '*Polyhedron Models*,' Cambridge University Press (1971).

[15] V. A. Zalgaller, 'Convex polyhedra with regular faces,' Consultants Bureaux, 1969.

Identification of Unknown Coefficient in a Nonlinear Diffusion Equation

Abdullah Shidfar Hossein Azary

Department of Mathematics

Iran University of Science and Technology

Narmak, Tehran–16, Iran

Abstract

The determination of an unknown diffusion coefficient in a nonlinear diffusion equation from overspecified data measured at the boundary will be considered. We also design a new method for identifying mesh points and seek a member of admissible coefficients which minimize a given error functional.

Key words: Diffusion coefficient, Overspecified data, Inverse problem, Error functional

1 Introduction

For $T > 0$ let $Q_T = \{(x,t) \mid 0 < x < 1, \quad 0 < t < T\}$ and seek both the functions $u(x,t)$ and $a(u)$ from the initial boundary value problem

$$\begin{aligned}
\partial_t u(x,t) &= \partial_x(a(u(x,t))\partial_x u(x,t)), \; 0 < x < 1, \; 0 < t < T, && (1)\\
u(x,0) &= 0, \qquad 0 < x < 1, && (2)\\
-a(u(0,t))\partial_x u(0,t) &= g(t), \qquad 0 < t < T, && (3)\\
\partial_x u(1,t) &= 0, \qquad 0 < t < T, && (4)\\
u(0,t) &= f(t), \qquad 0 < t < T. && (5)
\end{aligned}$$

Such problems arise in a variety of physical situations. For example, we might think of this problm as the problem of determining the diffusivity of an unknown porous medium [1]. In this context, we consider a horizontally placed cylinder of unit length which is filled with a homogeneous porous material. Parallel to the cylinder axis there is a liquid flow which partially saturates the porous medium.

Equation (1) will now be obtained by combining Darcy's law and continuity equation. Equations (2) and (4) represent the initial moisture content and moisture content at $x = 1$, respectively. The flux of moisture and moisture content at $x = 0$ are identified by equations (3) and (5), respectively.

If $a(u)$ is given, then the problem (1) would constitute a well–posed problem for the function $u(x,t)$. For an unknown function $a(u)$, we must therefore provide additional information namely (5) to provide a unique solution $(u, a(u))$

to the inverse problem (1)–(5).

Nonlinear inverse problems including equation (1) have been previously treated by many authors who considered certain special cases of this type of problem (cf. [2]–[7]). In [3], for example, Cannon and DuChateau defined an auxiliary inverse problem and sought a class of admissible coefficients $a(u)$ which minimize an error functional. They have shown the existence of a solution to their auxiliary problem in a specified admissible class of functions.

In this article, under certain conditions on $g(t)$ and $f(t)$, we shall identify both the moisture content and diffusivity coefficient at any time by using the overspecified condition (5) and initial–boundary conditions (2)–(4).

Suppose $a(s)$ defined for all $s > 0$ satisfies $0 < \alpha_0 \leq a(s) \leq \alpha_1$ for any given constants α_0, α_1. Now using the following transformation

$$T_a(s) = \int_0^s a(\tau)d\tau, \qquad s \geq 0,$$

which was used by Cannon [3], the probem (1)–(5) reduces to one with the unknwon coefficient not in divergence form. Note that $T_a'(s) = a(s) \geq \alpha_0 > 0$, so that $T(s)$ is invertible. Now, for any solution $u(x,t)$ of (1)–(5), if we define

$$v(x,t) = T_a(u(x,t)) = \int_0^{u(x,t)} a(s)ds \tag{6}$$

then $v(x,t)$ satisfies

$$\partial_t v(x,t) = A(v)\partial_{xx}v(x,t) \qquad 0 < x < 1, \quad 0 < t < T, \tag{7}$$

$$v(x,0) = 0, \qquad 0 < x < 1, \tag{8}$$

$$-\partial_x v(0,t) = g(t), \qquad 0 < t < T, \tag{9}$$

$$\partial_x v(1,t) = 0, \qquad 0 < t < T, \tag{10}$$

where

$$A(v) = a(T_a^{-1}(v))$$

and (5) becomes

$$v(0,t) = T_a(f(t)) = \int_0^{f(t)} a(s)ds = F(t), \qquad 0 < t < T. \tag{11}$$

The plan of this paper is as follows. In section 2, we consider some numerical procedure. In the final section we introduce some new numerical methods and their advantages with respect to other methods.

2 Numerical procedures

In this section, we consider some methods introduced by Cannon [3] and DuChateau [4]. For this purpose, let us suppose A to be a class of admissible coefficients defined as follows

$$A = \{a(u) = (u + \alpha_1)^{\alpha_2} \mid 1 \le \alpha_1 \le 2, \ \ 1.1 \le \alpha_2 \le 1.5\}. \tag{12}$$

For any $a(s)$ given by (12), we have

$$T_a(x) = \int_0^x (s + \alpha_1)^{\alpha_2} ds = \frac{(x + \alpha_1)^{\alpha_2+1} - \alpha_1^{\alpha_2+1}}{\alpha_2 + 1},$$

and

$$A(v) = a(T_a^{-1}(v)) = \left[(\alpha_2 + 1)v + \alpha_1^{\alpha_2+1}\right]^{\alpha_2/(\alpha_2+1)}. \tag{13}$$

Then for

$$g(t) = 2t, \qquad\qquad 0 < t < T, \tag{14}$$

let $v(x, t; \alpha_1, \alpha_2)$ denote the solution of (7)–(10) for

$$0.1 \leq \alpha_1 \leq 10 \quad \text{and} \quad 0 \leq \alpha_2 \leq 10. \tag{15}$$

The numerical experiments were conducted in the following fashion. A pair of values (α_1^*, α_2^*) satisfying (15) were arbitarily selected. In particular, we choose

$$\alpha_1^* = 1.1 \quad \text{and} \quad \alpha_2^* = 0.80. \tag{16}$$

Then for $g(t)$ given by (14), we solve (7)–(10) numerically to obtain $v(x, t; \alpha_1^*, \alpha_2^*)$. We then compute

$$F(t) = v(0, t; \alpha_1^*, \alpha_2^*), \qquad 0 \leq t \leq T,$$

for $T = 15$secs.

Now for $0 \leq T_1 < T_2 \leq T$. let

$$J(\alpha_1, \alpha_2) = \int_{T_1} |v(0, t; \alpha_1, \alpha_2) - F(t)|^2 dt.$$

We used a nonlinear optimization routine to minimize J over the compact set of values (α_1, α_2) indicated in (16). The optimizer used was the program ZXMIN of the IMSL package, and the partial differential equation was solved by a straightforward finite difference approach. Table 1 shows a selection of starting guesses, and the number of guesses required to converge back to $(\alpha_1^*, \alpha_2^*) = (1.1, 0.80)$.

Table 1:

Starting guess	T_1	Number of guesses
$(1.0, 0.7)$	10	31
$(1.0, 0.7)$	1	24
$(1.5, 1.0)$	10	60
$(1.5, 1.0)$	1	46
$(0.8, 0.6)$	10	46
$(1.5, 1.2)$	10	failed to converge

Of course, the results shown in this table are highly dependent on both the optimization routine and the pde solver one uses.

As a second numerical method for obtaining (α_1, α_2) we use the intersecting graph technique such that the results are displayed in Figures 1, 2, 3 and 4 (details in [3,4]).

3 Some numerical methods

In this section we are going to demonstrate numerically, some of numerical results for diffusion coefficient in the direct diffusion problem (7)–(11). For this purpose, let us define diffusion coefficient $A(v)$ in (7)–(11) as follows

$$A(v) = c_0 + c_1 v + c_2 v^2 + \cdots + c_n v^n \quad n \in \mathbf{N}$$

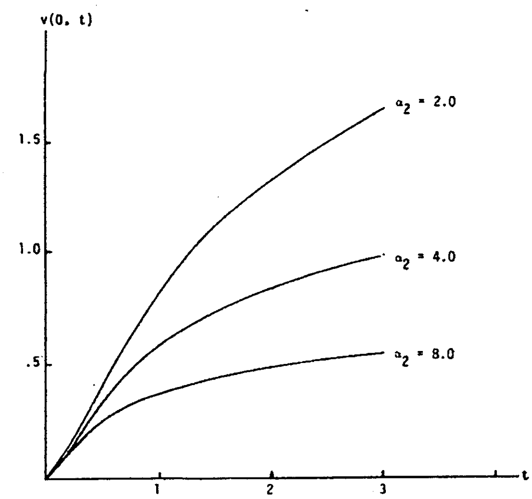


Figure 1: $v(0,t) = t$ for $\alpha_1 = 1$, with $\alpha_2 = 2,\ 4,\ 8$.

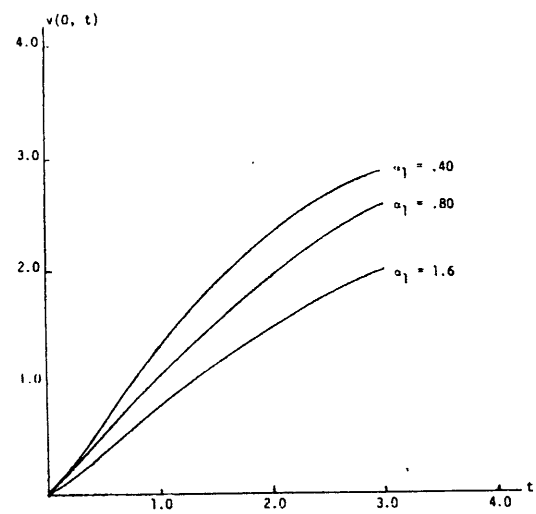


Figure 2: $v(0,t)$ $vs.$ t for $\alpha_2 = 1.1$, with $\alpha_1 = 0.4,\ 0.8,\ 1.6$.

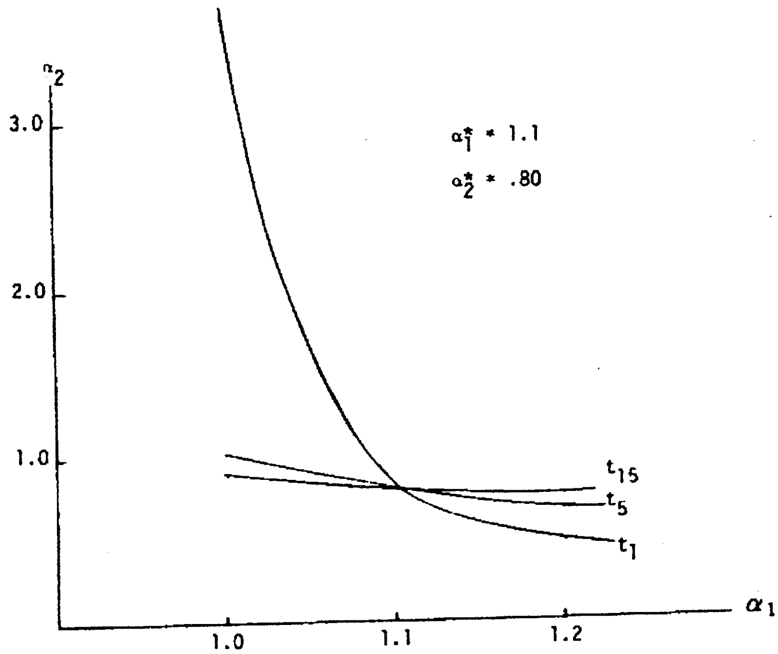


Figure 3: $\alpha_2 = h(\alpha_1; t_i)$ for t_1, t_5 O t_{15}.

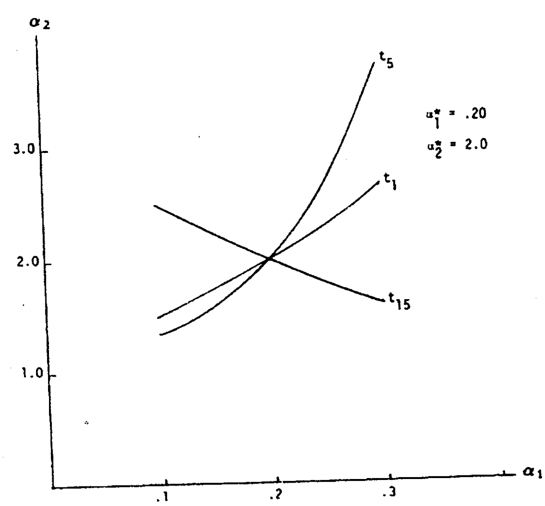


Figure 4: $\alpha_2 = h(\alpha_1; t_i)$ for t_1, t_5 O t_{15}.

where $c_0, c_1, \cdots, c_n$ are constants which remain to be determined in the following steps

1. In the first step, we select $A(v) = c_0$ and determine c_0 such that the following functional becomes minimum.

$$J(c_0) = \int_T |v(0,t : c_0) - F(t)|^2 dt,$$

2. In the second step, we choose $A(v) = c_0 + c_1 v$, where c_0 is given by first step and c_1 minimize the following functional

$$J(c_0, c_1) = \int_T |v(0,t; c_0, c_1) - F(t)|^2 dt,$$

3. In this step, we select $A(v) = c_0 + c_1 v + c_2 v^2$, where c_0 and c_1 are given in the previous steps and we choose c_2 such that $A(v) = c_0 + c_1 v + c_2 v^2$ minimizes the following functional

$$J(c_0, c_1, c_2) = \int_T |v(0,t; c_0, c_1, c_2) - F(t)|^2 dt.$$

This method may be carried on upto when the minimum of the above functional becomes greater than the previous one.

Now, let us consider the inverse problem (1)–(5). The difference equation for (1) may be written in the form

$$\frac{1}{k}\Delta_t u_{i,j} = \frac{1}{2}\left\{\frac{\delta_x^2(a(u_{i,j})u_{i,j}) + \delta_x^2(a(u_{i,j+1})u_{i,j+1})}{h^2}\right\}.$$

For $\delta x = 0.1$ and $\delta t = 0.01$, we find

$$-2a(u_1)u_1 + (2 + 2a(u_0))u_0 = 0, \qquad i = 0, \;\; j = 0, \tag{17}$$

$$-a(u_{i-1})u_{i-1}+(2+2a(u_i))u_i-a(u_{i+1})u_{i+1}=0 \quad i=1,\cdots,9, \;\; j=0, \quad (18)$$

$$-2a(u_9)u_9+(2+2a(u_{10}))u_{10}=0 \qquad i=10, \;\; j=0. \qquad (19)$$

For any given $a(u)$ this non–linear system may be solved by the method given in [ ] or by using the MAPLE V software.

Now, if $a(u)$ is unknown, then the diffusion coefficient $a(u)$, may be considered as a polynomial of u of degree n with the unknown coefficients $a_0, a_1, \cdots, a_n$, to determine these coefficients, we may procedure as follows

1. In this step, we choose $a(u) = a_0$, then by overspecified condition (5) u_0 is given. Now, $a_0, u_1, u_2, \cdots, u_{10}$ can be computed from (17) –(19).

2. In the second step, we select $a(u) = a_0 + a_1 u$, where a_0 is given in the previous step and by the system (17)–(19) we identify a_1.

In any step, we calculate the functional

$$\int_T |u(0,t) - f(t)|^2 dt, \qquad 0 \le t \le T.$$

The method will be stoped, when in one step, this functional becomes greater that the previous one.

References

[1] J. Bear, "*Dynamics of fluids in porous media*", Elsevier, New York, N.Y., 2nd ed., 1975.

[2] J.R. Cannon, "*The One-Dimensional Heat Equation*", Addison-Wesley Co. Menlo park, California, 1984.

[3] J.R. Cannon and P. DuChateau, "*An inverse problem for a nonlinear diffusion equation*", SIAM J. Appl. Math., 39(2), (1980), 272-289.

[4] P. DuChateau, "*Monotonicity and uniqueness results in identifying an unknown coefficient in a nonlinear diffusion equation* ", SIAM J. Appl. Math. 41(2), (1981), 310-323.

[5] P. DuChateau and D. Zachmann , "*Partial Differential Equations*", Schaum's Outlines 1986.

[6] A. Friedman, "*Partial Differential Equations of Parabolic Type,*", Prentice-Hall, Englewood Cliffs. NJ, 1964.

[7] O.A. Ladyzhenskaya, V.A. Sollonikov and N.N. Uralceva,"*Linear and Quasilinear equations of parabolic type*", AMS, Providence RI, 1968.

COMPUTATION OF ESSENTIALLY DIFFERENT PUISEUX EXPANSIONS VIA EXTENDED HENSEL CONSTRUCTION

Kousuke Shiihara

Institute of Mathematics, University of Tsukuba
Tsukuba-shi 305, Japan
kousuke@math.tsukuba.ac.jp

Abstract

This paper describes an efficient method for computing essentially different Puiseux expansions, $y = f_i(x)$ $(1 \le i \le d = \deg_y(F))$, which are roots of bivariate polynomial $F(x, y) \in \mathbf{C}[x, y]$. Essentially different Puiseux expansions play an important role in algebraic geometry, e.g., for calculating the genus of an algebraic plane curve and for computing the ramification index. Our method is based on extended Hensel construction proposed by Sasaki & Kako recently [SK]. In this paper, first, we give a close relation between Sasaki-Kako's extended Hensel construction and essentially different Puiseux expansions. Next, we explain how to compute essentially different Puiseux expansions efficiently, using extended Hensel construction.

1 Introduction

For a long time, Newton polygon method (see [Ki], [Ta] for details) has been a standard method to perform Puiseux expansions. However, the method itself is very inefficient and conventionally has been performed by introducing algebraic numbers successively, which makes the method even more inefficient. Although the method proposed by D.Duval keeps away from introducing algebraic numbers if possible, its computation is also quite heavy due to exact computation.

In many cases, we need only the properties of Puiseux expansion rather than the expansion itself. For example, for calculating the genus of an algebraic plane curve, only branched index is necessary. For this purpose, the computation can be performed with floating-point arithmetic and [SK] shows that Sasaki-Kako's extended Hensel construction method is reliable enough to guarantee correct answer. The method based on floating-point arithmetic is very efficient compared with classical Newton polygon method or the method proposed by D.Duval, which widen the applicability of the method. In this paper, we improve the method of extended Hensel construction. We prove that if two initial factors of extended Hensel construction are also conjugate to each other then the corresponding factors lifted by the construction are also conjugate to each other. This fact leads directly to a revision of extended Hensel construction. This improvement makes computation of essentially different Puiseux expansions easy.

We assume that the given polynomial $F(x, y)$ is irreducible over **C**, monic w.r.t. variable y and $F(0, y) = y^d$. The generality of this assumption is guaranteed by the following reasons.

- If $F(x, y)$ is not irreducible, we factorize $F(x, y)$ over **C** and treat each factor separately; note that we have an approximate factorization algorithm which is applicable for multivariate polynomials with floating-point number coefficients.

- If $F(x, y) = c_d(x)y^d + \cdots + c_0(x)$ is not monic, we define

$$\tilde{F}(x, y) = c_d^{d-1} \cdot F(x, y/c_d).$$

  Then, $\tilde{F}(x, y)$ is monic w.r.t. variable y, and the root $f(x)$ of $F(x, y)$ and the root $\tilde{f}(x)$ of $\tilde{F}(x, y)$ are related with each other as $f(x) = \tilde{f}(x)/c_d(x)$.

- If $F(0, y) \neq y^d$, we factorize it over **C** as follows.

$$F(0, y) = F_1 \cdots F_r, \quad F_i = (y - \alpha_i)^{e_i}, \quad (i = 1, \ldots, r).$$

  Then, we can apply Hensel construction to F_i and F_j, $(i \neq j)$.

This paper is organized as follows. In **2**, we give some notations. In **3**, we describe Sasaki-Kako's extended Hensel construction method briefly. In **4**, we give main result of this paper. In **5**, we give an illustrative example. Finally in **6**, conclude.

2 Notations

$\mathbf{C}\langle\langle\cdot\rangle\rangle$, $\mathbf{C}[\cdot]$ and $\mathbf{C}\{\cdot\}$ denote a field of Puiseux series over **C**, a ring of polynomials over **C** and a ring of formal power series over **C**, respectively.

3 Extended Hensel construction

In this section, we describe extended Hensel construction proposed by T.Sasaki and F.Kako, briefly.

Definition 3.1 *For each non-zero term $c \cdot y^i x^j$ of $F(x,y)$, we plot a dot at the point (i,j) in the two-dimensional Cartesian coordinate system. Let L be a straight line such that it passes the point $(d,0)$ as well as another dot plotted and that no dot is plotted below L. The line L is called Newton's line for F. The sum of all the terms of $F(x,y)$, which are plotted on Newton's line is called Newton's polynomial for F. We denote Newton's polynomial by $F^{(0)}(x,y)$.* ■

Remark 3.2 *Newton's line is uniquely determined by $F(x,y)$. Let δ be the $(\deg_y)$-coordinate of intersection of L and $(\deg_y)$-axis, hence Newton's line is $(\deg_x)/d + (\deg_y)/\delta = 1$. Therefore Newton's polynomial consists of some of the terms*

$$y^d,\ y^{d-1}x^{\delta/d},\ y^{d-2}x^{2\delta/d}, \ldots,\ x^{d\delta/d}.$$

Let $\hat{\delta}$ and $\hat{d}$ be positive integers such that

$$\hat{\delta}/\hat{d} = \delta/d, \quad \gcd(\hat{\delta}, \hat{d}) = 1.$$

■

Suppose that $F^{(0)}(1,y)$ is factorized over $\mathbf{C}$ as

$$F^{(0)}(1,y) = (y-\zeta_1)^{m_1}\cdots(y-\zeta_r)^{m_r}, \quad \zeta_1,\ldots,\zeta_r \in \mathbf{C}, \quad \zeta_i \neq \zeta_j \text{ for any } i \neq j.$$

Since $F^{(0)}(x,y)$ is homogeneous polynomial in y and $x^{\delta/d}$, we can write $F^{(0)}(x,y)$ in the following form.

$$F^{(0)}(x,y) = G_1^{(0)}\cdots G_r^{(0)}, \quad G_i^{(0)} = (y - \zeta_i x^{\hat{\delta}/\hat{d}})^{m_i}. \tag{3.1}$$

We define an ideal I_k as

$$I_k = (y^d x^{(k+0)/\hat{d}},\ y^{d-1}x^{(k+\hat{\delta})/\hat{d}},\ y^{d-2}x^{(k+2\hat{\delta})/\hat{d}}, \ldots,\ y^0 x^{(k+d\hat{\delta})/\hat{d}}).$$

Then, we have following lemma.

Lemma 3.3 (Lagrange's interpolation polynomials) *For each value of $l = 0,\ldots,d-1$, there exists only one set of polynomials $\left\{ W_i^{(l)}(x,y) \mid i = 1,\ldots,r \right\}$ satisfying*

$$W_1^{(l)}[G_1^{(0)}\cdots G_r^{(0)}/G_1^{(0)}] + \cdots + W_r^{(l)}[G_1^{(0)}\cdots G_r^{(0)}/G_r^{(0)}] = y^l x^{(\hat{\delta}/\hat{d})(d-l)}$$
$$\deg_y(W_i^{(l)}(x,y)) < \deg_y(G_i^{(0)}(x,y)), \quad (i = 1,\ldots,r).$$

■

The next theorem follows from the above lemma.

Theorem 3.4 *For all $k \in \mathbf{Z}$, we can construct $G_i^{(k)}(x,y)$, $(i=1,\dots,r)$, satisfying*

$$\begin{aligned} F(x,y) &\equiv G_1^{(k)}(x,y)\cdots G_r^{(k)}(x,y) \pmod{I_{k+1}} \\ G_i^{(k)}(x,y) &\equiv G_i^{(0)}(x,y) \pmod{I_1}, \quad (i=1,\dots,r). \end{aligned}$$

■

The above construction is performed by the following procedure. Suppose that $G_i^{(k-1)}(x,y)$, $(i=1,\dots,r)$ are calculated, we express $F(x,y)-G_1^{(k-1)}\cdots G_r^{(k-1)}$ as

$$F(x,y)-G_1^{(k-1)}\cdots G_r^{(k-1)} \equiv f_{d-1}^{(k)}\cdot y^{d-1}x^{\hat{\delta}/\hat{d}}+\cdots+f_0^{(k)}\cdot y^0 x^{d\hat{\delta}/\hat{d}} \pmod{I_{k+1}}. \quad (3.2)$$

Then, we construct $G_i^{(k)}(x,y)$ by putting

$$G_i^{(k)}(x,y)=G_i^{(k-1)}(x,y)+\sum_{l=0}^{d-1} W_i^{(l)}(x,y) f_l^{(k)}(x), \quad (i=1,\dots,r). \quad (3.3)$$

4 Essentially different Puiseux expansions

In this section, we describe an improvement on extended Hensel construction, and show how to compute essentially different Puiseux expansions.

Definition 4.1 *Let G_1, $G_2 \in \mathbf{C}[x^{1/\hat{x}}, y]$. G_1 and G_2 are said to be conjugate to each other if*

$$\exists \theta \in \mathbf{Z} \quad s.t. \quad G_1(x^{1/\hat{x}},y)=G_2(e^{2\pi i\theta/\hat{d}}x^{1/\hat{d}},y).$$

■

Definition 4.2 *For all θ, $\hat{d} \in \mathbf{Z}$, we define a mapping $\pi_{\theta,\hat{d}}$*

$$\begin{aligned} \pi_{\theta,\hat{d}} \ : \mathbf{C}[x^{1/\hat{d}},y] &\longrightarrow \mathbf{C}[x^{1/\hat{d}},y] \\ G(x^{1/\hat{d}},\ y) &\mapsto G(e^{2\pi i\theta/\hat{d}}x^{1/\hat{d}},\ y), \end{aligned}$$

$$\begin{aligned} \pi_{\theta,\hat{d}} \ : \mathbf{C}\langle\langle x\rangle\rangle &\longrightarrow \mathbf{C}\langle\langle x\rangle\rangle \\ g(x^{1/\hat{d}}) &\mapsto g(e^{2\pi i\theta/\hat{d}}x^{1/\hat{d}}). \end{aligned}$$

■

Definition 4.3 *We say that two series $f,\ g \in \mathbf{C}\langle\langle x\rangle\rangle$ are essentially different if there are no integers $\alpha,\ \beta$ such that $\pi_{\alpha,\beta}(f) = g$.* ■

Lemma 4.4 *Let $G_j^{(0)},\ G_k^{(0)} \in \mathbf{C}[x^{1/\hat{d}}, y]$, $(j \neq k)$. If there exists $\theta \in \mathbf{Z}$ such that $\pi_{\theta,\hat{d}}\left(G_j^{(0)}\right) = G_k^{(0)}$, then*

$$\pi_{\theta,\hat{d}}\left(W_j^{(l)}\right) = e^{-2\pi i\theta s/\hat{d}} \cdot W_k^{(l)}$$

for all $s \in \mathbf{Z}^+$ satisfying $\frac{s}{\hat{d}} + \left(\frac{\hat{\delta}}{\hat{d}}\right)(d-l) \in \mathbf{Z}$.

Proof. From the proof of Lagrange's interpolation polynomials, we have $\tilde{W}_j^{(l)}$, $\tilde{W}_k^{(l)}$, $(l = 0, \ldots, d-1)$ which satisfying

$$\tilde{W}_j^{(l)} G_j^{(0)} + W_j^{(l)}[G_1^{(0)} \cdots G_r^{(0)}/G_j^{(0)}] = y^l x^{(\hat{\delta}/\hat{d})(d-l)}, \tag{4.1}$$
$$\tilde{W}_k^{(l)} G_k^{(0)} + W_k^{(l)}[G_1^{(0)} \cdots G_r^{(0)}/G_k^{(0)}] = y^l x^{(\hat{\delta}/\hat{d})(d-l)}. \tag{4.2}$$

Multiplying the above equations by $x^{s/\hat{d}}$, we obtain

$$x^{s/\hat{d}} \cdot \tilde{W}_j^{(l)} G_j^{(0)} + x^{s/\hat{d}} \cdot W_j^{(l)}[G_1^{(0)} \cdots G_r^{(0)}/G_j^{(0)}] = y^l x^{s/\hat{d}+(\hat{\delta}/\hat{d})(d-l)}. \tag{4.3}$$

Then, applying $\pi := \pi_{\theta,\hat{d}}$ to (4.3), we have

$$e^{2\pi i\theta s/\hat{d}} \cdot x^{s/\hat{d}} \cdot \pi\left(\tilde{W}_j^{(l)}\right) \cdot G_k^{(0)} + e^{2\pi i\theta s/\hat{d}} \cdot x^{s/\hat{d}} \cdot \pi\left(W_j^{(l)}\right) \cdot [G_1^{(0)} \cdots G_r^{(0)}/G_k^{(0)}] = y^l x^{s/\hat{d}+(\hat{\delta}/\hat{d})(d-l)}.$$

We divide the above equation by $x^{s/\hat{d}}$ and obtain

$$e^{2\pi i\theta s/\hat{d}} \cdot \pi\left(\tilde{W}_j^{(l)}\right) \cdot G_k^{(0)} + e^{2\pi i\theta s/\hat{d}} \cdot \pi\left(W_j^{(l)}\right) \cdot [G_1^{(0)} \cdots G_r^{(0)}/G_k^{(0)}] = y^l x^{(\hat{\delta}/\hat{d})(d-l)} \tag{4.4}$$

Comparing the second term of the left-hand side of (4.2) with (4.4), we get the following equation by the uniqueness of the interpolation polynomials.

$$e^{2\pi i\theta s/\hat{d}} \cdot \pi\left(W_j^{(l)}\right) = W_k^{(l)}.$$

This completes the proof. ■

Lemma 4.5 $G_1^{(k)} \cdots G_r^{(k)} \in \mathbf{C}[x, y]$, *for all $k \in \mathbf{Z}^+$.* ■

Theorem 4.6 *Let $G_j^{(0)}, G_k^{(0)} \in \mathbf{C}[x^{1/\hat{d}}, y]$, $(j \neq k)$. If there exists $\theta \in \mathbf{Z}$ such that $\pi_{\theta,\hat{d}}\left(G_j^{(0)}\right) = G_k^{(0)}$, then*

$$\pi_{\theta,\hat{d}}\left(G_j^{(s)}\right) = G_k^{(s)}$$

for all $s \in \mathbf{Z}^+$ satisfying $\pi_{\theta,\hat{d}}\left(G_j^{(0)}\right) = G_k^{(0)}$. That is, the following diagram commutes.

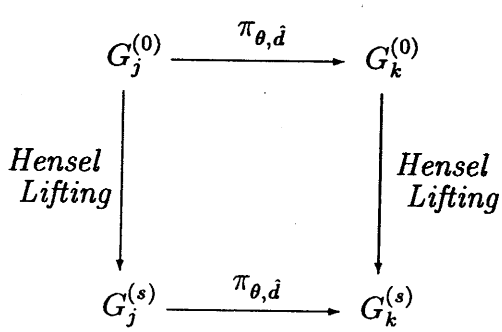


Figure 1:

Proof. Suppose that the theorem is valid up to the $(s-1)$-st step. Furthermore, as the induction assumption, we assume that

$$G_j^{(s)} = G_j^{(s-1)} + \sum_{l=0}^{d-1} W_j^{(l)} c_l^{(s)} x^{s/\hat{d}}, \tag{4.5}$$

where

$$f_{d-1}^{(s)} y^{d-1} x^{\hat{\delta}/\hat{d}} + \cdots + f_0^{(s)} y^0 x^{d\hat{\delta}/\hat{d}} \equiv F(x,y) - G_1^{(s-1)} \cdots G_r^{(s-1)} \pmod{I_{s+1}}$$

$$f_l^{(s)} = c_l^{(s)} x^{s/\hat{d}}, \quad c_l^{(s)} \in \mathbf{C} \quad (l = 0, \ldots, d-1). \tag{4.6}$$

$c_l^{(s)} = 0$ $(l = 0, \ldots, d-1)$ satisfying $\frac{s}{\hat{d}} + \left(\frac{\hat{\delta}}{\hat{d}}\right)(d-l) \notin \mathbf{Z}$ follows from (4.6) and lemma 4.5. Hence, equation (4.5) can be written as

$$G_j^{(s)} = G_j^{(s-1)} + \sum_{\substack{l=0,\ldots,d-1 \\ s/\hat{d}+(\hat{\delta}/\hat{d})(d-l)\in\mathbf{Z}}} W_j^{(l)} c_l^{(s)} x^{s/\hat{d}} \tag{4.7}$$

Applying $\pi := \pi_{\theta,\hat{d}}$ to the above equation and using the assumption and lemma 4.4, we have

$$\pi\left(G_k^{(s)}\right) = \pi\left(G_k^{(s-1)}\right) + \sum \pi\left(W_k^{(l)}\right) \cdot c_l^{(s)} \cdot \pi\left(x^{s/\hat{d}}\right)$$

$$
\begin{aligned}
&= G_j^{(s-1)} + \sum \pi\left(W_k^{(l)}\right) \cdot c_l^{(s)} \cdot e^{2\pi i \theta s/\hat{d}} \cdot x^{s/\hat{d}} \\
&= G_j^{(s-1)} + \sum e^{2\pi i \theta s/\hat{d}} \cdot \pi\left(W_k^{(l)}\right) \cdot c_l^{(s)} \cdot x^{s/\hat{d}} \\
&= G_j^{(s-1)} + \sum W_j^{(l)} \cdot c_l^{(s)} \cdot x^{s/\hat{d}} \\
&= G_j^{(s)}.
\end{aligned}
$$

This completes the proof. ■

Corollary 4.7 *Let $G_j^{(0)}, G_k^{(0)} \in \mathbf{C}[x^{1/\hat{d}}, y]$, $(j \neq k)$. Suppose that there exists $\theta \in \mathbf{Z}$ such that $\pi_{\theta,\hat{d}}\left(G_j^{(0)}\right) = G_k^{(0)}$. If $y = \chi(x)$ is one of the roots of $G_j^{(\infty)}$, then $y = \pi_{\theta,\hat{d}}\big(\chi(x)\big)$ is also one of the roots of $G_k^{(\infty)}$.* ■

Corollary 4.8 *$G_1^{(0)}, G_2^{(0)}, \ldots, G_t^{(0)}$ satisfies*

$$
\begin{cases}
G_1^{(0)} G_2^{(0)} \cdots G_t^{(0)} \in \mathbf{C}[x, y], \\
\text{Square-free factor of } G_1^{(0)} G_2^{(0)} \cdots G_t^{(0)} \text{ is irreducible.}
\end{cases}
$$

Then, the following equation is satisfied.

$$
\begin{aligned}
&\#\left\{ \text{Essentially different Puiseux expansions generated by } G_i^{(\infty)} \right\} \\
&\qquad = \#\left\{ \begin{array}{l} \text{Essentially different Puiseux expansions} \\ \text{generated by } G_1^{(\infty)} G_2^{(\infty)} \cdots G_t^{(\infty)} \end{array} \right\}.
\end{aligned}
$$

■

5 Example

Let $F(x, y)$ be

$$F(x, y) = y^5 + xy^4 - 2xy^3 - 2x^2y^2 + (x^2 - x^3)y + x^3 + x^{10}.$$

Initial factors of extended Hensel construction become as follows.

$$G_1^{(0)} = y, \quad G_2^{(0)} = (y + x^{1/2})^2, \quad G_3^{(0)} = (y - x^{1/2})^2.$$

Hence, we can compute $G_1^{(\infty)}$, $G_2^{(\infty)}$ and $G_3^{(\infty)}$, using extended Hensel construction as follows.

$$
\begin{aligned}
G_1^{(\infty)} &= y + x + \cdots \\
G_2^{(\infty)} &= (y + x^{1/2})^2 - (\frac{x^{3/2}y}{4} + \frac{x^2}{2}) + \cdots \\
G_3^{(\infty)} &= (y - x^{1/2})^2 + (\frac{x^{3/2}y}{4} - \frac{x^2}{2}) + \cdots
\end{aligned}
$$

Then, we find that $G_2^{(\infty)}$ and $G_3^{(\infty)}$ are conjugate to each other. Therefore, it is sufficient to check only Puiseux expansions generated by $G_1^{(\infty)}$ and $G_2^{(\infty)}$ for calculating essentially different Puiseux expansions. We apply extended Hensel construction to $G_2^{(\infty)}$ and $G_3^{(\infty)}$, and obtain the following essentially different Puiseux expansions of F.

$$
\begin{aligned}
y_1 &= x + x^2 + \cdots \\
y_2 &= x^{1/2} - \frac{x}{2} - \frac{x^{3/2}}{8} + \cdots \\
y_3 &= x^{1/2} + \frac{x}{2} - \frac{x^{3/2}}{8} + \cdots
\end{aligned}
$$

6 Conclusion

We prove that if two initial factors of extended Hensel construction are conjugate to each other then the corresponding factors lifted by the construction are also conjugate to each other. This fact leads directly to a revision of extended Hensel construction. By this, we can find essentially different Puiseux expansions easily.

Acknowledgments

The author is greatly indebted to Professor Tateaki Sasaki, Dr. Takuya Kitamoto and my colleagues for helpful discussion.

References

[BrKn] E.Brieskorn and H.Knörrer, *Plane Algebraic Curves.* Birkhäuser Verlag, Basel-Boston-Stuttgart, 1986.

[Du] D.Duval, Rational Puiseux expansions. Compositio Mathematica, 70, pp. 119-154, 1989.

[Hi] D.L.Hilliker, An algorithm for computing the values of the ramification index in the puiseux series expansions of an algebraic function. Pacific Journal of Mathematics, vol. 118, No. 2, pp. 427-435, 1985.

[KFS] T.Kitamoto, F.Kako and T.Sasaki, On numerical power series roots of a multivariate polynomial (in Japanese). Presented at RIMS Symposium, Kyoto University, November 1993. 7 pages.

[Ki] F.Kirwan, *Complex Algebraic Curves.* London Mathematical Society, Student Texts 23, 1992.

[KT] H.T.Kung and J.F.Traub, All algebraic functions can be computed fast. J. ACM 25, pp. 245-260, 1978.

[Ku] T.C. Kuo, Generalized Newton-Puiseux theory and Hensel's lemma in $\mathbf{C}[\![x,y]\!]$. Can. J. Math., Vol. XLI, No. 6, pp. 1101-1116, 1989.

[SIAS] T.Sasaki, K.Imai, T.Asano and K.Sugihara, *Algebraic computation and Computational Geometry* (In Japanese), Iwanami Shoten 1993.

[SK] T.Sasaki and F.Kako, Solving multivariate algebraic equation by Hensel construction. Preprint of Univ. Tsukuba, January 1993. 22 pages.

[SKK] T.Sasaki, T.Kitamoto and F.Kako, Error analysis of power series roots of multivariate algebraic equation, Preprint of Univ. Tsukuba, April 1994, 31 pages.

[Ta] J.Tamura, *Analytic Functions* (in Japanese). Selected Books on Mathematics Vol.3. Shoukabou, Tokyo, 1983.

[Wa] B.Wall, Puiseux expansion : an annotated bibliography. Technical Report, RISC-Linz Series no. 91-46.0, October 23, 1991

ON A CLASS OF TWO-DIMENSIONAL ALGEBRAS

by

Sebastian Vattamattam and R. Sivaramakrishnan

Department of Mathematics, University of Calicut,
India 673 635

ABSTRACT

Given a,b elements of a field of characteristic $\neq 2$, we construct a two-dimensional algebra $A_F(a,b)$ (over F) with unity. $A_F(a,b)$ is shown to be isomorphic to $A_F(d,0)$ where $d = b^2+4a$. In the special case when F = Q, the field of rational numbers, the structure of A_F (d,0) throws light on the splitting field of an irreducible quadratic polynomial over Q.

ON A CLASS OF TWO-DIMENSIONAL ALGEBRAS

by

Sebastian Vattamattom and R. Sivaramakrishnan
Department of Mathematics, UNIVERSITY OF CALICUT, INDIA 673 635

1. INTRODUCTION:

$\mathbb{R}$ denotes the field of real numbers. Under complex multiplication $\mathbb{R}^2$ is a division algebra. That is to say:

In addition to the vector-space structure of $\mathbb{R}^2$ over $\mathbb{R}$, multiplication (.) is such that for $u, v, w \in \mathbb{R}^2$

i) $u.(v+w) = u.v + u.w$ and $(u+v).w = u.w + v.w$

ii) $a(u.v) = (au).v = n.(av)$ for all $a \in \mathbb{R}$

iii) for every v and $u = 0$ in R^2, the equation

$$u.x = v \quad \text{and} \quad y.u = v$$

have unique solutions in $\mathbb{R}^2$.

In [1], Stephen C. Althoen and Lawrence D. Kugler have given an interesting account of how and when $\mathbb{R}^2$ could be made a division algebra by defining multiplication suitably. It is known [1] that every two-dimensional real division algebra has at least one idempotent and so contains a subalgebra isomorphic to $\mathbb{R}$. .

Let $\mathbb{R}^2$ be an algebra with unity e_1.

If $\{e_1, e_2\}$ is a basis for R^2, its multiplication table is given by

(1.1)

	e_1	e_2
e_1	e_1	$a_{12}e_1 + b_{12}e_2$
e_2	$a_{21}e_1+b_{21}e_2$	$a_{22}e_1 + b_{22}e_2$

where a_{ij} (i, j = 1,2) are real numbers.

The algebra described by (1.1) is a division algebra if and only if

(1.2) $$(b_{22} - A_1)^2 < 4\, b_{12}A_2$$

where

$$A_1 = \det \begin{bmatrix} a_{12} & b_{12} \\ a_{21} & b_{21} \end{bmatrix} \quad \text{and} \quad A_2 = \det \begin{bmatrix} a_{21} & b_{21} \\ a_{22} & b_{22} \end{bmatrix}$$

(1.2) is rewritten as

(1.3) $$(b_{22} - a_{12}b_{21} + a_{21}\, b_{12})^2 - 4\, b_{12}\, (a_{21}b_{22} - a_{22}b_{21}) < 0$$

We call the left side of (1.3), the discriminant of the algebra $\mathbb{R}^2$.

In the case of $\mathbb{R}^2$ with the following multiplication table

(1.4)

	e_1	e_2
e_1	e_1	e_2
e_2	e_2	$a\,e_1 + be_2$

where a, b $\in \mathbb{R}$, the discriminant has the form $b^2 + 4a$. So, the

2-dimensional algebra determined by (1.4) is a real division algebra if and only if $b^2 + 4a < 0$. By the proposition following theorem 4 in [1], we note that the condition $b^2+4a < 0$ implies and is implied by the fact that $\mathbb{R}^2$ given by the multiplication table (1.4) has exactly one idempotent. The case $a = -1$, $b = 0$ in (1.4) gives the familiar field of complex numbers.

In a more general setting, let F be an arbitrary but fixed field of characteristic $\neq 2$ and having the unity element 1_F. We write $e_1 = (1_F, 0)$, $e_2 = (0, 1_F)$. e_1 will be the unity element of the proposed algebra obtained from F^2. The multiplication table is given by

(1.5)

	e_1	e_2
e_1	e_1	e_2
e_2	e_2	$ae_1 + be_2$

where $a, b \in F$.

We denote the algebra obtained from the multiplication table (1.5) by $A_F(a,b)$. The structure of $A_F(a,b)$ is known, when $F = \mathbb{R}$. That $A_F(a,b)$ is a commutative algebra is clear from the multiplication table (1.5). The purpose of this note is to study the structure of $A_F(a,b)$ via the ´discriminant´ $b^2 + 4a$ and to point out an application in the special case $F = \mathbb{Q}$, the field of rational numbers. The outcome is the structure of the splitting field of a quadratic polynomial over Q.

2. THE STRUCTURE OF $A_F(a,b)$:

Firstly, we make the role of 'b^2+4a' clear in the following

THEOREM: 1. $A_F(a,b)$ is isomorphic to $A_F(d,0)$ where $d = b^2+4a$.

Proof: Whereas the multiplication table of $A_F(a,b)$ is given by (1.5), the multiplication table of $A_F(d,0)$ is

(2.1)

	e_1	e_2
e_1	e_1	e_2
e_2	e_2	de_1

Let $\phi : A_F(a,b) \longrightarrow A_F(d,0)$ be an algebra homomorphism defined by

$$\phi(e_1) = e_1, \qquad \phi(e_2) : x e_1 + y e_2 \text{ where } x, y \in F$$

Then,

$$\phi(e_2^2) = (\phi(e_2))^2 = (xe_1 + ye_2)^2 = x^2 e_1 + y^2 de_1 + 2xye_2$$

Therefore,

(2.2) $$\phi(e_2^2) = (x^2 + y^2 d) e_1 + 2xye_2$$

As $e_2^2 = ae_1 + be_2$ in $A_F(a,b)$, we also have

(2.3) $$\phi(e_2^2) = \phi(ae_1 + be_2) = (a + bx)e_1 + bye_2$$

From (2.2) and (2.3), we deduce that

(2.4) $$\begin{aligned} x^2 + y^2 d &= a + bx \\ 2xy &= by \end{aligned}$$

If $y = 0$, cancellation of y in $2xy = by$ gives $x = \frac{1}{2}b$ where $\frac{1}{2} = \dfrac{1_F}{1_F + 1_F}$. Then, we have

$$x = \tfrac{1}{2}b, \quad y = \pm\tfrac{1}{2}$$

So, ϕ maps e_2 into $\frac{1}{2}b\,e_1 \pm \frac{1}{2}e_2$.

Fixing the sign, let $\phi(e_2) = \frac{1}{2}be_1 + \frac{1}{2}e_2$.

We observe that $\phi(re_1 + se_2) = \phi(r'e_1 + s'e_2)$, $r, r', s, s' \in F$ implies

$$re_1 + s\left(\tfrac{1}{2}be_1 + \tfrac{1}{2}e_2\right) = r'e_1 + s'\left(\tfrac{1}{2}be_1 + \tfrac{1}{2}e_2\right)$$

or

$$(r + \tfrac{1}{2}sb)\,e_1 + \tfrac{1}{2}se_2 = (r' + \tfrac{1}{2}s'b)\,e_1 + \tfrac{1}{2}s'e_2$$

This gives $s = s'$ and $r = r'$, proving that ϕ is an isomorphism. To see that ϕ is surjective, we observe that when $r'e_1 + s'e_2$ is a vector in $A_F(d,0)$ there exists vector $(r' - bs')e_1 + 2s'e_2$ in $A_F(a,b)$ such that

$$\phi((r' - bs')\,e_1 + 2s'e_2) = r'e_1 + s'e_2 \in A_F(d,0)$$

here 2 stands for $1_F + 1_F$ in F.

This proves that $A_F(a,b)$ and $A_F(d,0)$ are isomorphic when $d = b^2 + 4a$.

From theorem 1, we see that we need to study the structure of $A_F(d,0)$ only instead of $A_F(a,b)$.

We call an element t in $A_F(d,0)$, a non-singular element if it is invertible. They are, of course, the units in the ring structure of $A_F(d,0)$.

THEOREM: 2. Let $t = x_1e_1 + x_2e_2 \in A_F(d,0)$. t is nonsingular if and only if $x_1^2 - d\,x_2^2 \neq 0$.

Proof: If $t = x_1e_1 + x_2e_2$ is nonsingular, there exists $t' = y_1e_1 + y_2e_2$ in $A_F(d,0)$ such that $t.t' = e_1$.

Using the multiplication table (2.1) of $A_F(d,0)$, we see that

$$(x_1e_1 + x_2e_2) \cdot (y_1e_1 + y_2e_2) = e_1$$

implies and is implied by

$$\begin{aligned} x_1y_1 + x_2d\,y_2 &= 1_F \\ x_2y_1 + x_1\,y_2 &= 0 \end{aligned} \tag{2.5}$$

That is,

$$\begin{bmatrix} x_1 & x_2d \\ x_2 & x_1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1_F \\ 0 \end{bmatrix} \tag{2.6}$$

For (2.6) to have a non-trivial solution in y_1 and y_2, a necessary and sufficient condition is that

$$\det \begin{bmatrix} x_1 & x_2d \\ x_2 & x_1 \end{bmatrix} \neq 0$$

Therefore t is nonsinglar if and only if $x_1^2 - dx_2^2 \neq 0$.

Remark: $x_1^2 - dx_2^2 = 0$ gives the condition for $x_1e_1 + x_2\,e_2$ to be a singular element or an element not possessing a multiplicative inverse.

Corollary: In the case of F = Q, the field of rational numbers, $A_Q(d,0)$ is a field if either d is not a perfect square or is negative.

For, when d is not a perfect square or is negative, $x_1^2 - dx_2^2 = 0$ is possible only for $x_1 = x_2 = 0$, or $A_Q(d,0)$ has no non-trivial singular elements when either d is not a perfect square or when d is negative.

We recall that an element w in $A_F(d,0)$ is an idempotent if $w^2 = w$. Set $w = xe_1 + ye_2$; $x, y \in F$. The criterion for w to be an idempotent yields

$$(2.7) \qquad \begin{aligned} x &= x^2 + dy^2 \\ y\,(2x - 1_F) &= 0 \end{aligned}$$

When y = 0, one has x = 0 or $x = 1_F$. They correspond to the trivial idempotents. The non-trivial idempotents are

$$(2.7) \qquad w_1 = \tfrac{1}{2} e_1 + \frac{1}{2\sqrt{d}} e_2 \text{ and } w_2 = \tfrac{1}{2} e_1 - \frac{1}{2\sqrt{d}} e_2$$

if $\sqrt{d} \in F$.

If $\sqrt{d}$ is not an element of F, $A_F(d,0)$ has only the trivial idempotents. We state this as

THEOREM: 3. The non-trivial idempotents (if any) of $A_F(d,0)$ are the solutions of

$$2x - 1_F = 0 \quad \text{and} \quad x^2 - x + dy^2 = 0$$

in F^2.

3. APPLICATION TO SPLITTING FIELDS:

In the case when $F = Q$, the field of rational numbers, we consider an irreducible quadratic polynomial $p(x) = x^2 + a_1x + a_2$ where $a_1, a_2 \in Q$. If $d = a_1^2 - 4a_2$ is not a perfect square of a rational number, the splitting field of $p(x)$ over Q is $Q(\sqrt{d})$.

We look at $A_Q(d,0)$ where $d = a_1^2 - 4a_2$. If d is not a perfect square, $A_Q(d,0)$ is a field and as a vector space over Q, it is of dimension 2. Now, $\frac{1}{2}(-a_1e_1 \pm e_2)$ are elements of $A_Q(d,0)$.

If $u = \frac{1}{2}(-a_1e_1 + e_2)$, $v = \frac{1}{2}(-a_1e_1 - e_2)$

u and v will make the polynomial in A, namely,

$$A^2 + a_1A + a_2e_1 = 0 \text{ in } A_Q(d,0).$$

That is, $A_Q(d,0)$ contains the zeros of the polynomial $p(x)$ and it is the smallest such field as $[A_Q(d,0): Q] = 2$. Therefore $A_Q(d,0)$ is a splitting field of $p(x) = x^2 + a_1x + a_2$ and is isomorphic to $Q(\sqrt{d})$ where $d = a_1^2 - 4a_2$.

We remark that in the case of the splitting field K of a cubic polynomial over Q, as $[K: Q] \leq 3!$, the degree of the extension of Q is in general greater than 2. Therefore, we do not get an analogue of A_Q (d,0) for evolving a commutative algebra of dimension 3 or more in a single stroke. However, it is possible to construct a finite-dimensional commutative algebra which is isomorphic to the splitting field of the given cubic polynomial. The working appears to be more involved than the direct method of obtaining the splitting field.

REFERENCE

1. Steven C. Athoen and Lawrence D. Kugler : When is $\mathbb{R}^2$ a division algebra? Amer. Math. Monthly 90 (1983) 625-635.

February 26, 1995.

ON THE SYMBOLIC-NUMERICAL METHODS FOR FINDING THE ROOTS OF AN ARBITRARY SYSTEM OF NON-LINEAR ALGEBRAIC EQUATIONS

Tran Quoc-Nam *

RISC–Linz,

Research Institute for Symbolic Computation

Johannes Kepler University, A–4040 Linz, Austria

E-mail: tqnam@risc.uni-linz.ac.at

Abstract

The problem of finding the roots (real and complex) of a system of non-linear algebraic equations is investigated. For finding all of the roots of an univariate polynomial, we describe a modified version of Uspensky's method which is more efficient and practical. For systems of algebraic equations with zero-dimensional zero-set, we find a Gröbner Basis with respect to the total degree ordering, and transform it to a basis with respect to the pure lexicalgraphic ordering to solve the problem symbolically. For arbitrary systems of equations with non-zero-dimensional zero-set, where all of the current methods fail to find a particular root of the system, we describe a method of the author's which allows us to find some of the roots of the system iteratively.

Keywords: Symbolic Computation, Polynomial System Solving, Hybrid Symbolic-Numerical Methods.

1 Introduction

In this paper, we investigate the problem of finding the roots (real and complex) of a system of non-linear algebraic equations. In the case of one equation in one variable we can use, for example, Uspensky's method or Sturm's sequence, etc., to isolate and approximate to an arbitrary precision all of the real roots (see [Usp48, Loo82, Joh91]) and all of the complex roots (see [Leh61, Pin73]) of the equation. In this paper, we describe a modified version of Uspensky's method which is more efficient and practical. The main ideas of the algorithm are the use of interval arithmetic and avoidance of recursive calls. The speed up of the new algorithm is considerable, especially for applications such as algebraic curve plotting, in which we need to count the number of real roots or isolate them within many different intervals; or root isolation for polynomials with small minimal separation of roots.

For systems of multivariate equations, when the zero-set of the system is zero-dimensional, we can use one of the elimination methods such as resultants, characteristic sets or Gröbner Bases to solve the problem symbolically. In this paper, we use the Gröbner Bases method. But, instead of compute a Gröbner Basis with respect to the pure lexicalgraphic ordering directly, we find a Gröbner Basis with respect to the total degree ordering (which is much easier to compute) and transform it to a basis with respect to the pure lexicalgraphic ordering to solve the problem symbolically (see [Buc65, Buc85, Buc92, Win88], and [Win94, Win95, BW93, CLO92, Laz89, FGLM93]).

When the zero-set of the system has more than zero dimension then none of the existing methods can be used for finding a particular root of the system (what we get is a "reduced" system of algebraic equations). Unfortunately, this case appears frequently in many problems such as: Surface-to-Surface intersection problem in CAGD, robotic, etc. (see [Nam95b, Nam95a]). In this paper, we present a new technique for determining the roots of an arbitrary system of equations iteratively, where the equations can be nonlinear algebraic or transcendental. The number of equations of the system can be greater than, less than or equal to the number of variables. When the number of equations is equal to the number of variables, the method is similar to Newton's method. The method is wildly applicable. It is especially useful for applications in which we need to find some particular roots from a non-zero-dimensional zero-set of a system of equations. The method is not a pure numerical method; it is a hybrid symbolic–numerical method.

The structure of the paper is as follows. In Section 2, we describe a modified version of Uspensky's method for root (real and complex) isolation of polynomials with integral coefficients as well as algebraic number coefficients, which is more efficient and practical. In Section 3, we describe the Gröbner Bases transformation technique and its application for solving zero-dimensional systems of multivariate algebraic equations symbolically. In Section 4, we give some basic definitions and theorems of a new method of the author's for determining the roots of an arbitrary system of equations, especially when the zero-set of the system has more than zero dimension. We prove some

*Supported by the Austrian Fonds zur Förderung der wissenschaftlichen Forchung, Project POSSO, Nr. P9181-TEC.

theorems which are concerned with the correctness of the method. In Section 5, we report experimental results.

2 Modified Uspensky's Method

There are several methods for isolating real roots, for example: the methods based on Sturm's theorem, Descartes' lemma, and Rolle's theorem. Of these the method based on Descartes' lemma is the most efficient in practice (see [Usp48, Joh91, BR90]). For isolating complex roots, we refer to [Pin73, Leh61].

Here, we recall some basic definitions and theorems for real root isolation of polynomials with real algebraic number coefficients. Readers can find proofs of most of the theorems in this section in, for instance [Usp48, Loo82].

Definition 2.1 (Variation) *Let $\bar{a} = (a_0, \ldots, a_n)$ be a sequence of real numbers. Then the* variation *(or number of sign changes) of $\bar{a}$, denoted $V(\bar{a})$, is the number of pairs $(i, i+k), k > 1$ such that:*

1. $a_i \cdot a_{i+k} < 0$.
2. $a_{i+r} = 0$ *for* $0 < r < k$.

Definition 2.2 (Root Bound) *Let $P(x)$ be a polynomial with complex coefficients. A* root bound *for the polynomial $P(x)$ is a real number B such that if $P(\alpha) = 0$ then $|\alpha| < B$.*

Theorem 2.1 (Knuth's Bound) :
Let $P(x) = \sum_{i=0}^{m} a_i x^i$ be a polynomial with complex coefficients. If $B = \max_{1 \le k \le m} |a_{m-k}/a_m|^{1/k}$, then $2B$ is a root bound for $P(x)$

Theorem 2.2 (Descartes' lemma) :
Let $P(x) = \sum_{i=0}^{m} a_i x^i$ in $\mathbb{R}[x]$ and let $V(P)$ be the variation of the sequence $(a_0, \ldots, a_m)$. Then the number $Z_+(P)$ of strictly positive real roots (counted with multiplicity) of $P(x)$ does not exceed $V(P)$, and is congruent to $V(P)$ mod 2.

A real algebraic number, for instance $\sqrt{2}$, is a real number which satisfies an integral polynomial equation. A real algebraic number α will be represented by a square-free polynomial $A(x)$ (we do not require $A(x)$ to be an irreducible polynomial) such that $A(\alpha) = 0$ and an isolating interval, I, which distinguishes α from its real conjugates. The extension field of the rationals generated by α is denoted by $\mathbb{Q}(\alpha)$. An element, β, in $\mathbb{Q}(\alpha)$ is represented by the polynomial $B(x)$ in the residue class ring $\mathbb{Q}[x]/\langle A(x)\rangle$ such that $\beta = B(\alpha)$ (the representation is not unique). Arithmetic for real algebraic numbers is performed using polynomial operations in $\mathbb{Q}[x]/\langle A(x)\rangle$. In the same manner, for polynomial evaluation and transformation, polynomials with algebraic coefficients will be viewed as bivariate polynomials in $\mathbb{Q}(\alpha)[y]$ (by clearing denominators we can assume the polynomials are in $\mathbb{Z}[\alpha][y]$). Sign computation is performed as follows: If $A(\alpha) = 0$, $\alpha \in I = (a, b]$, and $\beta = B(\alpha)$, the sign of β can be computed by refining the isolating interval I until it contains no root of $B(x)$. Hence, the sign of $\beta = B(\alpha)$ is equal to the sign of $B(x)$ for any $x \in I$.

Now we describe an algorithm for isolating positive real roots of a polynomial with algebraic coefficients P in which we avoid using recursive calls. For the other roots, one has to replace y by $-y$.

Algorithm 2.1 (*Descartes*$(A(x), I, P(\alpha, y))$) :

In: *$A(x)$ is the defining polynomial for a real algebraic number α. I is an isolating interval for α. $P(\alpha, y) \not\equiv 0$ is a square-free polynomial in $\mathbb{Z}[\alpha, y]$, and $y \not| P(\alpha, y)$;*

Out: *L is a list of isolating intervals for the real positive roots of $P(\alpha, y)$.*

1. *Let $B \ge 1$ be a root bound of $P(\alpha, y)$. Put $Q(\alpha, y) = P(\alpha, B \cdot y)$ (all the positive roots of Q are in $[0, 1]$).*
2. *$Q^*(\alpha, y) \leftarrow (y+1)^m Q(\alpha, 1/(y+1))$ (the roots of Q^* in $[0, \infty)$ correspond to the root of Q in $(0, 1]$). $TList \leftarrow cons([Q^*(\alpha, y)], [])$; $L \leftarrow []$.*
3. **while** *$TList \ne []$ do*
   (a) *$GList \leftarrow TList$; $TList \leftarrow []$;*
   (b) **for** *$g \in GList$, compute $Var(Q^*) = Var(A(x), I, Q^*)$:*
      i. *If $Var(Q^*) = 0$ then Q^* has no positive root, so the same is true for P.*
      ii. *If $Var(Q^*) = 1$ then Q^* has one positive root, so the same is true for P. We put it to L.*
      iii. *If $Var(Q^*) > 1$, test if $Q(1/2) = 0$, and put:*
         A. *$TList \leftarrow cons(2^m Q(\alpha, y/2), TList)$.*
         B. *$TList \leftarrow cons(Q_1(\alpha, y+1) = 2^m Q(\alpha, (y+1)/2), TList)$.*

   endfor.

   endwhile.

Termination of the algorithm is guaranteed by the following theorem:

Theorem 2.3 *Let $P(x)$ be a square-free polynomial with real coefficients of degree m. If $P(x)$ has a single root in the interval $(0, 1)$ and all of the remaining real and complex roots are outside the circle of radius 1 centered at $(0, 0)$ and $(1, 0)$, then $Var((x+1)^m P(1/(x+1)) = 1$.*

An algorithm based on Descartes' lemma for isolating real roots of a polynomial within a given interval can be found in [Usp48, Joh91] or [BR90]. In that, we have first to transform the given interval into the unit interval $(0, 1)$. The transformation, with the use of subdivision method or Horner's method, is efficient. However, the length of the coefficients of the transformed polynomial (both numerator and denominator) roughly multiplies by the degree of the polynomial thus making the algorithm impractical.

We describe an algorithm for isolating real roots of a polynomial P within a given interval J. Our algorithm does not use that transformation. In fact, it uses interval arithmetic. The algorithm is much more efficient than the transformation approach as well as Sturm's sequence method.

Algorithm 2.2 (*Descartes*$(A(x), I, P(\alpha, y), J)$) :

In: $A(x)$ *is the defining polynomial for a real algebraic number* α. I *is an isolating interval for* α. $P(\alpha, y) \not\equiv 0$ *is a square-free polynomial in* $\mathbb{Z}[\alpha, y]$, *and* $y \not| P(\alpha, y)$; J *is an open interval.*

Out: L *is a list of isolating intervals for the real roots of* $P(\alpha, y)$ *within the open interval* J.

1. *Let Lp be a list of isolating intervals for all of the real roots (positive and negative) of the polynomial in the real axis. We can efficiently compute Lp by using the Algorithm 2.1.*
2. **for** $l \in Lp$,
   (a) *If* $l \in J$ *then* $L \leftarrow cons(l, L)$.
   (b) *Else if* l *and* J *are disjointed then* **next**.
   (c) *Else refine the interval* l *using, say bisection method, until it is contained in* J *or disjointed from* J.

   endfor

Table 1 compares computing times of the Algorithm 2.2, transformation approach and Sturm's sequence. Listed are averages of computing times in seconds for isolating real roots within an open interval of width 2^{-100}. All input polynomials had randomly generated 32 bits coefficients. The degrees were varied from 10 to 100. The algorithms were implemented in Maple V.3. Timing results were obtained on a SGI/R4000A (RISC) machine under IRIX 5.2.

Table 1: Computing times comparison

deg	new method	transf. method	Sturm's seq.
10	0.14	0.87	2.74
20	0.36	10.06	49.05
30	0.68	45.26	258.54
50	1.34	315.24	2208.43
100	3.52	3976.66	39605.05

The algorithm can be extended for isolation of complex roots up to an arbitrary precision. In that case, instead of produces disjoint isolating interval in the real axis, the algorithm produces disjoint isolating rectangles for all of the roots of the polynomial in the complex plane, i.e. each such rectangle will contain exactly one root of the polynomial, and each root of the polynomial will be contained in one rectangle. The sides of the rectangles will be parallel to the coordinate axis.

3 Gröbner Bases Transformation

In this section, we will investigate how the Gröbner Bases transformation technique can be applied to solve systems of polynomial equations in several variables. Especially when the system has finitely many roots. For basic facts on the theory of Gröbner Bases itself, we refer to [Buc65, Buc85, CLO92, Win95]. Hereafter, let K be a field of characteristic 0.

3.1 Computation in the vector space $K[X]_{/I}$

The ring $K[X]_{/I}$ of polynomials modulo the ideal I is a vector space over K. A Gröbner Basis G provides a basis for this vector space.

Theorem 3.1 *The irreducible power product modulo* G, *viewed as polynomials with coefficient* 1, *form a basis for the vector space* $K[X]_{/I}$ *over* K.

Proof: Let B be the set of irreducible power products modulo G, viewed as polynomials with coefficient 1. Clearly B generates $K[X]_{/I}$, since every polynomial can be reduced to a linear combination of element of B with coefficients in K. Moreover, B is linearly independent, because any nontrivial linear combination of elements of B is irreducible modulo G and therefore different from 0 in $K[X]_{/I}$. ∎

3.2 Solution of Algebraic Equations

We consider a system of equations:

$$\begin{cases} f_1(x_1, \ldots, x_n) & = & 0, \\ \vdots & & \\ f_m(x_1, \ldots, x_n) & = & 0, \end{cases}$$

where $f_1, \ldots, f_m \in K[X]$. The system is called a system of polynomial or algebraic equations. First let us decide whether the system has any solution in $\bar{K}^n$, $\bar{K}$ being the algebraic closure of K. Let $I = \langle f_1, \ldots, f_m \rangle$.

Theorem 3.2 *Let* G *be a normed Gröbner Basis of* I. *The system is unsolvable in* $\bar{K}^n$ *if and only if* $1 \in G$.

Proof: If $1 \in G$ then $1 \in \langle G \rangle = I$, so every solution of the system is also a solution of $1 = 0$. Hence there can be no solution.

On the other hand, assume that the system is unsolvable. Then the polynomial 1 vanishes on every common root of the system. So by Hilbert's Nullstellensatz $1 \in radical(I)$ and therefore also $1 \in I$. Since G is a normed Gröbner Basis if I, we must have $1 \longrightarrow_G 0$. This is only possible if $1 \in G$. ∎

Now suppose that the system is solvable. We want to determine whether there are finitely or infinitely many solution of the system or, in the other words, whether or not the ideal I is zero-dimensional.

Theorem 3.3 *Let G be a Gröbner Basis of I. Then the system has finitely many solutions (i.e. I is zero-dimensional) if and only if for every i, $1 \leq i \leq n$, there is a polynomial $g_i \in G$ such that $lpp(g_i)$ is a pure power of x_i.*

Proof: I is zero-dimensional if and only if $K[X]_{/I}$ has finite vector space dimension over K (see, e.g, [Grö49]). By the previous theorem, that is the case if and only if the number of irreducible power products modulo G is finite, i.e. for every variable x_i there is a pure power of it in $lpp(G)$. ■

3.3 Elimination Property of Gröbner Bases

The role of Gröbner Bases method in solving system of algebraic equations is the same as that of Gaussian elimination in solving system of linear equations, namely to triangularize the system, or carry out the elimination process. The crucial observation, first stated in [Tri78], is the elimination property of Gröbner Bases. It states that if G is a Gröbner Basis of I with respect to the lexicalgraphic ordering with $x_1 < \ldots < x_n$, then the i-th elimination ideal of I, i.e. $I \cap K[x_1, \ldots, x_i]$, is generated by those polynomials in G that depend only on the variables $x_1, \ldots, x_i$

Theorem 3.4 (The Elimination Theorem) *:*
Let G be a Gröbner Basis of I with respect to the lexicalgraphic ordering with $x_1 < \ldots < x_n$. Then

$$I \cap K[x_1, \ldots, x_i] = \langle G \cap K[x_1, \ldots, x_i] \rangle$$

Proof: Obviously the right hand side is contained in the left hand side. On the other hand, let $f \in I \cap K[x_1, \ldots, x_i]$. Then $f \longrightarrow^{*}_{G} 0$. All the polynomials in this reduction depend only on the variables $x_1, \ldots, x_i$. So we get a representation of f as a linear combination $\sum h_j g_j$, where $g_j \in G \cap K[x_1, \ldots, x_i]$ and $h_j \in K[x_1, \ldots, x_i]$. ■

Let us restrict our attention to the case where we eliminate just the last variable x_n. Thus we want to know if a partial solution $(a_1, a_2, \ldots, a_{n-1}) \in V(I \cap K[x_1, \ldots, x_{n-1}])$ can be extended to a solution $(a_1, a_2, \ldots, a_{n-1}, a_n) \in V(I)$. The following theorem tells us when this can be done.

Theorem 3.5 (The Extension Theorem) *:*
Let $I = \langle f_1, \ldots, f_s \rangle \in K[X]$ and let I_{n-1} be the (n-1)-th elimination ideal of I, i.e. $I \cap K[x_1, \ldots, x_{n-1}]$. For each $1 \leq i \leq s$, write f_i in the form $f_i = g_i(x_1, \ldots, x_{n-1}) \cdot x_1^{N_i} +$ terms in which x_1 has degree $< N_i$, where $N_i \geq 0$ and $g_i \in K[x_1, \ldots, x_{n-1}]$ is non-zero. (We set $g_i = 0$ when $f_i = 0$.) Suppose that we have a partial solution $(a_1, a_2, \ldots, a_{n-1}) \in V(I \cap K[x_1, \ldots, x_{n-1}])$ then there exists $a_n \in K$ such that $(a_1, a_2, \ldots, a_{n-1}, a_n) \in V(I)$.

Proof: see [CLO92]. ■

3.4 Transformation of Gröbner Bases

In case the system has finitely many solutions, we can solve the system symbolically or determine up to an arbitrary precision all of the solutions of the system.

Example 3.1 Consider the system of equations:

$$\begin{cases} f_1 &= 4xz - 4xy^2 - 16x^2 - 1 = 0, \\ f_2 &= 2y^2z + 4x + 1 = 0, \\ f_3 &= 2x^2z + 2y^2 + x = 0 \end{cases}$$

are polynomials in $\mathbb{Q}[x, y, z]$. We are looking for solutions of this system of algebraic equations in $\bar{\mathbb{Q}}^3$ where $\bar{\mathbb{Q}}$ is the field of algebraic numbers.

Let $<$ be the lexicalgraphic ordering with $x < y < z$. The Buchberger's algorithm applied to $F = \{f_1, f_2, f_3\}$ yields (after reducing the result) the reduced Gröbner Basis $G = \{g_1, g_2, g_3\}$, where

$$\begin{cases} g_1 &= 65z + 64x^4 - 432x^3 + 168x^2 - 354x + 104, \\ g_2 &= 26y^2 - 16x^4 + 108x^3 - 16x^2 + 17x, \\ g_3 &= 32x^5 - 216x^4 + 64x^3 - 42x^2 + 32x + 5 \end{cases}$$

By the previous theorems the system is solvable. Furthermore the system has finitely many solutions. The Gröbner Basis G yields an equivalent triangular system in which the variables are completely separated. So we can get solutions by solving the univariate polynomial g_3 and propagating the partial solutions upwards to solutions of the full system. The univariate polynomial g_3 is irreducible over $\mathbb{Q}$, and the solutions are $(\alpha, \pm\frac{\sqrt{\alpha}\sqrt{16\alpha^3+16\alpha-17}}{\sqrt{26}}, -\frac{64\alpha^4-432\alpha^3+128\alpha^2-354\alpha+104}{65})$ where α is a root of g_3. We can also use Uspensky's method to determine up to an arbitrary precision all of the solutions of the system. ■

However, the pure lexicalgraphic ordering leads to computations which are much more time and memory space consuming than with total degree ordering and even are often intractable; the corresponding complexity has been proved to be $d^{O(n^3)}$ when the number of solutions is finite ([NGH88]). Hereafter, we discuss an algorithm which allows us to compute a Gröbner Basis w.r.t the total degree ordering and after that transform it to a Gröbner Basis w.r.t the pure lexicalgraphic ordering (see [Laz89, FGLM93]). The method is practical and in many case, the only way for computing the lexicalgraphical Gröbner Basis ; the algorithm reduces the corresponding complexity to $d^{O(n^2)}$. The other technique on the efforts to improve the Buchberger's algorithm we should mention here is Sugar strategy ([GMN+91]).

Proposition 3.1 *Let I be a zero dimensional ideal and G_1, the reduced Gröbner Basis with respect to an admissible ordering $<_1$. Given a different ordering $<_2$, it is possible to construct the Gröbner Basis G_2 with respect to the ordering $<_2$ with $O(nd^3)$ arithmetic operations.*

Proof: see [FGLM93]. ■

Procedure 3.1 (Basis Transformation) .

In: $<$ *a new admissible ordering.*
oldBasis, a Gröbner Basis of a zero dimensional ideal, with respect to some ordering.

Out: newBasis, the reduced Gröbner Basis of the ideal generated by oldBasis with respect to the ordering $<$.

- *Subfunctions* :
  - *NormalForm(polynom), returns the reduced form of a polynomial with respect to oldBasis.*
  - *NextMonom, removes the first element of ListOfNexts and returns it, returns nil if the list is empty.*
  - *InsertNexts(monom), adds to ListOfNexts the products of monom by all variables, sorts this list by increasing ordering for* $<$ *and remove duplicates.*
- *Localvariables* :
  - *staircase, the list of leading monomials of the elements of newBasis.*
  - *MBasis, a list of pairs* $[a_i, b_i]$ *where* $[\ldots, a_i, \ldots]$ *is the list of monomials which are in normal form with respect to newBasis and* $b_i = NormalForm(a_i)$*, the normal form of* a_i *with respect to oldBasis. We select each pair with selectors first and second;*
  - *ListOfNexts, the list of "next" monomials to be considered sorted by increasing ordering for* $<$.

1. *MBasis* := []; *staircase* := []; *newBasis* := []; *ListOfNexts* := [];
2. *monom* := 1
3. **while** *monom* $\neq$ *nil*
   (a) *if monom is not a multiple of some element of staircase then*
      i. *vector* := *NormalForm(monom)*;
      ii. *if there exist a linear relation* $vector + \sum_{v \in MBasis} \lambda_v \cdot second(v) = 0$ *then*
         A. $pol := monom + \sum_{v \in MBasis} \lambda_v first(v)$
         B. *newBasis* := *cons(pol, newBasis)*;
         C. *staircase* := *cons(monom, staircase)*;
      else
         A. *MBasis* := *cons([monom, vector], MBasis)*;
         B. *InsertNexts(monom)*;
   (b) *monom* := *NextMonom*

   endwhile.
4. *Return.*

3.4.1 Termination of the Algorithm

We have now to prove that the algorithm stops. We are in zero-dimensional case, so the maximal number of linearly independent irreducible polynomials is the (finite) dimension of $K[X]_{/I}$, i.e. the number of irreducible monomials for an Gröbner Basis . Thus the number of iterations which increase *MBasis* is finite. When *MBasis* is complete *ListOfNexts* may no more increase, thus the number of remaining iterations is bounded.

3.4.2 Correctness of the Algorithm

We first prove that the elements of *MBasis* are linearly independent modulo the ideal generated by *oldBasis*; if this were false, there would be a linear combination $P = \sum \lambda_i \cdot m_i$ of elements of *MBasis* which is a polynomial in $\langle oldBasis \rangle$. Then

$$NormalForm(P) = \sum \lambda_i \cdot NormalForm(m_i) = 0$$

and this gives a relation which implies that the algorithm would put in *staircase* rather than in *MBasis* the greatest of the m_i such that $\lambda_i \neq 0$. The elements of *newBasis* are in $\langle oldBasis \rangle$: if $P = monom + \sum \lambda_v \cdot m_v$ is such an element, then $NormalForm(P) = NormalForm(monom) + \sum \lambda_v \cdot NormalForm(m_v) = 0$ by construction of P.

Finally the elements of *staircase* are the leading monomials of the elements of *newBasis*. This is clear because the loop works with increasing monomials.

Now, each monomial which is not in *MBasis* nor in *staircase* is multiple of some element of *staircase*. This is clear if monomial appeared in *ListOfNexts*; otherwise suppose that the monomial is of the form $m_1 m_2$ where m_1 is in *MBasis* and maximal for this property. In this case the monomial, if not in *MBasis*, is a multiple of some element which appeared in *ListOfNexts* and was not put in *MBasis*, has been put in *staircase* or is a multiple of some element of *staircase*. Thus, the normal form of a polynomial with respect to *newBasis* is a linear combination of elements of *MBasis* and the normal form of an element of $\langle oldBasis \rangle$ is zero (first assertion above).

This proves that *newBasis* is a Gröbner Basis for the new ordering; it is minimal and reduced because, by construction, none of its monomials is a multiple of a leading monomial other than itself.

4 Extended Newton Method

In this section, we give some basic definitions and theorems of a new method of the author for determining the roots of an arbitrary system of equations iteratively, where the equations can be nonlinear algebraic or transcendental. The number of equations of the system can be greater than, less than or equal to the number of variables. Details of this method can be found in [Nam94a, Nam94b].

Let $x = [x_1, x_2 \ldots, x_k]^T$ be a k-dimensional column vector with components $x_1, x_2, \ldots, x_k$ in $\mathbb{C}$ and $f(x) = [f_1(x), \ldots, f_h(x)]^T$ an h-dimensional vector valued function, i.e., a column vector with components $f_1(x), f_2(x) \ldots, f_h(x)$ in $\mathbb{C}$. To deal with the norm of a vector, we can choose any one of the norms, for example: $\|x\|_\infty \equiv \max_{1\leq i\leq k} |x_i|$, $\|x\|_1 \equiv \sum_{i=1}^k |x_i|$ or $\|x\|_2 \equiv \sqrt{\sum_{i=1}^k |x_i|^2}$ We define the Jacobian matrix of f with respect to x to be the $h \times k$ matrix:

$$J_f(x) \equiv J(x) \equiv \left(\frac{\partial f_i(x)}{\partial x_j}\right), \; i = 1..h, \; j = 1..k. \quad (1)$$

When $h = k$ and with the assumption that det $|J(x)| \neq 0$ for x in a closed and bounded set $\mathcal{S}$, we will denote the inverse Jacobian of f at x by $J_f^{-1}(x)$. In this case, Newton's method is a well-known iterative method for determining the roots of the equation $f(x) = 0$. The iteration can be written in the form $x_{n+1} = g(x_n)$, where

$$g(x) = x - J_f^{-1}(x) \cdot f(x). \quad (2)$$

Unfortunately, when $h \neq k$, the Jacobian matrix of f is not a square matrix. Therefore, it cannot be inverted. Hence, Newton's method fails to apply to such cases. We extend Newton's method such that the extended method can be applied to these cases as well.

To get around the problem of a non-square matrix, we will use what we call the pseudo inverse of the Jacobian matrix.

Definition 4.1 *A pseudo inverse of a matrix A is a matrix $A^\perp$ satisfying*

$$\begin{aligned} AA^\perp A &= A \\ A^\perp A A^\perp &= A^\perp \\ (A^\perp A)^* &= A^\perp A \\ (AA^\perp)^* &= AA^\perp \end{aligned}$$

where $(\cdot)^*$ denotes the conjugate transpose of the matrix (in case of real coefficient functions, we will use $(\cdot)^T$ for $(\cdot)^*$).

Theorem 4.1 *For any matrix A there exists a pseudo inverse matrix $A^\perp$. Moreover the pseudo inverse is unique.*

Proof: see [BO71]. ■

There are several techniques for obtaining the pseudo inverse of a numerical matrix, for example Hestenes' technique or Greville's technique (see [BO71]). However, those techniques are pure numerical techniques, i.e., they can only be applied to each individual numerical matrix; and we have to take into the account the problems of rounding error, word length, precision of arithmetic, and so on. Moreover, when we use such techniques in our iterative method for determining the roots of a system of equations, we have to apply the approximate process for obtaining the pseudo inverse matrix $J^\perp$ at each iterate $x^{(n)}$. That means that it is time consuming and has accumulated errors.

To avoid such problems, we try to find a symbolic formula and use symbolic computation for the method. For example:

$$J^+ = J^T \cdot (J \cdot J^T)^{-1} \quad \text{or} \quad J^- = (J^T \cdot J)^{-1} \cdot J^T$$

Lemma 4.1 *For all x, if $(J \cdot J^T)$ is nonsingular then $J^+(x)$ is the pseudo inverse of matrix $J(x)$*

Proof: The matrix $J^+(x)$ is defined when $(J(x) \cdot J^T(x))$ is nonsingular; and it is easy to check that $J^+(x)$ satisfies the definition 4.1. ■

Lemma 4.2 *For all x, if $(J^T \cdot J)$ is nonsingular then $J^-(x)$ is the pseudo inverse of matrix $J(x)$*

Lemma 4.3 *For all x, if both $J^+(x)$ and $J^-(x)$ exist then they are identical.*

Matrix J^+ or J^- is a matrix of order $k \times h$. Hence we can define the function g as follows:

$$g(x) = x - J^+ \cdot f(x) \quad \text{or} \quad g(x) = x - J^- \cdot f(x) \quad (3)$$

where $g(x)$ and x are vectors of the same dimension k. We find a fixed point of the function g with respect to the initial point $x^{(0)}$ by the following algorithm:

Algorithm 4.1 (Extended Newton Method) .

In: a function $f : \mathbb{C}^k \to \mathbb{C}^h$, an initial point $x^{(0)}$, a tolerance ϵ and a maximum number of iterates M.

Out: a complex fixed point of the function g.

1. $x^{(n+1)} \leftarrow x^{(0)}; L \leftarrow 0$
   If $h < k$ *then* $g(x) = x - J^+ \cdot f(x)$
   else if $h > k$ *then* $g(x) = x - J^- \cdot f(x)$
   else $g(x) = x - J^{-1} \cdot f(x)$
2. *do*
   If $L > M$ *then return("failure").*
   $x^{(n)} \leftarrow x^{(n+1)}$;
   $x^{(n+1)} \leftarrow g(x^{(n)})$;
   $L \leftarrow L + 1$;
   While $\|x^{(n+1)} - x^{(n)}\| \geq \epsilon$
3. *return*($x^{(n+1)}$).

Under some conditions for the function f, as will be shown in the Section 4.1, the sequence $\{x^{(0)}, x^{(1)}, \ldots, x^{(n)}, \ldots\}$ will converge to the fixed point of g in a neighborhood of $x^{(0)}$. Let's call it α. We will have $f(\alpha) = 0$.

When $h = k$, in case J^+, J^- and J^{-1} exist, we have $J^+ = J^- = J^{-1}$. That means our method is similar to Newton's method when the system has n equations in n variables. Furthermore, we have $J \cdot J^+ = I$, $J^- \cdot J = I$. Therefore, we can consider J^+ and J^- to be *half-inverse* matrices of the Jacobian matrix J.

4.1 Correctness and Convergence of the Algorithm

Given a function $f : \mathbb{C}^k \to \mathbb{C}^h$, and a point $x^{(0)} = [x_1^{(0)}, x_2^{(0)} \ldots x_k^{(0)}]^T$, we construct a sequence:

$$x^{(n+1)} = g(x^{(n)}) \tag{4}$$

where g is defined by equation (3). Now we consider some conditions in which the sequence will converge:

Theorem 4.2 *Let $g(x)$ satisfy*

$$||g(x) - g(y)|| \leq \lambda ||x - y|| \tag{5}$$

for all vectors x, y such that $||x - x^{(0)}|| \leq \rho$, $||y - x^{(0)}|| \leq \rho$ with the Lipschitz constant, λ, satisfying

$$0 \leq \lambda < 1.$$

Let the initial iterate, $x^{(0)}$, satisfy

$$||g(x^{(0)}) - x^{(0)}|| \leq (1 - \lambda) \cdot \rho \tag{6}$$

Then

(i) all iterates, as in (4), satisfy

$$||x - x^{(0)}|| \leq \rho \tag{7}$$

(ii) the iterates converge to some vector, say α which is the root of the equation $f(x) = 0$;

(iii) α is the only root of f in $||x - x^{(0)}|| \leq \rho$.

Proof: see [Nam94b]. ■

Theorem 4.3 *Let (4) have a root $x = \alpha$. Let the components $g_i(x)$ have continuous first partial derivatives and satisfy*

$$\left| \frac{\partial g_i(x)}{\partial x_j} \right| \leq \frac{\lambda}{k}, \quad \lambda < 1; \tag{8}$$

for all x in

$$||x - \alpha||_\infty \leq \rho \tag{9}$$

Then:

(i) For any $x^{(0)}$ satisfying (9) all the iterates $x^{(n)}$ of (4) also satisfy (9).

(ii) For any $x^{(0)}$ satisfying (9), the iterates converge to the root α of $x = g(x)$ which is unique in (9).

Proof: For any two points x, y in (9) we have by Taylor's theorem:

$$g_i(x) - g_i(y) = \sum_{j=1}^{k} \frac{\partial g_i(\xi^{(i)})}{\partial x_j} \cdot (x_j - y_j), \quad i = 1..k \tag{10}$$

where $\xi^{(i)}$ is a point on the open line segment joining x and y. Thus $\xi^{(i)}$ is in (9), and using (8) yields

$$\begin{aligned} g_i(x) - g_i(y) &\leq \sum_{j=1}^{k} \frac{\partial g_i(\xi^{(i)})}{\partial x_j} \cdot (x_j - y_j) \\ &\leq ||x - y||_\infty \cdot \sum_{j=1}^{k} \frac{\partial g_i(\xi^{(i)})}{\partial x_j} \\ &\leq \lambda \cdot ||x - y||_\infty . \end{aligned}$$

Since the inequality holds for each i, we have

$$||g(x) - g(y)||_\infty \leq \lambda \cdot ||x - y||_\infty, \tag{11}$$

and thus we have proven that $g(x)$ is Lipschitz continuous in the domain (9), with respect to the indicated norm. Now note that for any $x^{(0)}$ in (9),

$$\begin{aligned} ||x^{(1)} - \alpha||_\infty &= ||g(x^{(0)}) - g(\alpha)||_\infty \\ &\leq \lambda \cdot ||x^{(0)} - \alpha||_\infty \\ &\leq \lambda \cdot \rho \end{aligned}$$

and so $x^{(1)}$ is also in (9). By an obvious induction we have then

$$\begin{aligned} ||x^{(n)} - \alpha||_\infty &= ||g(x^{(n-1)}) - g(\alpha)||_\infty \\ &\leq \lambda \cdot ||x^{(n-1)} - \alpha||_\infty \\ &\vdots \\ &\leq \lambda^n \cdot ||x^{(0)} - \alpha||_\infty \\ &\leq \lambda^n \cdot \rho \end{aligned} \tag{12}$$

and hence all $x^{(n)}$ is also in (9). The convergence immediately follows from (12) since $\lambda < 1$. The uniqueness follows as in the previous theorem. ■

If the initial iterate $x^{(0)}$ is sufficiently close to the root α of $f(x) = 0$, then the previous theorems can be used to prove that the iterates, $x^{(n)}$, defined in (4) converge to the root. However, we do not know from these results if a given initial iterate $x^{(0)}$ is close enough to the unknown root, α. Instead, from Theorems 4.2 and 4.3, we can derive some conditions for the method to be convergent which depend only on properties of the original function f instead of g; and these conditions may be explicitly checked without a knowledge of α.

Corollary 4.1 *Let the initial iterate $x^{(0)}$ be such that the Jacobian matrix $J(x^{(0)})$ defined in (1) has a half-inverse with norm bounded by*

$$||J^{\perp}(x^{(0)})|| \leq a \tag{13}$$

Let the difference of the first two iterates be bounded by

$$||x^{(1)} - x^{(0)}|| = ||J^{+}(x^{(0)}) \cdot f(x^{(0)})|| \leq b \tag{14}$$

Let J be normal, i.e., $J \cdot J^T = J^T \cdot J$; and let the components of $f(x)$ have continuous second derivatives which satisfy

$$\sum_{j'=1}^{k} \left| \frac{\partial^2 f_i(x)}{\partial x_j \, \partial x_{j'}} \right| \leq \frac{c}{k}, \tag{15}$$

for all x in $||x - x^{(0)}|| \leq 2b$; $i, j = 1..k$. If the constants a, b and c are such that

$$a \cdot b \cdot c \leq 1/2. \tag{16}$$

Then

(i) The iterates are uniquely defined and lie in the "2b-sphere" about $x^{(0)}$:

$$||x^{(n)} - x^{(0)}|| \leq 2b.$$

(ii) The iterates converge to some vector, say α, for which $f(\alpha) = 0$, and

$$||x^{(n)} - \alpha|| \leq \frac{2b}{2^n}. \quad (17)$$

Lemma 4.4 *If the function $f(x)$ has two derivatives and $(J^T \cdot J)$ is non-singular at the root then the convergence of the extended method is quadratic.*

Proof: If the function $g(x)$ is such that at a root α,

$$G(\alpha) \equiv \left(\frac{\partial g_i(\alpha)}{\partial x_j}\right) = 0, \quad i,j = 1,2,\ldots,k,$$

and these derivatives are continuous near the root, then (8) can be satisfied for some $\rho > 0$. If, in addition, the second derivatives $\frac{\partial^2 g_i(x)}{\partial x_j \, \partial x_{j'}}$ all exist in a neighborhood of the root, then by Taylor's theorem

$$g_i(x) - g_i(\alpha) = \frac{1}{2}\sum_{j=1}^{k}\sum_{j'=1}^{k} \frac{\partial^2 g_i(\xi^i)}{\partial x_j \, \partial x_{j'}}(x_j - \alpha_j)(x_{j'} - \alpha_{j'}).$$

Now in the iteration, we find

$$||x^{(n)} - \alpha||_\infty \leq M||x^{(n-1)} - \alpha||^2_\infty, \quad (18)$$

where M is chosen such that

$$\max_{i,j,j'}\left|\frac{\partial^2 g_i(x)}{\partial x_j \, \partial x_{j'}}\right| \leq \frac{2M}{k^2}.$$

We use these iterations for the method

$$x^{(n+1)} = g(x^{(n)}) = x^{(n)} - J^-(x^{(n)}) \cdot f(x^{(n)}).$$

The jth column of $G(x)$ is given by

$$\begin{aligned}\frac{\partial g(x)}{\partial x_j} &= \frac{\partial x}{\partial x_j} - \frac{\partial}{\partial x_j}[J^-(x) \cdot f(x)] \\ &= \frac{\partial x}{\partial x_j} - J^-(x)\frac{\partial f(x)}{\partial x_j} - \frac{\partial J^-(x)}{\partial x_j} \cdot f(x).\end{aligned}$$

By setting $x = \alpha$ in the above and recalling that $f(\alpha) = 0$ and $J = (\partial f_i(x)/\partial x_j)$, we get

$$G(\alpha) = I - J^-(\alpha)J(\alpha) - 0 = 0.$$

Therefore, from (18) the convergence of the method is quadratic. ■

5 Experiments

The results in this section are base upon our implemented package which have been written in *Mathematica* 2.2 and run on SGI's machines.

5.1 Finding the Roots of a System of Transcendental Equations

Given a function $f : \mathcal{C}^2 \rightarrow \mathcal{C}$ defined by

$$f(x,y) = x\cos(y) + y\sin(x)$$

The half-inverse Jacobian matrix of f will be:

$$J^+ = \begin{pmatrix} \frac{y\cos(x)+\cos(y)}{(y\cos(x)+\cos(y))^2+(\sin(x)-x\sin(y))^2} \\ \frac{\sin(x)-x\sin(y)}{(y\cos(x)+\cos(y))^2+(\sin(x)-x\sin(y))^2} \end{pmatrix}$$

The iterative function g will be:

$$g(x,y) = \begin{pmatrix} x - \frac{(y\cos(x)+\cos(y))(x\cos(y)+y\sin(x))}{(y\cos(x)+\cos(y))^2+(\sin(x)-x\sin(y))^2} \\ y - \frac{(x\cos(y)+y\sin(x))(\sin(x)-x\sin(y))}{(y\cos(x)+\cos(y))^2+(\sin(x)-x\sin(y))^2} \end{pmatrix}$$

In Table 2 are some roots of the function f which have been obtained by our new method. Note that we cut of the number of digits after the decimal point of the results while they have an arbitrary precision.

5.2 Finding the Roots of a System of Non-linear Algebraic Equations

Given a function $f : \mathcal{C}^3 \rightarrow \mathcal{C}^2$ defined by

$$f(x,y) = \begin{pmatrix} f_1 = (x^2+y^2)^3 - 4x^2y^2 - z \\ f_2 = x^2 - y^3 + z^2 \end{pmatrix}$$

In Table 3 are some roots of the system of equations f which have been obtained by our new method. Note that we cut of the number of digits after the decimal point of the results while they have an arbitrary precision.

6 Conclusion

We investigated the problem of finding the roots (real and complex) of a system of non-linear algebraic equations. For one equation in one variable case, we describe a modified version of Uspensky's method which is more efficient and practical. The main ideas of the algorithm are the use of interval arithmetic and avoidance of recursive calls. The speed up of the new algorithm is considerable compared with both the transformation approach and Sturm's sequence, especially for applications such as algebraic curve plotting, in which we need to count the number of real roots or even isolate them within a lot of different intervals; or when the polynomial has very near roots. The idea can be extended to complex root isolation and we are doing so.

For systems of multivariate equations, when the zero-set of the system is zero-dimensional, instead of computing a Gröbner Basis with respect to the pure lexicalgraphic ordering directly, we find a Gröbner Basis with respect to the total degree ordering (which is much easier to compute) and transform it to the basis

Table 2: Example 5.1

starting point	timing (sec)	approximate root	accu.
$(0,0,0)$	0	$(0,0,0)$	0
$(1,5,3)$	10.11	$(0.0061016327,$ $1.0000041355,$ $0.9999875882)$	10^{-27}
$(7,3,8)$	10.25	$(0.0000420132,$ $1.000000000,$ $0.999999999)$	10^{-10}
$(1+2i,$ $2+i,5-i)$	19.64	$(0.1351451272+2.1862015587i,$ $2.7062430828+0.8010331484i,$ $4.7336305601+1.7422999187i)$	10^{-38}
$(i,1+7i,$ $8-5i)$	18.61	$(2.7171563948-0.1136309561i,$ $0.1242232605+3.8691969044i,$ $4.7764762212-5.9801301677i)$	10^{-39}

Table 3: Example 5.2

starting point	timing (sec)	approximate root	accu.
$(0,0)$	0.01	$(0,0)$	0
$(1,2)$	1.94	$(4.9157051307\,10^{-32},$ $2.0678891334)$	10^{-32}
$(7,3)$	1.96	$(10.5818632245,$ $1.4453993728)$	10^{-38}
$(1+i,2i)$	3.75	$(2.8332413914-0.7621733899i,$ $-1.3858095352+0.3868286655i)$	10^{-29}
$(5i,8+i)$	4.32	$(-5.13280\,10^{-61}+6.55666\,10^{-61}i,$ $8.2886327129+1.0533001303i)$	10^{-60}

with respect to the pure lexicalgraphic ordering to solve the problem symbolically. We are combining those two methods for isolation of real roots of a zero-dimensional system of algebraic equations.

For an arbitrary system of equations, we present a new technique for determining some of the roots of the system iteratively; where the equations can be nonlinear algebraic or transcendental. The number of equations of the system can be greater than, less than or equal to the number of variables. When the number of equations is equal to the number of variables, the method is similar to Newton's method. The method is wildly applicable. It is especially useful for applications in which we need to find some particular roots from a non-zero-dimensional zero- set of a system of equations. The method is not a pure numerical method; it is a hybrid symbolic–numerical method.

Acknowledgments

The author would like to thank Dr. Franz Winkler for his encouragements and his supports.

References

[BO71] T.L. Boullion and P.L. Odell, editors. *Generalized Inverse Matrices.* Wiley–Interscience, New York, 1971.

[BR90] R. Benedetti and Jean-Jacques Risler. *Real Algebraic and Semi-Algebraic Sets.* Hermann, Paris, 1990.

[Buc65] B. Buchberger. *An Algorithm for Finding a Basis for the Residue Class Ring of a Zero-dimensional Polynomial Ideal (in German).* PhD thesis, Institute of Mathematics, Univ. Innsbruck, Innsbruck, Austria, 1965.

[Buc85] B. Buchberger. Gröbner Bases: An Algorithmic Method in Polynomial Ideal Theory. In N. K. Bose, editor, *Recent Trends in Multidimensional Systems Theory*, chapter 6, pages 184–232. D. Reidel Publ. Company, 1985.

[Buc92] Bruno Buchberger. *Gröbner Bases: Elementary Theory and Applications*. Springer Verlag, 1992. To appear.

[BW93] Thomas Becker and Volker Weispfenning. *Gröbner Bases*. Graduate Texts in Mathematics, 141. Springer Verlag, 1993. In cooperation with Heinz Kredel.

[CLO92] David Cox, John Little, and Donal O'Shea. *Ideals, Varieties, and Algorithms*. Undergraduate Texts in Mathematics. Springer Verlag, 1992.

[FGLM93] F. C. Faugère, P. Gianni, D. Lazard, and T. Mora. Efficient Computation of Zero-dimensional Gröbner Bases by Change of Ordering. *Journal of Symbolic Computation*, 16:329–344, 1993.

[GMN+91] A. Giovini, T. Mora, G. Niesi, T. Robbiano, and C. Traverso. "One sugar cube, please" or selection strategy in Buchberger's algorithm. In *Proceeding of ISSAC-91*, pages 49–54, 1991.

[Grö49] W. Gröbner. *Moderne Algebraische Geometrie*. Springer Verlag, Wien-Innsbruck, 1949.

[Joh91] J. R. Johnson. *Algorithms for Polynomial Real Root Isolation*. PhD thesis, The Ohio State University, 1991.

[Laz89] D. Lazard. Solving Zero-dimensional Algebraic Systems. Research report series no. 89-48, Laboratoire Informatique Theorique et Programmation (LITP), 1989.

[Leh61] D. H. Lehmer. A machine method for solving polynomial equations. *Journal of the ACM*, 8(2):151–162, April 1961.

[Loo82] R. Loos. Computing in Algebraic Extensions. In B. Buchberger, G. E. Collins, and R. Loos, editors, *Computer Algebra*, pages 173–187. Springer Verlag, New York, 1982.

[Nam94a] Tran Quoc Nam. Extended Newton's method for finding the roots of an arbitrary system of equations and its applications. In M. H. Hamza, editor, *Proceedings of the Twelfth IASTED International Conference: Applied Informatics*. IASTED, 1994.

[Nam94b] Tran Quoc Nam. Extended Newton's method for finding the roots of an arbitrary system of equations and its applications. Technical Report 94–49, RISC-Linz, J. Kepler Universität, Linz, Austria, 1994.

[Nam95a] Tran Quoc Nam. A Hybrid Symbolic-Numerical Method for Tracing Plane and Space Algebraic Curves. In *11th European Workshop on Computational Geometry CG'95*. RISC-Linz, 1995.

[Nam95b] Tran Quoc Nam. A Hybrid Symbolic-Numerical Method for Tracing Surface-to-Surface Intersections. In *Proceedings of ISSAC-95*. ACM, 1995. To appear.

[NGH88] L. Naniglia, A. Galligo, and J. Heintz. Some new effective bounds in computational geometry. In *Proceedings of AAECC-6*, LNCS 357, pages 131–152. Springer Verlag, 1988.

[Pin73] J. R. Pinkert. *Algebraic algorithms for computing the complex zeros of Gaussian polynomials*. PhD thesis, University of Wisconsin, 1973.

[Tri78] W. Trinks. Über B. Buchbergers Verfahren, System Algebraischer Gleichungen zu Lösen. *J. Number Th.* , 10:475–488, 1978.

[Usp48] J.V. Uspensky. *Theory of Equations*. McGraw-Hill Book Company, Inc., Cambridge, 1948.

[Win88] Franz Winkler. A p-adic approach to the computation of Gröbner bases. *Journal of Symbolic Computation*, 6:287–304, 1988.

[Win94] Franz Winkler. Introduction to computer algebra. Technical report, RISC-Linz, J. Kepler Universität, Linz, Austria, July 1994. Lecture Notes of WS 93/94 and SS 94.

[Win95] Franz Winkler. *Introduction to Computer Algebra – Polynomial Algorithms*. Springer-Verlag, 1995. To appear in the series "Texts and Monographs in Symbolic Computation".

GATHER-AND-SIFT: A SYMBOLIC METHOD FOR SOLVING POLYNOMIAL SYSTEMS*

Lu Yang and Xiao-Rong Hou

Laboratory for Automated Reasoning & Programming
Chengdu Institute of Computer Applications, Academia Sinica
610041 Chengdu, People's Republic of China

e-mail: yanglu@bepc2.ihep.ac.cn

This paper gives a distinctive method to find the symbolic solutions of the parametric polynomial systems. "Gather" means to construct some ascending chains whose zeros contain all the zeros of the original systems. "Sift" means to remove extra zeros from the ascending chains such that only the required ones remain. Relevant programs written in MAPLE are implemented efficiently and some benchmark test examples are given.

1 Introduction

Solving a parametric polynomial system $PS := \{p_1, p_2, \cdots, p_n\}$ means finding all solutions of the system of algebraic equations:

$$\{p_j(u_1, u_2, \cdots, u_d, x_1, x_2, \cdots, x_n) = 0, \quad j = 1, \cdots, n\}, \tag{1}$$

where $u_1, \cdots, u_d$ are regarded as parameters; $x_1, \cdots, x_n$ as elements (indeterminates).

Definition 1.1. An n-tuple $(x_1(u_1, \cdots, u_d), \cdots, x_n(u_1, \cdots, u_d))$ where $x_j(u_1, \cdots, u_d)$ is an algebraic function of parameters $u_1, \cdots, u_d$, for $j = 1, \cdots, n$, is called a *zero* of parametric system $\{p_1, p_2, \cdots, p_n\}$ *in symbolic sense* if it satisfies (1), i.e.

$$\{p_j(u_1, \cdots, u_d, x_1(u_1 \cdots, u_d), \cdots, x_n(u_1, \cdots, u_d)) = 0, \quad j = 1, \cdots, n\}.$$

While a zero of a parametric system is mentioned hereafter, we always mean that in symbolic sense, throughout this paper.

*The work reported here was supported in part by National "Scaling Plan" of China and also supported in part by Academia Sinica Special Funds.

Sometimes, a system of algebraic equations is represented in a "triangular form" as follows:

$$\begin{cases} f_1(u_1, u_2, \cdots, u_d, x_1) = 0, \\ f_2(u_1, u_2, \cdots, u_d, x_1, x_2) = 0, \\ \cdots\cdots \\ f_s(u_1, u_2, \cdots, u_d, x_1, x_2, \cdots, x_s) = 0, \end{cases} \tag{2}$$

where $u_1, \cdots, u_d$ are regarded as parameters; x_j is called the *principal element of* f_j and the coefficient of the highest term of f_j with respect to it's principal element x_j the *initial* of f_j, for $j = 1, 2, \cdots, s$. And the set of polynomials $\{f_1, f_2, \cdots, f_s\}$ is also called a *triangular form.*

By resultant(p, q, x) denote the Sylvester resultant[2, pp.51-52] of any two polynomials, p and q, with respect to indeterminate x, which is a determinant in terms of the coefficients of $p(x)$ and $q(x)$ to verify if the two polynomials have a common solution. Given a triangular form $\{f_1, \cdots, f_s\}$ and another polynomial G, putting

$$\begin{aligned} r_{s-1} &:= \text{resultant}(G, f_s, x_s), \\ r_{s-2} &:= \text{resultant}(r_{s-1}, f_{s-1}, x_{s-1}), \\ &\cdots\cdots \\ r_0 &:= \text{resultant}(r_1, f_1, x_1), \end{aligned}$$

we call r_0 *the resultant of triangular form* $\{f_1, \cdots, f_s\}$ *with respect to polynomial* G and denote it simply by

$$\text{res}(f_1, \cdots, f_s, G).$$

A triangular form $\{f_1, \cdots, f_s\}$ is called a *normal ascending chain* if $I_1 \neq 0$ and for every $j = 2, 3, \cdots, s$,

$$\text{res}(f_1, \cdots, f_{j-1}, I_j) \neq 0, \tag{3}$$

where I_j denote the initial of f_j, i.e. the leading coefficient of f_j in x_j, for $j = 1, \cdots, s$.

While an ascending chain is mentioned hereafter, we always mean a normal one, throughout this paper.

An ascending chain $CS := \{f_1, f_2, \cdots, f_s\}$ is called a *characteristic set* of the system $PS := \{p_1, p_2, \cdots, p_n\}$ if
(i) every zero of PS must be a zero of CS;
(ii) $prem(f_1, \cdots, f_s, p_i) = 0, \quad$ for $i = 1, 2, \cdots, n$,
where $prem(f_1, \cdots, f_s, p_i)$ is the well-known *pseudo remainder of* CS *with respect to* p_i, obtained by doing Wu's[5] successive pseudo division.

It was inferred from (ii) that the zeros of CS are generically zeros of PS, too. So, it is often for people to solve the characteristic set CS instead of PS, the original system, i.e. to solve (2) instead of (1).

Given a set of constant values of the parameters $u_1, \cdots, u_d$, theoretically speaking, one can solve (2) by solving a series of equations with one unknown each. So, from the viewpoint of symbolic manipulation, if a characteristic set CS of a parametric polynomial system PS is found, then, we hold that PS is generically solved, and regard CS as the symbolic solution of PS.

In this paper, the authors suggest an mixed symbolic method for polynomial system solving. Given a system PS, establish at first some ascending chains $ASC_1, \cdots, ASC_k$ by an efficient method such that the union of the zeros of $ASC_1, \cdots, ASC_k$ contains all the zeros of PS, and then, remove the extra zeros from $ASC_1, \cdots, ASC_k$ using a "sieve" algorithm such that only the zeros of PS remain.

We call the method "Gather-and-Sift".

2 A sketch of "gather-and-sift"

Definition 2.1. A polynomial system $APS := \{q_1, \cdots, q_m\}$ where $q_j = q_j(u_1, \cdots, u_d, x_1, \cdots, x_n)$ for $j = 1, \cdots, m$ is called an *auxiliary system* of PS (and every q_j is called an *auxiliary polynomial*) provided that

(i) every zero of PS must also be a zero of APS;

(ii) counting the number of all the power products of all indeterminates except one, say, all the power products of $x_n, x_{n-1}, \cdots, x_2$ (including $x_n^0 x_{n-1}^0 \cdots x_2^0 = 1$) appeared in the expansions of $q_1, \cdots, q_m$, denoting the number by l, it holds that $l \leq m$.

Given a polynomial system $PS := \{p_1, \cdots, p_n\}$ as above, we suggest the following steps to find it's symbolic solutions.

Step 1. Construct an auxiliary system APS for PS.

Step 2. Sort every auxiliary polynomial q_j into a decreasing lexicographical order or a decreasing degree order.

Step 3. The auxiliary system can be regarded as a linear system:

$$\{q_j = c_{j1}\mu_1 + c_{j2}\mu_2 + \cdots + c_{jl}\mu_l, \quad j = 1, \cdots, m\} \tag{4}$$

where $\mu_1, \cdots, \mu_l$ represent the power products of $x_n, x_{n-1}, \cdots, x_2$ appeared in the expansions of $q_1, \cdots, q_m$, sorted into a decreasing order, while the "coefficients" c_{ji} are polynomials in $x_1, u_1, \cdots, u_d$.

Do a fraction-free Gaussian elimination for system (4), which results in a system:

$$\{h_1 = b_{11}\mu_1 + b_{12}\mu_2 + \cdots + b_{1l}\mu_l,\ h_2 = b_{22}\mu_2 + \cdots + b_{2l}\mu_l, \cdots, h_l = b_{ll}\mu_l\} \tag{5}$$

where b_{ji} are polynomials in $x_1, u_1, \cdots, u_d$, too.

Conversely, of course, we can write (5) as a system in $x_1, x_2, \cdots, x_n$ with parameters $u_1, \cdots, u_d$:

$$\{h_1(x_1, x_2, \cdots, x_n) = 0, \cdots, h_l(x_1, x_2, \cdots, x_n) = 0\}. \tag{6}$$

We denote the system (6) by GPS.

Step 4. Select a subsequence from the sequence $\{h_l, h_{l-1}, \cdots, h_1\}$, which is a triangular form in $\{x_1, \cdots, x_n\}$, and from that establish an ascending chain ASC (or ascending chains $ASC_1, \cdots, ASC_k$).

Since every previous step has lost no zeros, the set of zeros of ASC (or the union of the sets of zeros of $ASC_1, \cdots, ASC_k$) must contain all the zeros of PS.

For this step, however, no program has been given universally, because system APS is constructed with a great arbitrarility so that the coefficient matrix of (4) may be have lower rank, and some polynoimals of GPS may even be identically 0.

So, the method introduced here is a heuristic one, it is not a complete algorithm in the strict sense. As we observed in practice, however, it is still quite efficient in many cases, especially, for some benchmark problems.

Step 5. Employ a decomposition algorithm without factorization[6][7][8] (established by Jing-Zhong Zhang and authors), which we will describe later as a "sieve" algorithm, to remove the extra zeros from ASC (or $ASC_1, \cdots, ASC_k$) such that only the zeros of PS, i.e. the symbolic solutions remain.

So, we summarize the method briefly as follows:

- Constructing an auxiliary system.
- Sorting every auxiliary polynomial into a decreasing order.
- Doing Gaussian elimination.
- Establishing ascending chains from the system resulting from last step.
- Sifting out the extra zeros.

3 Constructing ascending chains from a Dixon set

As mentioned in last section, one can construct the auxiliary system in a variety of ways. For example, we can use dialytic method to create the needed auxiliary polynomials. The dialytic method was popular with scientists and engineers who want to solve polynomial systems. That was developed as a systematic algorithm by F.S. Macaulay[2, pp.55-58]. The computations based on dialytic method, however, is usually with a very high complexity and infeasible especially for parametric systems.

We would like to construct the auxiliary system in a way due to Dixon[1].

Given a system consisting of n polynomials in $n-1$ indeterminates,

$$\{p_1(x_1, \cdots, x_{n-1}), \cdots, p_n(x_1, \cdots, x_{n-1})\} \tag{7}$$

where the coefficients of $p_1, \cdots, p_n$ may depend upon some parameters which we have not written down explicitly, we define an $n \times n$ determinant as follows:

$$\Delta(x_1, \cdots, x_{n-1}, \alpha_1, \cdots, \alpha_{n-1}) := \begin{vmatrix} p_1(x_1, x_2, \cdots, x_{n-1}) & \cdots & p_n(x_1, x_2, \cdots, x_{n-1}) \\ p_1(\alpha_1, x_2, \cdots, x_{n-1}) & \cdots & p_n(\alpha_1, x_2, \cdots, x_{n-1}) \\ p_1(\alpha_1, \alpha_2, \cdots, x_{n-1}) & \cdots & p_n(\alpha_1, \alpha_2, \cdots, x_{n-1}) \\ \cdots & \cdots & \cdots \\ p_1(\alpha_1, \alpha_2, \cdots, \alpha_{n-1}) & \cdots & p_n(\alpha_1, \alpha_2, \cdots, \alpha_{n-1}) \end{vmatrix}, \tag{8}$$

where $\alpha_1, \cdots, \alpha_{n-1}$ are new variables and $p_i(\alpha_1, \cdots, \alpha_k, x_{k+1}, \cdots, x_{n-1})$ stands for uniformly replacing x_j by α_j for $1 \le j \le k$ in p_i.

Each of $x_i = \alpha_i$, for all $1 \le i \le n-1$, is a zero of Δ, so they can be removed by dividing Δ by $\prod_{i=1}^{n-1}(x_i - \alpha_i)$. Let

$$\delta(x_1, \cdots, x_{n-1}, \alpha_1, \cdots, \alpha_{n-1}) := \frac{\Delta(x_1, \cdots, x_{n-1}, \alpha_1, \cdots, \alpha_{n-1})}{(x_1 - \alpha_1) \cdots (x_{n-1} - \alpha_{n-1})}. \tag{9}$$

The polynomial δ is known as the **Dixon polynomial**[4].

Any zero of system (7) (say $x_1 = c_1, \cdots, x_{n-1} = c_{n-1}$) makes the Dixon polynomial vanish, no matter what the values of $\alpha_1, \cdots, \alpha_{n-1}$, hence all the coefficients of the various power products of $\alpha_1, \cdots, \alpha_{n-1}$ in the Dixon polynomial vanish. Let DPS be the set of all the polynomials in $x_1, \cdots, x_{n-1}$ which are coefficients of the power products of $\alpha_1, \cdots, \alpha_{n-1}$ in δ.

Definition 3.1. The polynomial system DPS, constructed as above, is called the *Dixon set* of $\{p_1, \cdots, p_n\}$ with respect to $\{x_1, \cdots, x_{n-1}\}$.

Now, let us return to our Gather-and-Sift method described in last section. Given a system PS as in (1), we may take the Dixon set of PS with respect to $\{x_2, \cdots, x_n\}$ as our auxiliary system, that is, take DPS as APS. Subsequently, by doing Step 2 and Step 3, we obtain a system GPS as (6) which is somewhat in an "upside-down triangular form". We have edited a generic program for the steps 1-3 based on Dixon set, we call it **GPS algorithm** which was written in MAPLE and efficiently implemented on a 486/33 personal computer.

Let us illustrate the GPS algorithm by the following example.

Example 1. Equations connecting sides and bisectors of a triangle.

Let ABC be a triangle, a, b and c the lengths of the sides opposite to angles A, B and C, a_i, a_e the lengths of internal and external bisectors of angle A, and b_e the length of the external angle bisector of angle B. The objective is to express a, b and c in terms of a_i, a_e and b_e.

This problem can be reduced to solving the following polynoimal system (see [3][4]) which is obtained from the formulae for computing bisectors,

$$\begin{cases} p_1 = a_i^2(b+c)^2 - cb(c+b-a)(c+b+a) = 0, \\ p_2 = a_e^2(c-b)^2 - cb(a+b-c)(c-b+a) = 0, \\ p_3 = b_e^2(c-a)^2 - ac(a+b-c)(c+b-a) = 0, \end{cases} \tag{10}$$

where a, b, c are indeterminates and a_i, a_e, b_e parameters.

Taking the Dixon set of $\{p_1, p_2, p_3\}$ with respect to $\{b, c\}$ as the auxiliary system, Sorting it into a decreasing degree order and doing a fraction-free Gaussian elimination, our program (**GPS algorithm**) gives a system GPS which consists of 13 polynomials, namely, $\{h_1, \cdots, h_{13}\}$. And then, it is very easy to get an ascending chain from the last three, $\{h_{13}, h_{12}, h_{11}\}$:

$$\begin{cases} P_1 = P_1(a_i, a_e, b_e, a) = 0, \\ P_2 = Q_1(a_i, a_e, b_e, a)b + Q_2(a_i, a_e, b_e, a) = 0, \\ P_3 = R_1(a_i, a_e, b_e, a, b)c + R_2(a_i, a_e, b_e, a, b) = 0. \end{cases} \tag{11}$$

where R_1, R_2, Q_1, Q_2 and P_1 are polynomials with 162, 390, 298, 162 and 330 terms, respectively. Since P_1 is an irreducible polynomial, P_2 is a linear form in b and P_3 a linear form in c, the ascending chain $\{P_1, P_2, P_3\}$ is irreducible, hence $\{P_1, P_2, P_3\}$ is just the symbolic solution of system (10). The Step 5 is no longer needed in the case that the output of Step 4 is irreducible ascending chains. The program for Example 1 took 635 seconds CPU time on a PC 486/33 to create GPS.

Gao and Wang in [3] reported that they solved this problem by successively eliminating b and c using a combination of pseudo division and resultant computation with some human intervention. They took 19 hours on their implementation in Lisp on a Sun 4 work station to get P_1, a polynomial with degree 20 in a.

Attempts were made to compute the Gröbner basis and the Macaulay resultant of system (10) on a Sun Sparc 10 work station, but neither ever finished while it went on for more than a day. (see [4])

In general cases, of course, the ascending chains produced by Step 4 may be reducible and have some extra zeros which we have to sift them out employing an algorithm[6][7] we will expound in the next section.

4 A distinctive decomposition algorithm

Something from [6][7] need be reviewed at first.

Definition 4.1. Given an polynomial system PS and a polynomial g, assume PS is the union of some ascending chains $ASC_1, \cdots, ASC_k$, (i.e. the set of zeros of PS is the union of the sets of zeros of $ASC_1, \cdots, ASC_k$), and by $prem_i$ and res_i denote the pseudo remainder and the resultant of ASC_i with respect to g for $i = 1, \cdots, k$, respectively. We said that PS is *simplicial to support* g if all of the $prem_1, \cdots, prem_k$ are identically 0, and PS is *simplicial to exclude* g if none of the $res_1, \cdots, res_k$ is identically 0. In both cases, PS is said to be *simplicial with respect to* g.

An irreducible ascending chain is always simplicial with respect to any polynomial. In general case, however, it is often required to decompose a reducible ascending chain into components simplicial with respect to some polynomial. An efficient algorithm is given by J.Z. Zhang and authors[6][7] for a normal ascending chain ASC and a polynomial g to decompose ASC into *simplicial components*, i.e. the subvarieties each of them is an ascending chain and their union is ASC, such that every component is simplicial with respect to g. Using this algorithm, we need not factorize ASC beforehand. In fact, we need neither do factorization over an algebraic number field, nor do that in any sense. It was sometimes known as "WR decomposition algorithm" since Wu's division[5] and sub-resultant computaion are employed.

Now we introduce notations. Let D be a domain, Q the quotient field of D. Put

$$f(x) := a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \in D[x],$$

$$g(x) := b_m x^m + b_{m-1} x^{m-1} + \cdots + b_1 x + b_0 \in D[x],$$

and $m_i(g,f,x)$ denote the following $(m+n-2i)\times(m+n-i)$ matrix,

$$m_i(g,f,x) := \begin{bmatrix} a_n & a_{n-1} & \cdots & \cdots & a_0 & & & \\ & a_n & a_{n-1} & \cdots & \cdots & a_0 & & \\ & & & \cdots & \cdots & & & \\ & & & a_n & a_{n-1} & \cdots & \cdots & a_0 \\ b_m & b_{m-1} & \cdots & \cdots & b_0 & & & \\ & b_m & b_{m-1} & \cdots & \cdots & b_0 & & \\ & & & \cdots & \cdots & & & \\ & & & b_m & b_{m-1} & \cdots & \cdots & b_0 \end{bmatrix}$$

where the first $m-i$ rows in terms of the coefficients of $f(x)$ and the rest $n-i$ rows in terms of those of $g(x)$, for $i=0,1,\cdots,\min(m,n)-1$. Then, by $M_j^i(g,f,x)$ denote the determinant of the sub-matrix which consists of the first $m+n-2i-1$ columns and the $(m+n-i-j)$-th column of $m_i(g,f,x)$ for $j=0,1,\cdots,i$.

The *sub-resultant polynomial series* of $f(x),g(x)$ means a polynomial series $\{P_i(g,f,x)\}_{i=0,1,\cdots,\min(m,n)-1}$ where $P_i(g,f,x) := \sum_{j=0}^{i} M_j^i(g,f,x)x^j$ for $i=0,1,\cdots,\min(m,n)-1$. For simplicity, put $s_i(g,f,x) := M_i^i(g,f,x)$ and call it the i-th *principal sub-resultant* of $f(x),g(x)$. We have:

Theorem 4.1. Defined notations as above, If at least one of a_n, b_m is non-vanishing and it holds that

$$s_0(g,f,x) = \cdots = s_{k-1}(g,f,x) = 0, \quad s_k(g,f,x) \neq 0,$$

then $P_k(g,f,x)$ is the greatest common divisor of $f(x),g(x)$ in $Q[x]$. The theorem is not new, for example, a proof can be found in [8].

Let $\{f_1(\vec{u},x_1),\cdots,f_s(\vec{u},x_1,\ldots,x_j)\}$ be a normal ascending chain consisting of polynomials over a domain D and $f(x) := a_ny^n + a_{n-1}y^{n-1} + \cdots + a_1y + a_0$, $g(x) := b_my^m + b_{m-1}y^{m-1} + \cdots + b_1y + b_0$ two polynomials over $D(\vec{u})[x_1,\cdots,x_j]$. And put $R_i := \mathrm{prem}(f_1,\cdots,f_j,s_i(g,f,y))$ where $s_i(g,f,y)$, the i-th principal sub-resultant of $f(y),g(y)$, is a polynomial in $D(\vec{u})[x_1,\cdots,x_j]$. Our algorithm employs the following

Theorem 4.2. If at least one of $\mathrm{res}(f_1,\cdots,f_j,a_n)$ and $\mathrm{res}(f_1,\cdots,f_j,b_m)$ doesn't equal 0 and

$$R_0 = \cdots = R_{k-1} = 0, \quad \mathrm{res}(f_1,\cdots,f_j,R_k) \neq 0,$$

then, $P_k(g,f,y)$ is the greatest common divisor of $f(y),g(y)$ over the quotient ring

$$D(\vec{u})[x_1,\cdots,x_j]/(f_1,\cdots,f_j).$$

Proof: Let $\{f_1^*,\cdots,f_j^*\}$ be an irreducible component of $\{f_1,\cdots,f_j\}$. Equalities $R_i = 0$ for $i=0,\cdots,k-1$ mean that

$$s_i(g,f,y) = 0 \mod(f_1,\cdots,f_j)$$

for $i = 0, \cdots, k-1$, hence

$$s_i(g, f, y) = 0 \mod(f_1^*, \cdots, f_j^*)$$

for $i = 0, \cdots, k-1$. On the other hand, $\text{res}(f_1, \cdots, f_j, R_k) \neq 0$ means

$$s_k(g, f, y) \neq 0 \mod(f_1^*, \cdots, f_j^*).$$

By assumption, at least one of $\text{res}(f_1, \cdots, f_j, a_n)$ and $\text{res}(f_1, \cdots, f_j, b_m)$ doesn't equal 0, so that at least one of a_n and b_m doesn't equal 0 in the field

$$D(\vec{u})[x_1, \cdots, x_j]/(f_1^*, \cdots, f_j^*).$$

Applying Theorem 4.1, we know that then, $P_k(g, f, y)$ is the greatest common divisor of $f(y), g(y)$ over the quotient ring

$$D(\vec{u})[x_1, \cdots, x_j]/(f_1^*, \cdots, f_j^*).$$

Since the choice for $\{f_1^*, \cdots, f_j^*\}$ is arbitrarily, it is also the g.c.d. of $f(y), g(y)$ over

$$D(\vec{u})[x_1, \cdots, x_j]/(f_1, \cdots, f_j).$$

Now, let us describe the steps of our WR algorithm in detail as follows.

For an ascending chain of single polynomial, $\{f_1\}$, it is simplicial with respect to g if either $\text{prem}(g, f_1, x_1) = 0$ or $\text{res}(g, f_1, x_1) \neq 0$. Otherwise, $\{f_1\}$ can be decomposed into $\{f_1'\}$ and $\{f_2'\}$ where f_1' is the greatest common divisor of f_1 and g; and f_2' is the pseudo quotient of f_1 by f_1'. In general, we have

WR Step 1. Compute the pseudo remainder of ASC with respect to g, namely, $\text{prem}(f_1, \cdots, f_s, g)$. If it equals 0, stop. Otherwise, let either $g^* := g$ or $g^* := \text{prem}(f_1, \cdots, f_s, g)$. If $\text{res}(f_1, \cdots, f_s, g^*) \neq 0$, stop.

WR Step 2. Assume $\text{res}(f_1, \cdots, f_s, g^*) = 0$. Let k be the smallest integer such that $\text{prem}(f_1, \cdots, f_{s-1}, s_k(g^*, f_s, x_s)) \neq 0$. If $\text{res}(f_1, \cdots, f_{s-1}, s_k(g^*, f_s, x_s)) \neq 0$, then by Theorem 4.2, over the quotient ring, $D(u_1, \cdots, u_d)[x_1, \cdots, x_{s-1}]/(f_1, \cdots, f_{s-1})$, we can find the g.c.d. of f_s and g^*, namely, $P_k(g^*, f_s, x_s)$, which we denote by f_s'. Let f_s'' be the pseudo quotient of f_s by f_s'. Then, return two ascending chains $\{f_1, \cdots, f_{s-1}, f_s'\}$, $\{f_1, \cdots, f_{s-1}, f_s''\}$ to be decomposed.

WR Step 3. Let k be the smallest number such that $\text{prem}(f_1, \cdots, f_{s-1}, s_k(g^*, f_s, x_s)) \neq 0$. If $\text{res}(f_1, \cdots, f_{s-1}, s_k(g^*, f_s, x_s)) = 0$, replace ASC and g by $\{f_1, \cdots, f_{s-1}\}$ and $s_k(g^*, f_s, x_s)$, i.e. to decompose $\{f_1, \cdots, f_{s-1}\}$ with respect to $s_k(g^*, f_s, x_s)$. Then, if $\{f_1, \cdots, f_{s-1}\}$ is well decomposed with respect to $s_k(g^*, f_s, x_s)$ and $\{\bar{f}_1, \cdots, \bar{f}_{s-1}\}$ is a component, we return the ascending chain $\{\bar{f}_1, \cdots, \bar{f}_{s-1}, f_s\}$ to be decomposed.

Since every return of the steps reduces the degrees of a polynomial in the ascending chain we deal with unless it's simplicial already, the algorithm ends in finite returns and gives a decomposition consisting of ascending chains every of which is simplicial with respect to g.

A generic program for WR decomposition algorithm written in MAPLE was efficiently implemented on a 486/33 personal computer.

5 Sifting extra zeros out

Now, let us describe how to use WR decomposition to do the Step 5 of the gather-and-sift algorithm sketched in Section 2.

Given a polynomial system $PS := \{p_1, \cdots, p_n\}$ to be solved, provided the Step 1 to Step 4 were completed, i.e. we have got an ascending chain ASC (or ascending chains $ASC_1, \cdots, ASC_k$) whose zeros contain all the zeros of PS, By WR decomposition we firstly find the components of ASC (or of all the $ASC_1, \cdots, ASC_k$) which are simplicial to support p_1, say, $ASC_1^{(1)}, \cdots, ASC_{i_1}^{(1)}$. Subsequently, we find (by WR too) the components of all the $ASC_1^{(1)}, \cdots, ASC_{i_1}^{(1)}$ which are simplicial to support p_2, say, $ASC_1^{(2)}, \cdots, ASC_{i_2}^{(2)}$. Doing like this successively, at last we find ascending chains $ASC_1^{(n)}, \cdots, ASC_{i_n}^{(n)}$ which are simplicial to support $p_1, \cdots, p_n$ so that $\mathcal{S} := \{ASC_1^{(n)}, \cdots, ASC_{i_n}^{(n)}\}$ is the symbolic solution of PS.

Example 2. (A neural network problem).

Find the symbolic solution of the following polynomial system:

$$\begin{cases} 1 - cx + xy^2 + xz^2 = 0 \\ 1 - cy + yx^2 + yz^2 = 0 \\ 1 - cz + zx^2 + zy^2 = 0 \end{cases} \tag{12}$$

where x, y, z are indeterminates and c is a parameter.

Employing **GPS algorithm** which implements steps 1-3 of the "gather-and-sift", we obtain a system $GPS := \{h_1, \cdots, h_7\}$ somewhat in an "upside-down triangular form", where $h_1 = x(1 - cz + zx^2 + zy^2)$, $h_4 = x^3(x - y)(x^2y + xy^2 + 1)$, $h_7 = x^2(-2x^3 + cx - 1)(2cx^4 - 2x^3 - c^2x^2 - 2cx - 1)(x^3 - cx - 1)^2(2x^4 - 3cx^2 + x + c^2)^2$.

It is easy to see that $x = 0$ cannot satisfy the first equation of (12), so that we can remove the factor x from h_7, h_4, h_1 and simplify them into an ascending chain $ASC := \{P_1, P_2, P_3\}$ that

$$P_1 = (-2x^3 + cx - 1)(2cx^4 - 2x^3 - c^2x^2 - 2cx - 1)(x^3 - cx - 1)(2x^4 - 3cx^2 + x + c^2),$$
$$P_2 = (x - y)(x^2y + xy^2 + 1),$$
$$P_3 = 1 - cz + zx^2 + zy^2.$$

The zeros of ASC contain all the zeros of system (12), and what we need do is to sift the extra zeros out.

Since P_1 and P_2 are explicitly the products of 4 and 2 factors, respectively, ASC can be easily decomposed into 8 components $ASC_1, \cdots, ASC_8$. Applying WR decomposition program to each of them, at last we receive 5 ascending chains, $CS_1, \cdots, CS_5$, each of which is simplicial to support every polynomial in (12), i.e. $\mathcal{S} := \{CS_1, \cdots, CS_5\}$ is the symbolic solution of (12). One can infer very easily from the following expressions that the number of the zeros of (12) does not exceed 21, provided the parameter c takes a constant value.

$$CS_1 = \{-2x^3 + cx - 1, x - y, 1 - cz + zx^2 + zy^2\},$$

$$
\begin{aligned}
CS_2 &= \{2x^4 - 3cx^2 + x + c^2, x - y, 1 - cz + zx^2 + zy^2\},\\
CS_3 &= \{x^3 - cx - 1, x^2y + xy^2 + 1, 1 - cz + zx^2 + zy^2\},\\
CS_4 &= \{2x^4 - 3cx^2 + x + c^2,\\
&\quad y(2c^4x^3 - 156cx^3 + 76c^3x^2 + 36x^2 - 6c^2x - 2c^5x - 42c^4)\\
&\quad + 8c^3x^3 - 36x^3 + c^5x^2 + 84c^2x^2 - 7c^4x - 42cx - c^6 - 42c^3,\\
&\quad 1 - cz + zx^2 + zy^2\},\\
CS_5 &= \{2cx^4 - 2x^3 - c^2x^2 - 2cx - 1,\\
&\quad y(-2c^{12}x^3 + 92c^9x^3 - 416c^6x^3 - 576c^3x^3 - 64x^3 + 100c^8x^2 - \cdots)\\
&\quad + 2c^{11}x^3 + 32c^8x^3 - 384c^5x^3 - 64c^2x^3 - c^{13}x^2 + 48c^{10}x^2 - \cdots,\\
&\quad 1 - cz + zx^2 + zy^2\}.
\end{aligned}
$$

References

[1] Dixon, A.L., 'The eliminant of three quantics in two independent variables', *Proc. London Math. Soc.*, **6**, 468-478, (1908).

[2] Donald,B.R., Kapur.D & Mundy, J.L.(eds), *Symbolic and Numerical Computation for Artificial Intelligence*, Academic Press, 1992.

[3] Gao, X.S. & Wang, D.K., 'On the automatic derivation of a set of geometric formulae', *Journal of Geometry*, to appear.

[4] Kapur, D., Saxena, T. & Yang L., 'Algebraic and geometric reasoning using Dixon resultants', Proceedings of ISSAC'94, ACM Press, 1994. pp.99-107.

[5] Wu, W.T., 'On the decision problem and the mechanization of theorem proving in elementary geometry', *Scientia Sinica* **21**, 157-179, (1978).

[6] Yang, L., Zhang, J.Z. & Hou, X.R., 'A criterion of dependency between algebraic equations and its applications', Proceedings of the 1992 International Workshop on Mathematics Mechanization, Wu & Cheng (eds), International Academic Publishers, (Beijing) 1992. pp.110-134.

[7] Yang, L., Zhang, J.Z. & Hou, X.R., 'An efficient decomposition algorithm for geometry theorem proving without factorization', in MM-Preprints, No.**9**, 1993. pp.115-131.

[8] Zhang, J.Z., Yang, L. & Hou, X.R., 'The sub-resultant method for automated theorem proving', *J. Sys.Sci. and Math.Sci.* (Chinese ed.), **15**:1, 10-15, (1995).

EXPERIMENTING INTEGRATION QUADRATURE WITH SOFTWARE

WEI-CHI YANG
DEPARTMENT OF MATHEMATICS AND STATISTICS
RADFORD UNIVERSITY, RADFORD, VA 24142, U.S.A.
E-MAIL: WYANG@RUNET.EDU
MYUNGSUK CHUNG
DEPARTMENT OF MATHEMATICS
VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
BLACKSBURG, VA 24061, U.S.A.
E-MAIL: MYUNGSUK@MATH.VT.EDU

1. INTRODUCTION

In this paper, we shall describe how we may incorporate the existing computer algebra systems (CAS) into experimenting integration quadratures. With the help of the math software, users can make conjectures easier than not using the software. Consequently, users will discover innovative methods of solving problems. The main section of this paper is in section 2. We discuss two closed type adaptive quadratures in one dimension, tail correction for the closed type. Next we talk about the open type adaptive quadrature in two dimensions. Finally, we discuss a closed type adaptive quadrature in two dimensions.

2. ADAPTIVE QUADRATURE

2.1. Closed type in one dimension. First we experiment a closed type quadrature in one dimension, which can be used in estimating the integral of a monotone function with one singularity at one end point. A closed type quadrature is to ignore the singularity, see [DR]. We shall see that an adaptive quadrature in treating this type of function is more efficient than quadratures which use uniform spaced intervals. We therefore consider the following definition which enables us to divide an interval unevenly.

Definition 1. *A matrix A with positive a_{nk} is called uniformly regular if the following conditions are satisfied:*

(i) $\lim_{n\to\infty} a_{nk} = 0$ uniformly over k.
(ii) $\sum_{k=1}^{n} a_{nk} = 1$.

For example, we may use the finite sum formula, $\sum_{k=1}^{n} k^m$, $m = 1, 2, ...$, to form uniform regular matrices. For $m = 1$, we define the matrix $a_{nk} = \frac{2k}{n(n+1)}$. We remark that in Scientific Workplace, we may use

$$a(n,k) = \frac{2k}{n(n+1)}$$

In Maple, we may use

$$ank = (n,k)- > 2 * k/(n * (n+1));.$$

In Mathematica, we use

$$ank[n\_, k\_] := 2 * k/\{n * (n+1)\}.$$

(Which one is more natural to the users?) To illustrate what a uniform regular matrix would look like. We use Scientific Workplace to show the matrix determined by a_{nk} when $n = 10$, but first we modify $a(n,k)$ as follows:

$$a(n,k) = \begin{cases} \frac{2k}{n(n+1)} & \text{if } \ k \le n \\ 0 & \text{if } \ k > n \end{cases}.$$

We obtain

$$\begin{array}{cccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{6} & \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{10} & \frac{1}{5} & \frac{3}{10} & \frac{2}{5} & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{15} & \frac{2}{15} & \frac{1}{5} & \frac{4}{15} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{21} & \frac{2}{21} & \frac{1}{7} & \frac{4}{21} & \frac{5}{21} & \frac{2}{7} & 0 & 0 & 0 & 0 \\
\frac{1}{28} & \frac{1}{14} & \frac{3}{28} & \frac{1}{7} & \frac{5}{28} & \frac{3}{14} & \frac{1}{4} & 0 & 0 & 0 \\
\frac{1}{36} & \frac{1}{18} & \frac{1}{12} & \frac{1}{9} & \frac{5}{36} & \frac{1}{6} & \frac{7}{36} & \frac{2}{9} & 0 & 0 \\
\frac{1}{45} & \frac{2}{45} & \frac{1}{15} & \frac{4}{45} & \frac{1}{9} & \frac{2}{15} & \frac{7}{45} & \frac{8}{45} & \frac{1}{5} & 0 \\
\frac{1}{55} & \frac{2}{55} & \frac{3}{55} & \frac{4}{55} & \frac{1}{11} & \frac{6}{55} & \frac{7}{55} & \frac{8}{55} & \frac{9}{55} & \frac{2}{11}
\end{array}$$

Consider the following closed type quadrature:

$$Q_n^1(f) = \frac{1}{2} a_{n1} f(u_{n1}) + \sum_{k=2}^{n} \frac{a_{nk}}{2} \left(f(u_{n,k-1}) + f(u_{nk}) \right).$$

We would like to experiment this quadrature with the Scientific Workplace (which uses Maple as a tool for computation). But first we need to make the following adjustments for computation purpose. We define the right and left endpoints as follows:

$$r(n,k) = \sum_{j=1}^{k} a(n,j),$$

and

$$l(n,k) = \sum_{j=0}^{k-1} a(n,j),$$

which correspond to $u_{n,k}$ and $u_{n,k-1}$ respectively.

We define our first closed type quadrature as follows:

$$Q^1(n) = (1/2)a(n,1)f(r(n,1)) + \sum_{k=2}^{n} \frac{a(n,k)}{2}(f(l(n,k)) + f(r(n,k))).$$

We note that the first term of $Q^1(n)$, $(1/2)a(n,1)f(r(n,1))$, is a tail term to take care of functions with a singularity, and the second term of $Q^1(n)$ is a trapezoidal sum. Thus, we may call the quadrature, $Q^1(n)$, to be the **adaptive trapezoidal sum.**

Example 2. *Consider the function $f(x) = \ln(1-\cos x)$, if $x \neq 0$, and $f(0) = 0$. (We notice that f has a singularity at $x = 0$.) Use $Q^1(n)$ to approximate $\int_0^1 \ln(1-\cos x)dx$.*

If we use **Evaluate numerically** with Scientific Workplace under "Maple", we get the following numeric results:

$$\begin{aligned} Q^1(300) &= -2.720856531 \\ Q^1(400) &= -2.720938148 \\ Q^1(430) &= -2.720950937 \end{aligned}$$

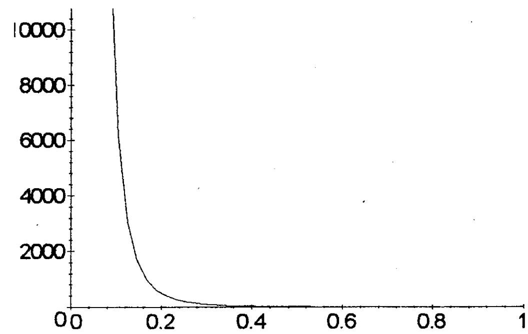
However, when we increase n, we will be warned of the existence of the singularity at $x = 0$. To further investigate the convergence or divergence of this integral, we could write a separate program to run our quadrature.

The next example shows that the tail term, $(1/2)a(n,1)f(r(n,1))$, could speed up the divergence too.

Example 3. *Let $f(x) = \frac{(\ln(1+x))^{1/7}}{x^4}$ if $x \neq 0$, and $f(0) = 0$. We would like to use $Q_1(n)$ to approximate $\int_0^1 f(x)dx$. The table below is a comparison between the quadrature with and without a tail.*

	without the tail $\frac{1}{2}a_{n1}f(u_{n1})$	with the tail $\frac{1}{2}a_{n1}f(u_{n1})$
$n = 100$	$0.3963290351 \times 10^{11}$	$0.1265245186 \times 10^{17}$
$n = 150$	$0.3982843325 \times 10^{12}$	$0.4030436606 \times 10^{18}$
$n = 200$	$0.2051511485 \times 10^{13}$	$0.4711286825 \times 10^{19}$

We conjecture from the table above that $\int_0^1 \frac{(\ln(1+x))^{1/7}}{x^4} dx = \infty$. We include the graph of f as follows:



We now describe the error bound of this quadrature. The following two theorems can be found in [Y].

Theorem 4. *Let f be improper Riemann integrable on $(0,1]$, f'' exist over $(0,1)$ and a_{nk} be a uniformly regular matrix. If the quadrature $Q_n^1(f)$ is used on f over the interval $[0,1]$. Then we have*

$$\left|E_n^1(f)\right| = \left|\int_0^1 f - Q_n^1(f)\right| \le \left|\int_0^{u_{n1}} f - u_{n1}f(u_{n1})\right| + \left|\frac{1}{2}a_{n1}f(u_{n1})\right| + \left|\sum_{k=2}^{n} -\frac{f''(c_{nk})a_{nk}^3}{12}\right|.$$

In view of Example 3 and Theorem 4, for a monotone function with singularity at one end point, to speed up the rate of convergence or divergence, we add an additional tail term, $\frac{1}{2}a_{n1}f(u_{n1})$ in $Q_n^1(f)$, or $(1/2)a(n,1)f(r(n,1))$ in $Q_1(n)$.

2.2. Closed type tail correction. We may modify the tail term, $\frac{1}{2}a_{n1}f(u_{n1})$ in $Q_n^1(f)$, and, thus, introduce a second quadrature. In essence, it is to replace the tail, $\frac{1}{2}a_{n1}f(u_{n1})$, by a derivative term, $a_{n1}f(u_{n1}) - \frac{1}{2}a_{nk}^2 f'(u_{n1})$, which is obtained by considering the area of the trapezoid formed by tangent line at $x = u_{n1}$, x-axis and y-axis. Hence, if we set

$$Q_n^2(f) = a_{n1}f(u_{n1}) - \frac{1}{2}a_{nk}^2 f'(u_{n1}) + \sum_{k=2}^{n} \frac{a_{nk}}{2}\left(f(u_{n,k-1}) + f(u_{nk})\right).$$

We obtain the following information with the help of Maple V Release 3:

$$\begin{aligned} Q_{300}^2(f) &= -2.721138187 \\ Q_{400}^2(f) &= -2.721091402 \\ Q_{430}^2(f) &= -2.721081466. \end{aligned}$$

Again, we can't increase the number n, but we can write a separate program to run this quadrature and we conjecture that the integral is convergent. Similar to $Q_n^1(f)$, we have the following

Theorem 5. *Let f be improper Riemann integrable on $(0,1]$, f'' exist over $(0,1)$ and a_{nk} be a uniformly regular matrix. If the quadrature (8) is used on f over the interval $[0,1]$. Then we have*

$$\begin{aligned} \left|E_n^2(f)\right| &= \left|\int_0^1 f - Q_n^1(f)\right| \le \left|\int_0^{u_{n1}} f - u_{n1}f(u_{n1})\right| + \left|\frac{1}{2}a_{n1}f'(u_{n1})\right| + \\ &\quad \left|\sum_{k=2}^{n} -\frac{f''(c_{nk})a_{nk}^3}{12}\right|. \end{aligned}$$

Remark 1. *It is interesting to see that for $f(x) = 1/\sqrt{x}$, if $x \neq 0$, and $f(0) = 0$, if we use $a_{nk} = \frac{2k}{n(n+1)}$, then using $Q_n^2(f)$ is better than using $Q_n^1(f)$, meaning that $Q_n^2(f)$ uses less number of points for calculation and yet obtain desired accuracy. However, if we use $b_{nk} = \frac{6k^2}{n(n+1)(2n+1)}$, then $Q_n^1(f)$ is better than $Q_n^2(f)$. We illustrate these by observing the following data*

a_{nk}	b_{nk}
$Q_{80}^1 = 1.973737918$	$Q_{80}^1 = 1.999305311$
$Q_{80}^2 = 1.991306127$	$Q_{80}^2 = 2.001703454$
$Q_{90}^1 = 1.976647478$	$Q_{90}^1 = 1.999442063$
$Q_{90}^2 = 1.992274385$	$Q_{90}^2 = 2.001453908$

.

2.3. Open type in two dimensions. Before we extend the closed type quadrature to two dimensions, we consider an open type quadrature in two dimensions by using two uniformly regular matrices, c_{nk}, d_{ml}, and denote them by $c(n,k)$ and $d(m,l)$ for computation purpose. Now set $c(n,k) = \frac{2k}{n(n+1)}$, $d(m,l) = \frac{2l}{m(m+1)}$, and consider the function $g(x,y) = \frac{1}{\sqrt{xy}}$ if $x \neq 0$, and $y \neq 0$, and $g(x,y) = 0$ if $x = y = 0$. We define the following open quadrature which can be experimented directly inside Scientific Workplace:

$$\begin{aligned} Q(m,n) &= \sum_{l=2}^{m}(\sum_{k=2}^{n} \frac{c(n,k)d(m,l)}{4}(g(r(n,k),r(m,l)) + \\ &\quad g(l(n,k),r(m,l)) + g(r(m,l),l(n,k)) + \\ &\quad +g(l(n,k),l(m,l)))). \end{aligned}$$

We obtain the following information:

$$\begin{aligned} Q(10,10) &= 3.100296603 \\ Q(30,30) &= 3.681601032 \\ Q(40,40) &= 3.759392108 \end{aligned}$$

We note that the quadrature should converge to 4. However, if we compute $Q(50,50)$, we get $Q(50,50) = \text{bytes used} = 29295868$, alloc $= 7731832$, which indicates that the Scientific Workplace (using Maple) stops computation.

Remark 2. *We could use different uniformly regular matrices for experimenting, for example, we may replace* $a(n,k) = \frac{2k}{n(n+1)}$ *by* $b(n,k) = \frac{6k^2}{n(n+1)(2n+1)}$ *in* $Q_1(n)$.

Remark 3. *The open type quadrature is so called "avoiding the singularity" see [DR], and can be also used in estimating the integrals of functions which are highly oscillatory, such as* $f(x) = \frac{1}{x}\sin\frac{1}{x}$ *and* $f(x,y) = \frac{1}{xy}\sin\frac{1}{xy}$, *see [LY].*

2.4. Closed type in two dimensions. To speed up the rate of convergence for this type of function, naturally, we consider a closed type quadrature, which is an extension of $Q_n^1(f)$, as follows:

$$\begin{aligned} Q_n^3(f) &= \sum_{l=1}^{m}\sum_{k=1}^{n} \frac{a_{nk}b_{ml}}{4}(f(u_{n,k-1},v_{m,l-1}) + f(u_{nk},v_{m,l-1}) + \\ &\quad f(u_{n,k-1},v_{ml}) + f(u_{nk},v_{ml})) + \frac{a_{n1}b_{m1}}{4}f(u_{n1},v_{m1}) \end{aligned}$$

$$\sum_{k=2}^{n}\frac{a_{nk}b_{ml}}{4}(f(u_{n,k-1},v_{m1})+f(u_{nk},v_{m1}))+$$
$$\sum_{l=2}^{m}\frac{a_{nk}b_{ml}}{4}(f(u_{n,1},v_{m,l-1})+f(u_{n1},v_{ml}))$$

If we use $a_{nk}=\frac{6k^2}{n(n+1)(2n+1)}$ and $b_{ml}=\frac{6l^2}{m(m+1)(2m+1)}$,we obtain the following information from Maple V Release 3:

$$Q^3_{70}(f) = 3.999361277$$
$$Q^3_{80}(f) = 3.999619399$$
$$Q^3_{90}(f) = 3.999780084.$$

By comparing the open type and closed type quadratures, we see that closed type quadrature is more efficient in this case.

3. Conclusion

Computer algebra systems are great tools for teaching and research. Users can use them to explore mathematics, making conjectures, verifying conjectures, and consequently formulating exciting new theorems. However, there are reasons why some people are afraid of using them too. There are so many CAS available, learning a particular software may not be an easy task. Up to now, software such as Scientific Workplace can help users do wordprocessing and computation at the same time, which release lots of burden on the users, but it lacks the programming capability. This suggests that users sometimes need to exit one program and enter the other. Indeed, one area that most software developers are striving for nowadays is to combine their powerful numeric and symbolic software together with a wordprocessor. Hopefully, someday we have a product that can handle all kinds of demands.

References

[DR] Davis and Rabinowitz, method of Numerical Integration, 2nd ed., Academic Press 1983.

[LY] P.Y. Lee and W.-C. Yang, Henstock Integral and Numerical Integration, preprint.

[Y] W.-C. Yang, The Errors for the Closed and Open Type Adapted Quadratures, preprint.

Scientific WorkPlace

Roger Hunter, TCI Software Research

1 Introduction

Scientific WorkPlace is a tool for the creation and use of scientific documents. By a document we mean an organized collection of information designed to communicate ideas. Under this definition, a document can exist electronically, on paper, or in any other medium. Maple and Mathematica worksheets, spreadsheets, Web pages, the files that underlie multimedia CD-ROMs, as well as traditional paper documents, all fall under this broad definition of document.

Scientific WorkPlace is optimized for handling mathematical content. Entering and editing mathematical expressions is simple, direct, and uses standard, natural, notation. All expressions are live—you can compute directly with mathematics in standard form.

Scientific WorkPlace complements other mathematical products like Maple, Mathematica, and Matlab. The creators of those products focus primarily on algorithms and computations. Scientific WorkPlace focuses on interface and interaction, providing simple and direct access to the capabilities of those products. At the same time, documents can be typeset with Scientific WorkPlace, producing the highest possible quality in paper form, another aspect of interaction with documents.

The emphasis on interface and interaction makes Scientific WorkPlace much more than a front end to Maple or LaTeX; it is a front end to scientific content and communication.

The use of "WorkPlace" in the name indicates how we see the product being used. It is designed as the product of choice for anyone who deals with scientific documents, including students, teachers, researchers, and professionals. By emphasizing content and logical structure, Scientific WorkPlace increases productivity, creates archival documents, and improves communication.

2 The Mathematics Editing Interface

2.1 Natural Notation

Natural mathematical notation is the standard notation used in printed articles, in books, and on the blackboard. Two important features of this notation are:

1. Two dimensionality. An expression like $\int_0^1 \frac{2x^2+\sin x}{\ln x}dx$ uses symbol size and positioning to convey information.

2. Single letter variable names. This makes it possible to represent multiplication with juxtaposition, increasing the information density by reducing the number of symbols needed.

Both of these features increase the speed of comprehension over the forms required by many systems. Optical bandwidth is many orders of magnitude higher than aural bandwidth, and the linear forms provide something similar to aural bandwidth. With natural notation, you can see the basic structure of very complex expressions at a glance, while it may take several minutes or more in linear form.

Developers of mathematical software systems realize the importance of natural notation, and are moving toward it for displaying expressions, although most still use linear form of notation for entering and editing.

2.2 Free-Form Editing

Scientific WorkPlace is not the only computational product to use natural notation for entering and editing expressions. Mathcad, for example, maintains expressions in what is essentially natural notation, although variable names are not assumed to be single symbols.

The unique contribution of Scientific WorkPlace is complete adherence to natural notation, together with free-form editing. The primary alternative to free-form editing is the use of complete templates. Requiring complete templates for mathematical constructs is like requiring entry of whole words when entering or editing text. Although there are only 26 characters and some punctuation in the English language, there are over 500,000 words. Similarly, there are a small number of mathematical "characters" and a large number of mathematical structures. The difference in ease-of-use provided by free-form over templates is in a similar ratio.

For example, the template for an integral is $\int_{\square}^{\square} \square d\square$. The template is a single object with four separate fields. The only way to enter the template is to select it from a palette. Apart from the huge number of different templates necessary, this example illustrates another basic problem with templates—in natural notation, there is no clear indication of the end of the last field. The user must be aware of the boundary between the end of the last field and any following material, and must actively move out of the field. Typically, the field must be indicated on screen as a box, cluttering the screen and obscuring the notation.

In Scientific WorkPlace, the integral template is replaced by a sequence of "characters." The integral begins with the operator symbol $\int_{\square}^{\square}$ with two optional fields. The rest of the expression is a sequence of characters and symbols. You can insert characters at any position, and there is no need to distinguish the end of any field. The initial operator symbol with its two fields is the most complex part of the expression.

An expression consists of sequences of characters and symbols, together with what could be called character templates—fragments of expressions that represent the minimal building blocks of two-dimensional structures. The basic building blocks are shown in the following table.

Name	Picture
Operator	$\int_{\square}^{\square}$
Radical	$\sqrt[\square]{\square}$
Fraction	$\frac{\square}{\square}$
Script	${}^{\square}_{\square}$
Bracket	$(\square)$
Matrix	$\begin{matrix}\square & \square \\ \square & \square\end{matrix}$
Decoration	$\overrightarrow{\square}$
Labelled	$\overset{\square}{\square}$

Of course, there are trade-offs. With templates, you cannot enter a malformed expression; with free-form, you can. This trade-off is analogous to the fact that you are free to misspell words in a word processor—the system does not force you to be correct. The ease-of-use in a free-form system so far outweighs the problems of entering incorrect expressions that there is simply no contest.

Computer programmers and users of traditional computational systems assume that multi-character variable names are the norm, and are surprised at a system that assumes variable names are single characters. In fact, there is a tradition in mathematics of multi-character names for special functions. The primary examples are the transcendental functions, such as sin, cos, and ln.

Rather than assume that variables are normally multi-character, and requiring that the ends of names be deduced using spaces or symbols for operations, Scientific WorkPlace assumes that all variable names are single-character, and requires you to encapsulate the characters of a multicharacter name. This process is made much easier by automatic substitution. If you type the letters s, i, and n in mathematics, Scientific WorkPlace replaces them by the single multicharacter object, sin. You can create your own multi-character names and have them recognized automatically.

2.3 Use of Color and Fonts

Scientific WorkPlace uses color to distinguish logical elements of a document. Normal text uses a black upright font while mathematics uses a red italic font. This makes the distinction between text and mathematics very clear, and at the same time allows you to insert mathematics anywhere that you can insert plain text.

3 Performing Computations

3.1 Basic Computations

Basic computations are performed by placing the insertion point in an expression and choosing an operation from a menu. The results of the computation follow the expression. For example, if you place the insertion point anywhere in the expression $\int x^2 dx$ and choose

Evaluate from the Maple menu, the result is inserted immediately following the integral sign, preceded by an equals sign: $\int x^2 dx = \frac{1}{3}x^3$. The insertion point is placed at the end of the result.

Scientific WorkPlace makes no distinction between the result of a computation and any other part of your document. There is no concept of input and output cells in Scientific WorkPlace, and performing a computation has no effect on any other part of your document. This has several advantages:

1. You can use the result directly for further computation.
2. You can edit the result in any way you want.
3. There are no special areas or cells in the document that behave differently.
4. There are no side effects that ripple through the document.

3.2 In-Place Computations

If you select any part of an expression and choose an operation from the menu while holding down the ctrl key, the result of the computation replaces the selection. This in-place computation provides a very powerful way to manipulate expressions. For example, by selecting the numerator of

$$\frac{(x+1)^3}{(x-1)^2}$$

and choosing Expand, you can expand the numerator without changing the denominator:

$$\frac{x^3+3x^2+3x+1}{(x-1)^2}$$

3.3 Definitions

To assign the value 27 to the variable x, you place the insertion point in the expression $x = 27$ and choose New Definition from the menu. The same operation applied to the expression $f(x) = x^2+1$ defines the function f. Definitions can be saved with a document and restored when it is opened.

3.4 About the Model

Scientific WorkPlace's basic computational model is directed toward easy, direct access. In this regard, it is closest to Maple and Mathematica. Programming capabilities provide a way of automating the basic model, and are only loosely related to document structure. Definitions are temporal—the most recent definition of a variable is in effect throughout the document, and will be used for subsequent computations until it is changed.

The other prevalent model is represented by Mathcad and Theorist, where definitions are spatial—the current value of a variable depends on where you are in a document.

Programming is structural as it is in spreadsheets. In Mathcad, for example, a form of programming is achieved by linking the output of a computation in one region to the input of a computation in another.

4 Features of the Interface

4.1 Understanding Mathematics

Because free-form natural mathematics entry reduces the burden on the user significantly, there is a corresponding increase in the work required by the program to understand what the user enters. For example, the variable of integration in an expression like $\int x^2 dx$ is never distinguished other than by the preceding letter *d*, yet we understand the expression instantly. Our design goal was to use natural notation exclusively, never asking the user to supply more information. These are some of the issues that arose from this decision:

1. Many equivalent forms must be recognized—the expressions 3×2, $3 \cdot 2$, and $3 * 2$ must all be equivalent.

2. Inconsistent notation must be handled correctly—$\sin^2 x$ and $\sin^{-1} x$ are fundamentally different in meaning yet use the same notation.

3. Natural notation is often ambiguous, with ambiguities that must be resolved differently depending on the elements of an expression. For example, the expressions $\sin \pi/2$ and $\sin a/\cos b$ have the same form but should be interpreted as $\sin \frac{\pi}{2}$ and $\frac{\sin a}{\cos b}$ respectively.

These issues, and many more, are addressed during an internal translation from free-form to what amounts to mathematically complete template form. This is done every time an expression is used in a computation and it is the most complex part of Scientific Workplace's mathematics hàndling.

Our experience with this project bears out the statement that the more you make an interface simple and direct, the more complex it is to program. As time goes by, the percentage of computing resources devoted to interfacing will far outstrip the percentage devoted to the underlying computations. This trend is clearly evident in the rapidly increasing size of Windowing programs.

Choosing the part of an expression on which to perform a computation is another issue that arises from the decision to make mathematics a seamless part of a document. Scientific WorkPlace uses its knowledge of the operation you want to perform to determine how much of an expression it chooses for the operation. For example, suppose your insertion point is within the left-hand side of the equation $x^3 + 1 = 0$ and you choose Solve Exact. Because you chose Solve, Scientific WorkPlace selects the entire equation for the operation, returning Solution is : $\{x = -1\}, \{x = \frac{1}{2} + \frac{1}{2}i\sqrt{3}\}, \{x = \frac{1}{2} - \frac{1}{2}i\sqrt{3}\}$

On the other hand, if you choose Factor, Scientific WorkPlace selects just the left hand side of the expression and inserts the factorization within the expression, yielding $x^3 + 1 = (x + 1)(x^2 - x + 1) = 0$.

4.2 Incomplete Expressions

One of the more controversial features of Scientific WorkPlace is its ability to work with incomplete or malformed expressions. There were two motivations for this feature:

1. Deliberate shorthand should be accepted. For example, $\int x$ is interpreted as $\int x dx$.

2. When the intent is obvious, minor errors should be tolerated. For example, $\frac{d}{dx^2}x^3$ should be accepted to mean $\frac{d^2}{dx^2}x^3$.

We believe this makes Scientific WorkPlace more powerful and easier to use. We're all familiar with the frustration caused by systems that insist on very rigid input syntax and add insult to injury by beeping at the slightest transgression.

The main objection to this feature is that it may encourage students to use sloppy notation. At this time, we have no evidence supporting or refuting this objection.

Another objection is that the system may interpret an expression in a way that you didn't intend. This is a consequence of the general principle that the more powerful a tool, the more potential there is for misuse.

4.3 Plots

Plotting an expression is just like performing a computation. You select the expression and choose one of the plot commands from the menu. The expression is copied and stored with the plot record, so it is independent of the original expression in the document.

Plots are "live" and can be manipulated directly in place. You can use the mouse to pan or zoom a plot. To plot a second expression on an existing plot, you select the expression and drag it onto the plot.

5 Use for Teaching and Learning

At the time handheld calculators were first introduced, the computations they performed and the algorithms they employed were well known and available on the computers of the time. The real innovation was the method of delivery. Calculators rapidly replaced slide rules and their use led to significant changes in the actual material taught in schools.

The situation with symbolic computation is similar. At this time, the algorithms are well known and are embodied in products like Maple and Mathematica. However, classroom use is not yet widespread. There are several reasons for this, but the most important is that it takes at least the first month of each class to teach the syntax and operation of the systems before any actual mathematics can be introduced. Calculus students have to learn two languages and notations for the course instead of the one that was already a significant burden.

Because Scientific WorkPlace uses the natural notation of the textbooks and blackboard, and because it can be learned in a fraction of the time it takes to learn other systems, it can be used immediately. Like handheld calculators, Scientific WorkPlace constitutes a major new delivery system that makes the algorithms accessible to everyone.

The innovations in the Scientific WorkPlace interface make it the appropriate vehicle for the next major change in mathematics education—the widespread availability and use of symbolic computation. The impact on teaching calculus, for example, will be immense. Modern calculus classes spend most of their time on methods of differentiation and integration, exactly the mechanical aspects that are best handled by symbolic algorithms. With an accessible symbolic system available, teachers can focus on the real value of calculus—its application to solving problems. It is our conjecture that students who understand how to use the new tools to solve problems will be better at performing the mechanical computations without them.

There are many ways in which Scientific WorkPlace can be used in the teaching and learning. The system can be made available in student laboratories for exploration and assignments. It can be used in a lecture with an overhead projector. A brief outline prepared before class can be used to guide the exposition, and student "what if" questions can be answered directly by performing computations that are not obscured by interminable typing and syntax errors. The resulting document can be distributed to students, obviating the need for them to take notes.

Scientific WorkPlace is particularly suited to the trend toward getting mathematics students to write. Its word-processor format makes it easy to prepare and edit assignments, and the high-quality typeset output encourages students to take more care with their writing.

6 Future Directions

The general purpose new features planned for Scientific WorkPlace include availability on more platforms, support for non-English languages, and availability on CD-ROM.

The following planned improvements will enhance the use of the system in the classroom.

1. A scripting language. This will provide a form of programming that is independent of the underlying symbolic system. Our approach will be similar to that taken by Doug Childs in Calculus T/L.

2. Hypertext links. These will provide point and click movement within documents and to other documents.

3. Plot animation.

4. Reduced simplification. Currently, Scientific WorkPlace uses the full Maple simplification process. For teaching purposes, it would be useful to reduce the amount of simplification that is done, allowing students to guide the system through the steps as they learn how to perform them.

COOPERATIVE LEARNING THROUGH ASSIGNMENTS AND MATHEMATICS ACHIEVEMENT

(An Experimental Design)

Charita A. Luna, Ph.D.
Professor 6
MINDANAO POLYTECHNIC STATE COLLEGE
Cagayan de Oro City, Philippines

INTRODUCTION:

Current mathematics teaching practices have come under increasing scrutiny in recent years. Recommendations for reforms from research-based knowledge of how students learn mathematics and how mathematics can be taught effectively are forwarded to improve achievement and develop critical thinking (Cobb, et al 1991). One of such recommended teaching techniques is cooperative learning. Romiszowski (1981) in his book Designing Instructural System stated that group learning is a well-favored procedure of learning. Group learning or presence of group helps the individual to perceive what concepts are to be assimilated and stored for future use. The mutual sharing of insights among learners in a group and the actual overcoming of misconceptions can enhance reflective process and improve the quality of conceptual learning.

One of the difficulties of teaching at all levels lies in the ability of the students to grasp new concepts and to make learning more lasting to those whose rate of assimilation is not as fast as expected. This basic difficulty indicates the need for reinforcement to make ideas clear and lasting, thereby helping students acquire the basic repertoire of behavior to formulate concepts and generalization. A student learns more when each step is immediately strengthened or reinforced through assignments (Eurich, 1962). Inevitably some loss of knowledge may occur after an interval of time, but surprisingly it is resistant to extinction if provided by some forms of reinforcement.

In the study of Emley (1986) on cooperative learning versus individualized instruction in college level remedial mathematics course revealed that group work helps students change attitude towards mathematics significantly. Happs (1985) said that if learning is conceived primarily as a change of behavior due to reinforced practice, instruction should be designed to provide differential reinforcement to correct behavior in the presence of appropriate environment stimuli. He further stressed that the learner is seen to be acquiring knowledge through interaction with his environment. Goldberg (1981) revealed that cooperative problem solving in small groups is helpful both to students and to the teachers for students developed better understanding to the work and had greater incentive to do their assignments. Bryant (1981) reported that team assisted instruction (group work) is superior over the traditional method of instruction (lecture method). Alston (1990) revealed that students in small group problem solving activity exhibited a significantly higher achievement over the control group in traditional method. Miller(1990) reported that cooperative learning groups have significantly changed the students attitude toward mathematics although their achievement was not significantly high.

It could be inferred from previous studies that group learning and reinforcement can enhance learning. This study verified the effect of cooperative learning method through assignments on the achievements of students.

METHOD:

The freshmen students of the school of Industrial Education of Mindanao Polytechnic State College of Cagayan de Oro City school year 1991-1992 taking College Algebra in the second semester were the subjects of the study. The section was randomly selected from two sections as experimental group and the other as control group. The experimental group were grouped in threes to do their assignments while the control group did their assignment individually. An IQ test using OLMAT form J and a pretest in College Algebra were given at the start of second semester in November

1991 to both groups. A posttest was given to the same group at the end of the semester in March 1992. An analysis of variance was used to analyze the IQ raw scores to determine if the two groups were comparable. An analysis of covariance (ANCOVA) was used to analyze the pretest and posttest scores which comprised the data of this study.

RESULTS:

Table 1,2 and 3 showed the data gathered before and after the experiment. Table 1 showed the mean and standard deviation of the pretest and posttest scores in College Algebra of both the control and experimental group. Table 2 showed the mean, standard deviation, variance and F-ratio of the IQ raw scores of both groups. Table 3 showed the analysis of covariance of the pretest and posttest of both the control and experimental groups.

Table 1

Mean and Standard Deviation of Pretest and Posttest

	Control Group		Experimental Group	
	Mean	(SD)	Mean	(SD)
Pretest	24.04	(5.78)	27.52	(7.30)
Posttest	34.09	(7.69)	43.57	(6.21)

Table 2

Analysis of Variance (ANOVA) of OLMAT

Control Group			Experimental Group			Computed	Table	Deci-
Mean	(SD)	S^2	Mean	(SD)	S^2	F-ratio	F-ratio	sion
26.87	(6.72)	45.21	30.43	(6.03)	36.35	1.24	1.94	NS

Table 3

Analysis of Covariance (ANCOVA) of the Pretest and Posttest Scores

Sources of Variation	Adjusted Sum of Squares	df	Adjusted Mean Squares	Computed F-ratio	Table Value of F	Decision
Treatment	535.47	1	535.47	18.69	4.08	P < 0.05
Error w/in	1232.08	43	28.65	-----	-----	-----
Total	1767.55	44	-----	-----	-----	-----

DISCUSSIONS:

The analysis of variance of the IQ scores of the subjects in the control and experimental groups yielded a Computed F-ratio of 1.24 which is less than the table value of 1.94 at 0.05 level. This implies that the two groups were comparable in their IQ.

The analysis of covariance on the pretest and posttest scores of the control group (assignments done individually) and the experimental group (assignments done by group) yielded a computed F-ratio of 18.69 which is greater than the table value of F 4.08 at 0.05 level. This means that the experimental group with a posttest mean of 43.57 is better in achievement than the control group whose posttest mean is 34.09. This result revealed that doing assignment by group helped the students to retain mathematical concepts as shown in their achievements scores. This findings confirmed the studies conducted by Goldberg, Bryant, Alston and Romiszowski that cooperative learning can enhance students achievement.

RECOMMENDATION:

Based on the findings the following recommendation are hereby forwarded:

1. All mathematics teachers at all level should let students perform assignments by group.
2. A study where the students are grouped not only in doing assignments but even during regular classroom instruction.

REFERENCES

1. Alston, Alice., "Effects of Small Groups Problem Solving Activities using concrete Embodiments of Mathematical Properties on Individual Seventh-Grade Children Success in Solving Numerical Problem with the same Structures" DAI. January 1991, Vol. 51 No. 7 2301-A.

2. Artzt, Alice., and Newman, Claire M., "Cooperative Learning" Mathematics Teacher, September 1990 pp. 448-452.

3. Artzt, Alice., "Students Teams in Mathematics Class"., Mathematics Teacher, October 1979 pp. 505-508.

4. Bryant, Revi Ruth., "Effect on Team Assisted Individualization on Attitude and Achievement of 3rd, 4th and 5th grade Students in Mathematics" DAI. July 1982, Vol. 43, No. 1 43-A.

5. Cobb, et. al., Assessment of Problem Centered Second Grade Mathematics Project" Journal of Research in Mathematics Education. January 1991, Vol. 22. No. 1 pp. 3-29.

6. Emley, William Patrick., "The Effectiveness of Cooperative Learning Versus Individualized Instruction in a College Level Remedial Mathematics Course with Relation to Attitude towards Mathematics and Myer - Briggs Personality Type"., DAI., July 1987, Vol. 48., No. 1 70-A.

7. Eurich, A.C., "Technology in Education" New Society, December 1962, pp. 11-15.

8. Goldberg, D.J., "Problem Solving in Small Groups" International Journal Mathematics Education Science Technology, 1981., Vol. 12, No. 2 pp, 211-217.

9. Happs, John C. "Cognitive Learning Theories and Classroom Complexity" Research in Science and Technology Education Vol. 3. November 1985.

10. Miller, Wynn Arthur., "Effects of a Cooperative Learning Format on Content, Attitude and Learning Preferences in Mathematics for a Population of Junior High School Summer Academy Students" DAI. November 1990 Vol. 51, 1540-A.

11. Romiszowski, Alexander L., "Designing Instructional System Decision Making in Course Planning and Curriculum Design" Kogan Paze., London Nichols Publishing, New York, 1981.

MATHEMATICS EDUCATION FOR BIOLOGISTS

V. P. Saxena
School of Mathematics and Allied Sciences,
Jiwaji University, Gwalior - 474 011 (India)

ABSTRACT

Biological sciences are going more and more mathematical in recent times. The modern trends in research and teaching in biology indicate that the knowledge of mathematical and computational techniques is necessary. Therefore, mathematics has become an important component of the course curriculum for biologists. However, the mathematics education in present form, is not yet well accepted by the students and researchers of biological sciences. This is due to biological orientation of their minds and traditional way of teaching of the subject. Thus it is necessary to evolve new technology and new methodology in this direction. Infect, the teaching of mathematics should go hand in hand with its applications to biology and its viability for computer use.

INTRODUCTION:-

Mathematics is a subject which has its roots in various other subjects and disciplines. Infect many fields of knowledge grow because of mathematics and new branches of mathematics continuously emerge due to developments in these fields. Physical sciences,engineering and economics are the well known users of mathematics. The teaching programs of these branches deeply involve mathematics education which is well integrated with the curricula. In physical sciences the mathematics is always considered as a major subject and there are constant efforts to improve its teaching (see [Lochhead, Page 55; Saxena, Page 90]). In engineering institutions also it is always treated as an important subject (see [Greber, Page 71]). Similarly, in social sciences and management programs the awareness for mathematics education is becoming more and more imperative (see [Norman, Page 75; Zionts, Page 81]). Some new mathematical subjects like Industrial Mathematics and Computer Science have also widened the scope of mathematics education (see [Saxena, Page 45; Saxena, Page 23; Scherlis Shaw, Page 89]).

A new challenge faced today by mathematical educationists is imparting the training and teaching of this subject to biologists in general and to ecologists, atmospheric and medical students in particular. Infect the biological sciences and mathematics are considered to be non- interacting streams except in the area of population dynamics and genetics. The statistical techniques are also used by many biology students to analyse and present their data. But these methods were not having much quantitative orientation of deterministic nature.

The development of biological sciences during last two decades or so has been a significant transformation of this field in respect of mathematical modeling and mathematical formulation. It is now well realised that in-depth study of biological processes, right from microbial entities to astronomical domains, needs theoretical and mathematical infrastructure. Many laws and principles applied to physical sciences so far, have been made applicable to biological activities and states to explain them in more explicit way.

In view of above the mathematics education for biologists has become necessary at all levels. However, the diversity in the approach of teaching of biology and mathematics make the task of mathematics teachers in biological centres difficult and unpopular. Therefore, an altogether new and acceptable methodology has to be deviced to make the mathematics teaching meaningful.

TEACHING METHODOLOGY

The mathematics courses have already been deviced in various universities for the students of biological sciences. These courses are based on some standard texts available in the market under the title Biomathematics. These courses generally start with elementary calculus, discrete mathematics and linear algebra. The students who do not have the background of these topics do not welcome these teaching programs whole heartedly. Generally they think that something new is imposed in their curriculum which may or may not be useful to them in their course topics. By the time they realise the importance of Biomathematics in their advanced teaching and research it becomes too late for them to restudy these basic subjects.

To impart the teaching of these subjects it is necessary that the mathematics, whatsoever, should emerge from the biological concepts in such a way that the beneficiary pupils should accepts it as a part of concerned biological description. For this purpose the mathematics should be taught in the following order.

(a) Discrete Mathematics
- (i) Numbers and Number System
- (ii) Numerical Sequences and Series
- (iii) Finite Difference and Interpolation
- (iv) Difference Equations

(b) Calculus
- (i) Functions and Their Behavior
- (ii) Limit of Functions
- (iii) Rate of Change of Functions
- (iv) Graphs of Functions

(v) Derivatives
(vi) Integrals
(vii) Differential Equations

(c) Linear Algebra
(i) Determinants and Matrices
(ii) Linear Equations
(iii) Linear System of Difference Equations
(iv) Linear System of Ordinary Differential Equations

(d) Boundary Value Problems
(i) Initial Value Problems as O.D.E.
(ii) Boundary Value Problems as O.D.E.
(iii) Partial Differential Equations in Biology
(iv) Boundary Value Problems in P.D.E.

(e) Mathematical Methods
(i) Series Solutions and Special Functions
(ii) Laplace Transform
(iii) Other Integral Transform
(iv) Numerical Solutions of B.V.P.
(v) Finite Element Method

As indicated above all these topics should be introduced through the description of biological processeces and their quantification. Some examples of this approach in basic topics are given below.

BIOLOGICAL INTRODUCTION OF MATHEMATICAL TOPICS

(a) Discrete Mathematics

The introduction to numbers through biology is not difficult. This can be done through the structure and patterns in plants, flowers and fruits. Numbers and their occurrence with monotonicity and periodicity can be illustrated through the life patterns of animals and their behavior. Similarly the introduction to numerical sequences can be done through the distribution of petals in flowers and leaves and branches in plants. For example the distribution of ridges in pineapple can be given by the formula

$F(k+1) = F(k-1) + F(k)$, $k = 2,3,4$

where $F(1) = F(2) = 1$.

This distribution forms a well-known numerical sequence known as Fibonacci sequence.

The finite differences and difference equations can be introduced through population studies of finite number of animals and human beings. Some examples on wild life and bird sanctuaries can make the students convinced regarding the use of quantitative studies. For example a population of finite size can be represented by $P(n)$ where n is the number of generation and its linear growth can be written as

$$P(n) = a\,P(n-1) + b$$

Above equation can be easily solved by induction method which does not need knowledge of any advanced mathematics. Interpolation can also be introduced by using any field data table and finding the intermediate value of field variable.

The introduction of calculus should be done carefully as this topic is most applicable in biological studies. Before going deep into the subject a thorough discussion on dependent and independent variables in biological world is necessary. All types of examples in nature are available as far as functions are concerned. This may include discrete and continuous functions, boundedness, monotonicity and periodicity etc.. Once the students understand the occurrence of such functions in nature, they can easily grass the concept of limit and continuity.

The rate of change of functions of all types can also be taught through field examples and biological processes. Such examples can be taken from the growth of bacteria, viruses and insects. This idea can be further extended to define the derivatives of continuous functions.

After gaining knowledge of differentiation and integration, mathematical modeling of many biological phenomena may lead to the definition of simple differential equations. The best example may be the equation

$$dp/dt = AP$$

which gives the quantitative idea of exponential growth. This idea can be modified to define many other equations. The non- linear concept can also be given through logistic growth and epidemiology.

Simultaneous ordinary differential equations emerge due to mathematical modeling of two or more species population models, enzyme kinetics and flow of substances in biological compartments.

If above part of mathematics teaching is well assimilated in the minds of biologists then the teaching of linear algebra does not remain difficult and can be given as a follow up topics.

Likewise other advanced topics like partial differential equations can be explain through bio diffusion and biofluid dynamics. Some of the specific examples in this regard are diffusion through membranes and tissues, blood flow, air and water pollution and so on.

SUPPORTING ACTIVITY

As biological sciences are based mostly on experimental investigations and field studies, it is necessary to include the application of mathematical tools directly to real problems by the students themselves. For this purpose small projects should be included in the mathematics education program on direct application of mathematical formula to biological problems. They should know how to generate quantitative data, to search appropriate mathematical Formula and to draw directed conclusions. This will also help them to make computer applications in their subsequent studies.

REFERENCES

[Lochhead] Jack Lochhead, The Mathematical Needs of Students in The Physical Sciences, Springer - Verlag New York 1983.

[Saxena] V.P.Saxena, New Challenges For College Mathematics,Mathematical Education 1988.

[Greber] Isaac Greber, Engineering Needs and The College Mathematics Core, Springer - Verlag New york 1983.

[Norman] Robert Norman, The First Two Years of Mathematics at a University as It Relates to The Mathematical Needs of Students in The Social Sciences, Springer - Verlag New York 1983.

[Zionts] Stanley Zionts, Mathematics in Business and Management, Springer - Verlag New York 1983.

[Saxena] V.P.Saxena, Application of Computers in Mathematics Education,Mathematical Education 1989.

[Saxena] V.P.Saxena, Industrial Mathematics: An Over View of The Base, Mathematical Education 1992.

[Scherlis Shaw] William L. Scherlis and Mary Shaw, Mathematics curriculum and the Needs of computer Science, Springer-Verlag New York 1983.

\ \ \ \ \ \ \ \ \ \ \

Identification of Unknown Terms in a Nonlinear Parabolic Problem

Abdullah Shidfar   Hossein Azary

Department of Mathematics
Iran University of Science and Technology
Narmak, Tehran–16, Iran

Abstract

In this paper we identify two unknown radiation terms in a linear heat equation with nonlinear boundary conditions from two overspecified data measured on the boundary. Explicit solutions for the temperature and radiation terms are derived. Some uniqueness, stability and asymptotic results are presented. Finally the numerical solution of the inverse problem will be compared with its analytical solution with some experimental results.

Key words: Radiation terms, Overspecified condition, Inverse problem, Heat equation, Numerical solution

1 Introduction

The problem of determining unknown parameters in parabolic differential equations has been treated previously by many authors [1–4]. Usually these problems involve the determination of a single unknown parameter from overspecified boundary data. In some applications, however, it is desirable to be able to determine more than one parameter from the given boundary data [2,6].

This paper seeks to determine some unknown radiation functions those depend only on the heat flux in a radiative heat transfer equation.

Hence, we may consider the following problem

$$\begin{aligned}
\partial_t u(x,t) &= \partial_{xx} u(x,t), \quad 0 < x < 1, \quad 0 < t < T, & (1)\\
u(x,0) &= f(x), \qquad 0 < x < 1, & (2)\\
u(0,t) &= \phi(u_x(0,t)) + g_0(t), \qquad 0 < t < T, & (3)\\
u(1,t) &= \psi(u_x(1,t)) + h_0(t), \qquad 0 < t < T, & (4)
\end{aligned}$$

and the overspecified condition

$$u_x(0,t) = g_1(t); \quad t > 0, \tag{5}$$

$$u_x(1,t) = h_1(t); \quad t > 0, \tag{6}$$

where T is a given constant and $f(x)$, $g_0(t)$, $h_0(t)$, $g_1(t)$, and $h_1(t)$ are given functions.

2 Existence and uniqueness results

To solve the inverse problem (1)-(6), let us consider the following auxiliary problem

$$\begin{aligned}
u_t &= u_{xx}, \qquad 0 < x < 1, \quad 0 < t < T, & (7)\\
u(x,0) &= f(x), \qquad 0 \le x \le 1, & (8)\\
u_x(0,t) &= g_1(t), \qquad t \ge 0, & (9)\\
u_x(1,t) &= h_1(t), \qquad t \ge 0. & (10)
\end{aligned}$$

For any piecewise–continuous functions f, g_1, and h_1 this problem has a unique solution [1]

$$\begin{aligned}
u(x,t) &= \int_0^1 \{\theta(x-\xi,t) + \theta(x+\xi,t)\} f(\xi) d\xi \\
&- 2\int_0^t \theta(x,t-\tau) g_1(\tau) d\tau + 2\int_0^t \theta(x-1,t-\tau) h_1(\tau) d\tau, \qquad (11)
\end{aligned}$$

where

$$\theta(x,t) = \sum_{m=-\infty}^{\infty} k(x+2m,t),$$

and

$$k(x,t) = \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{x^2}{4t}\right), \quad t > 0.$$

(3) and (5) yield

$$\phi(g_1(t)) = u(0,t) - g_0(t) \tag{12}$$

From $\theta(-\xi,t) = \theta(\xi,t)$, (11) and (12) we conclude that

$$\begin{aligned}\phi(g_1(t)) &= 2\int_0^1 \theta(\xi,t)f(\xi)d\xi \\ &- 2\int_0^t \theta(0,t-\tau)g_1(\tau)d\tau + 2\int_0^t \theta(-1,t-\tau)h_1(\tau)d\tau - g_0(t).\end{aligned}$$

If we assume that the function $s = g_1(t)$ is invertible, then we find

$$\begin{aligned}\phi(s) &= 2\int_0^1 \theta(\xi, g_1^{-1}(s))f(\xi)d\xi \\ &- 2\int_0^{g_1^{-1}(s)} \theta(0, g_1^{-1}(s)-\tau)g_1(\tau)d\tau \\ &+ 2\int_0^{g_1^{-1}(s)} \theta(-1, g_1^{-1}(s)-\tau)h_1(\tau)d\tau - g_0(g_1^{-1}(s)).\end{aligned} \tag{13}$$

Similar result for ψ, in the form

$$\begin{aligned}\psi(\nu) &= 2\int_0^1 \theta(\xi+1, h_1^{-1}(\nu))f(\xi)d\xi \\ &- 2\int_0^{h_1^{-1}(\nu)} \theta(1, h_1^{-1}(\nu)-\tau)g_1(\tau)d\tau \\ &+ 2\int_0^{h_1^{-1}(\nu)} \theta(0, h_1^{-1}(\nu)-\tau)h_1(\tau)d\tau - h_0(h_1^{-1}(\nu))\end{aligned} \tag{14}$$

for any invertible function $\nu = h_1(t)$ may be obtained from $\theta(1-\xi,t) = \theta(1+\xi,t)$ and conditions (4) and (6).

For any given Lipschitz continuous functions ϕ and ψ, and piecewise-continous functions f, g_0, and h_0, the problem (1)–(4) has a unique solution of the form [1].

$$\begin{aligned}u(x,t) &= w(x,t) - 2\int_0^t \frac{\partial\theta}{\partial x}(x,t-\tau)\{\phi(\phi_0(t)) + g_0(\tau)\}d\tau \\ &+ 2\int_0^t \frac{\partial\theta}{\partial x}(x-1,t-\tau)\{\psi(\phi_1(t)) + h_0(\tau)\}d\tau,\end{aligned} \tag{15}$$

where

$$w(x,t) = \int_0^1 \{\theta(x-\xi,t) + \theta(x+\xi,t)\} f(\xi) d\xi,$$

if and only if ϕ_0 and ϕ_1 are piecewise–continuous functions and satisfy

$$\begin{aligned}
\phi_0(t) &= w(0,t) - 2\int_0^t \frac{\partial\theta}{\partial x}(0,t-\tau)\{\phi(\phi_0(t)) + g_0(\tau)\}d\tau \\
&+ 2\int_0^t \frac{\partial\theta}{\partial x}(-1,t-\tau)\{\psi(\phi_1(t)) + h_0(\tau)\}d\tau
\end{aligned}$$

$$\begin{aligned}
\phi_1(t) &= w(1,t) - 2\int_0^t \frac{\partial\theta}{\partial x}(1,t-\tau)\{\phi(\phi_0(t)) + g_0(\tau)\}d\tau \\
&+ 2\int_0^t \frac{\partial\theta}{\partial x}(0,t-\tau)\{\psi(\phi_1(t)) + h_0(\tau)\}d\tau
\end{aligned}$$

In order to prove the uniqueness of solution $(u,(\phi,\psi))$, we use the system of Volterra integral equations

$$\begin{aligned}
\eta_0(t) &= -2\int_0^t \frac{\partial\theta}{\partial x}(0,t-\tau)\phi(g_1(\tau))d\tau \\
&+ 2\int_0^t \frac{\partial\theta}{\partial x}(-1,t-\tau)\psi(h_1(\tau))d\tau,
\end{aligned} \tag{16}$$

and

$$\begin{aligned}
\eta_1(t) &= -2\int_0^t \frac{\partial\theta}{\partial x}(1,t-\tau)\phi(g_1(\tau))d\tau \\
&+ 2\int_0^t \frac{\partial\theta}{\partial x}(0,t-\tau)\psi(h_1(\tau))d\tau,
\end{aligned} \tag{17}$$

where

$$\begin{aligned}
\eta_0(t) &= g_1(t) - 2\int_0^t \frac{\partial\theta}{\partial x}(\xi,t) f(\xi)d\xi \\
&+ 2\int_0^t \frac{\partial\theta}{\partial x}(0,t-\tau)g_0(\tau))d\tau - 2\int_0^t \frac{\partial\theta}{\partial x}(-1,t-\tau)h_0(\tau)d\tau
\end{aligned}$$

and

$$\begin{aligned}
\eta_1(t) &= h_1(t) - 2\int_0^t \frac{\partial\theta}{\partial x}(1+\xi,t) f(\xi)d\xi \\
&+ 2\int_0^t \frac{\partial\theta}{\partial x}(1,t-\tau)g_0(\tau))d\tau - 2\int_0^t \frac{\partial\theta}{\partial x}(0,t-\tau)h_0(\tau)d\tau.
\end{aligned}$$

The system of linear Volterra integral equations of the first kind (16) and (17) has a unique solution (ϕ, ψ). Now the unicity of may be easily concluded from (15) and [1,5].

Now we can summarize the above results in the following statement.

Theorem 2.1 *For any given piecewise–continuous functions and f, g_0, h_0, g_1, and h_1 and invertible functions g_1 and h_1, and Lipschitz continuous functions ϕ and ψ, the inverse problem (1)–(6) has a unique solution pair $(u, (\phi, \psi))$, where u, ϕ, and ψ may be obtained from (11), (13), and (14), respectively.*

3 Some monotonic results

In this section we consider some monotone asymptotic results. First, by demonstrating the following statement, we discuss the strcitly monotone of solutions.

Theorem 3.1 *If g_0 and $g_1 = h_1$ are strictly decreasing and continuous functions and $f = 0$, then ϕ is an strictly decreasing function.*

Proof. By differentiating from (15) with respect to t, we obtain

$$\begin{aligned} u_t(x,t) &= w_t(x,t) - 2\{\theta(x,0)g_1(t) + \int_0^t \frac{\partial\theta}{\partial x}(x, t-\tau)g_1(\tau)d\tau\} \\ &+ 2\{\theta(x-1,0)h_1(t) + \int_0^t \frac{\partial\theta}{\partial x}(x-1, t-\tau)h_1(\tau)d\tau\} \end{aligned}$$

Using the properties of $\theta(x,t)$ function

$$\lim_{\tau \uparrow t} \theta(x, t-\tau) = 0, \qquad 0 < x < 1,$$

$$\begin{aligned} u_t(0,t) &= -2\int_0^t \frac{\partial\theta}{\partial x}(0, t-\tau)g_1(\tau)d\tau \\ &+ 2\int_0^t \frac{\partial\theta}{\partial x}(-1, t-\tau)h_1(\tau)d\tau, \end{aligned}$$

$$\frac{\partial\theta}{\partial t} = -\frac{\partial\theta}{\partial\tau},$$

and integrating by parts yields

$$\begin{aligned} u_t(0,t) &= -2\theta(0,t)g_1(0) - 2\int_0^t \theta(0,t-\tau)g_1'(\tau)d\tau \\ &+ 2\theta(-1,t)h_1(1) + 2\int_0^t \theta(-1,t-\tau)h_1'(\tau)d\tau \end{aligned}$$

From $g_1(0) = h_1(0) = 0,\ 2\theta(0,t) > 1$ and $0 < 2\theta(-1,t) < 1$, we conclude that

$$-\beta(g_1(t) - g_1(0)) + (h_1(t) - h_1(0)) = g_1(t)(1-\beta) > 0, \tag{18}$$

where

$$\beta = \sup_{0\le t\le T} \{2\theta(o,t)\} > 1,$$

and $h_1 = g_1$ is an strictly decreasing functions. From (18), we obtain $u_t(0,t) > 0$. Now from (12) and $s = g_1(t)$, we find

$$\phi'(s) = \left(\frac{\partial u(0,t)}{\partial t} - \frac{\partial g_0(t)}{\partial t}\right)\frac{1}{g_1'(t)} < 0.$$

Then ϕ is an strictly decrasing function. In the case of h_0 and $g_1 = h_1$ are strictly decreasing and continuous functions and $f = 0$ then we may obtain a similar result for ψ.

4 Numerical procedure

This section compare the analytical solutions (13) and (14) of problem (1)–(6) with some experimental results.

We apply the Crank–Nicolson method. For this purpose, we choose $f = g_0 = h_0 = 0$, $g_1(t) = 100t,\ \ h_1(t) = 5t,\ \ \delta t = 0.0025$, and $\delta x = 0.05$. For calculation of $\theta(x,t)$, we use the first 51 terms of its series.

Then (13) can be written in the form

$$\phi(s) = \frac{-2s^{\frac{3}{2}}}{15\sqrt{\pi}} + \frac{5}{\pi}\left(\frac{-\left(4\sqrt{\pi} - 3\text{Gamma}(-\frac{3}{2}, 0, \frac{25}{s})\right)}{24}\right. \tag{19}$$

$$
\begin{aligned}
&\left. - \frac{s\left(2\sqrt{\pi} + \mathrm{Gamma}(-\frac{1}{2}, 0, \frac{25}{s})\right)}{200} \right) \\
&+ \frac{5}{\pi} \left(\sum_{m=1}^{50} \left\{ \frac{-(\sqrt{\pi}\mathrm{Abs}(-1+2m))}{6} \right. \right. \\
&+ \frac{2m\sqrt{\pi}\,\mathrm{Abs}(-1+2m)}{3} - \frac{2m^2\sqrt{\pi}\,\mathrm{Abs}(-1+2m)}{3} \\
&+ \frac{20m^2\,\mathrm{Abs}(m)\,\left(4\sqrt{\pi} - 3\,\mathrm{Gamma}(-\frac{3}{2}, 0, \frac{100m^2}{s})\right)}{3} \\
&+ \frac{\mathrm{Abs}(-1+2m)\,\mathrm{Gamma}(-\frac{3}{2}, 0, \frac{-25\left(-1+4m-4m^2\right)}{s})}{8} \\
&- \frac{m\,\mathrm{Abs}(-1+2m)\,\mathrm{Gamma}(-\frac{3}{2}, 0, \frac{-25\left(-1+4m-4m^2\right)}{s})}{2} \\
&+ \frac{m^2\,\mathrm{Abs}(-1+2m)\,\mathrm{Gamma}(-\frac{3}{2}, 0, \frac{-25\left(-1+4m-4m^2\right)}{s})}{2} \\
&+ \frac{s\,\mathrm{Abs}(m)\,\left(2\sqrt{\pi} + \mathrm{Gamma}(-\frac{1}{2}, 0, \frac{100m^2}{s})\right)}{5} \\
&\left. \left. + \frac{s\left(-(\sqrt{\pi}\mathrm{Abs}(-1+2m)) - \frac{\mathrm{Abs}(-1+2m)\mathrm{Gamma}(-\frac{1}{2},0,\frac{-25\left(-1+4m-4m^2\right)}{s})}{2}\right)}{100} \right\} \right) \right).
\end{aligned}
$$

The results for ϕ is plotted in Figures 1 and 2. After Four hundreds time steps, in Figure 3, the numerical and analytical results cover each others.

References

[1] J.R. Cannon, "*The One-Dimensional Heat Equation*", Addison-Wesely Co. Menlo park, California, 1984.

[2] J.R. Cannon and D.Zachmann ,"*Parameter determination in parabolic differential equations from overspecified boundary data*", Int. J. Engng. Sci. 20(1982), 779–788.

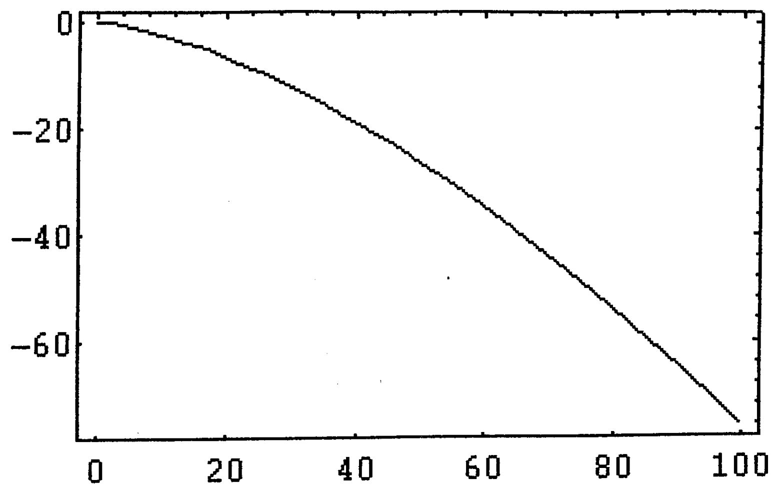


Figure 1: The graph of the analytical solution $\phi(s)$

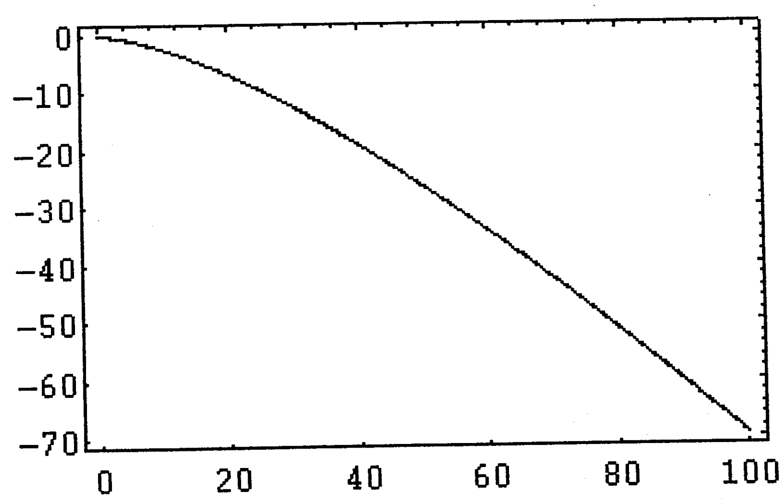


Figure 2: The graph of the numerical solution $\phi(s)$ by using C.N. method

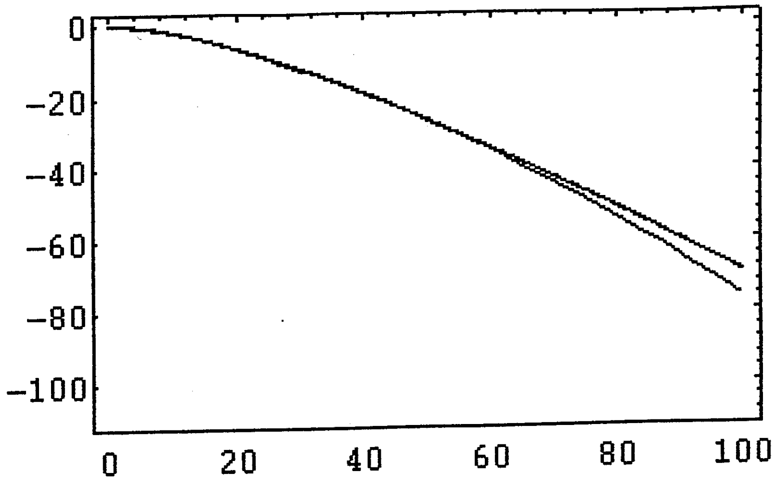


Figure 3: The graphes of analytical and numerical solution of $\phi(s)$

[3] J.R. Cannon and P. DuChateau, "*An inverse problem for a non–linear diffusion equation*", SIAM J. Appl. Math., 39 (1980), 272–289.

[4] J.R. Cannon and P. DuChateau, "*Determining unknown coefficients in a non–linear conduction problem*", SIAM J. Appl. Math., 24(1973), 298–314.

[5] P. DuChateau and D. Zachmann , "*Partial Differential Equations*", Schaum's Outlines 1986.

[6] A. Friedman, "*Partial Differential Equations of Parabolic Type,*", Prentice–Hall, Englewood Cliffs. NJ, 1964.

[7] O.A. Ladyzhenskaya, V.A. Sollonikov and N.N. Uralceva,"*Linear and Quasilinear equations of parabolic type*", AMS, Providence RI, 1968.

[8] A. Shidfar, "*Identifying an unknown term in an inverse problem of linear diffusion equation*", Int, J. Engng. Sci., 26 (1988), 753–755.

[9] A. Shidfar, "*Determination of an unknown radiation term in a heat conduction problem*", Differential and Integral Equations, 3 (1990), 1225–1229.

A CURIOUS CONNECTION BETWEEN CHEBYSHEV POLYNOMIALS AND FIBONACCI NUMBERS

by

O. Jayanth and R. Sivaramakrishnan
Department of Mathematics, University of Calicut,
India 673 635

ABSTRACT

We introduce the Chebyshev coefficients

$$T_{n,k} = \frac{n}{n-k}\binom{n-k}{k} 2^{n-2k-1}, \quad 0 \le k \le [n/2]$$

$$U_{n+1,k} = \binom{n-k}{k} 2^{n-2k-1}, \quad 0 \le k \le [n/2]$$

using the well-known Chebyshev polynomials.
We take $T_{0,0} = 1$, $U_{1,0} = 0$.
If $\{F_n\}$ denotes the Fibonacci sequence, it is shown that

(1) $$F_n = \sum_{k=0}^{[\frac{n}{3}]} (-1)^k T_{n-k,k}$$

and

(2) $$F_n = 1 + \sum_{k=0}^{[\frac{n-2}{3}]} (-1)^k U_{n-k-1,k} .$$

A CURIOUS CONNECTION BETWEEN CHEBYSHEV POLYNOMIALS AND FIBONACCI NUMBERS

by

O. Jayanth and R. Sivaramakrishnan
Department of Mathematics, University of Calicut
Kozhikode 673 635, INDIA

1. INTRODUCTION

It is well-known that the study of Chebyshev polynomials of the first and second kind arises in connection with the solution of

$$(1-x^2)\frac{d^2y}{dx^2} - x\frac{dy}{dx} + n^2y = 0$$

when n is an integer $\geqslant 0$ and $|x|<1$.

As in [1] or [2], we define

(1.1) $$T_n(x) = \cos(n \arccos x)$$

(1.2) $$U_n(x) = \frac{\sin(n+1)\arccos x)}{\sin(\arccos x)}$$

Then, $T_n(x)$ has the form [2]

(1.3) $$T_n(x) = \tfrac{1}{2}\left\{(x+\sqrt{x^2-1})^n + (x-\sqrt{x^2-1})^n\right\}$$

Similarly,

(1.4) $$(\sqrt{x^2-1})\,U_n(x) = \tfrac{1}{2}\left\{(x+\sqrt{x^2-1})^{n+1} - (x-\sqrt{x^2-1})^{n+1}\right\}$$

From (1.3) and (1.4), we have

$$T_o(x) = 1, \quad T_1(x) = x \text{ and } T_2(x) = 2x^2-1$$

and

$$U_o(x) = 1, \quad U_1(x) = x \text{ and } U_2(x) = 4x^2-1$$

The polynomial expressions for $T_n(x)$ and $U_n(x)$ are given by

$$(1.5) \quad T_n(x) = \frac{1}{2} \sum_{k=0}^{[\frac{n}{2}]} (-1)^k \frac{n}{n-k} \binom{n-k}{k} (2x)^{n-2k}$$

where the notation [x] stands for the greatest integer not exceeding x.

$$(1.6) \quad U_n(x) = \sum_{k=0}^{[\frac{n}{2}]} (-1)^k \binom{n-k}{k} (2x)^{n-2k}$$

It is known [1] that

$$(1.7) \quad T_n(x) + T_{n-2}(x) = 2x\, T_{n-1}(x), \quad n = 2, 3, \ldots..$$

$$(1.8) \quad U_n(x) - U_{n-2}(x) = 2\, T_n(x), \quad n = 2, 3, \ldots..$$

We introduce the Chebyshev coefficients $T_{n,k}$ and $U_{n,k}$ as follows:

$$(1.9) \quad T_{n,k} = \frac{n}{n-k} \binom{n-k}{k} 2^{n-2k-1}, \quad 0 \leq k \leq [n/2]$$

and

$$(1.10) \quad U_{n+1,k} = \binom{n-k}{k} 2^{n-2k-1}, \quad 0 \leq k \leq [n/2]$$

We assume that $T_{0,0} = 1$ and $U_{1,0} = 0$

The purpose of this note is to point out that $T_{n,k}$ and $U_{n,k}$ are connected with the Fibonacci sequence $\{F_n\}$ given by

$$(1.11) \qquad F_{n+1} = F_n + F_{n-1}, \qquad n \geqslant 1$$

with $F_0 = F_1 = 1$. See theorems 3 and 4 in section 3. The curious connection with $\{F_n\}$ is achieved via certain recurrence relations shown in section 2.

2. RECURRENCE RELATIONS:

We give below certain recurrence relations involving $T_{n,k}$ and $U_{n,k}$. They are direct consequences of the formulae (1.9) and (1.10).

THEOREM: 1. If $T_{n,k}$ and $U_{n,k}$ are defined as in (1.9) and (1.10) respectively, then

$$(2.1) \qquad T_{n,k} = \begin{cases} 2\, T_{n-1,k} + T_{n-2,k-1} & (n > 1,\ k > 0) \\ 2\, T_{n-1,k} & (n \geqslant 1,\ k = 0) \end{cases}$$

$$(2.2) \qquad U_{n,k} = \begin{cases} 2\, U_{n-1,k} + U_{n-2,k-1} & (n > 1,\ k > 0) \\ 2\, U_{n-1,k} & (n \geqslant 1,\ k = 0) \end{cases}$$

Proof: The simplification of $2\, T_{n-1,k} + T_{n-2,k-1}$ using (1.9) yields:

When $k \geqslant 1$

$$2\, T_{n-1,k} + T_{n-2,k-1} = 2 \frac{n-1}{n-k-1} \binom{n-k-1}{k} 2^{n-2k-2} + \frac{n-2}{n-k-1} \binom{n-k-1}{k-1} 2^{n-2k-1}$$

$$= \frac{n\,(n-k-1)\,!}{k!\,(n-2k)\,!}\,2^{n-2k-1}$$

$$= \frac{n}{n-k}\,\frac{(n-k)\,!}{k!\,(n-2k)\,!}\,2^{n-2k-1}$$

$$= \frac{n}{n-k}\binom{n-k}{k}2^{n-2k-1}$$

$$= T_{n,k}$$

If $k = 0$, $T_{n,0} = 2^{n-1} = 2\,T_{n-1,0}$

This proves (2.1). The proof of (2.2) follows on similar lines.

THEOREM: 2. If $T_{n,k}$ and $U_{n,k}$ are as defined in (1.9) and (1.10) respectively then for $n \geqslant 1$, and for admissible values of k,

(2.3) $$T_{n,k} + U_{n,k} = U_{n+1,k}$$

Proof:- For $n \geqslant 1$, $k \geqslant 0$, one has

(2.4) $$T_{n,k} + U_{n,k} = \frac{n}{n-k}\binom{n-k}{k}2^{n-2k-1} + \binom{n-k-1}{k}2^{n-2k-1}$$

where $k \leq [n-1/2]$. However, the right side of (2.4) reduces to

$$\binom{n-k}{k}2^{n-2k} = U_{n+1,k}$$

So (2.3) is valid for admissible values of k.

The following recurrence relations are stated without proof:

$$U_{n,k} = 2\, T_{n-1,k} - U_{n-2,k-1} \quad (n > 2,\ k > 1) \tag{2.5}$$

$$U_{n+1,k} + U_{n,k-1} = T_{n+1,k} \quad (n \geqslant 1,\ k \geqslant 1) \tag{2.6}$$

3. CONNECTION WITH FIBONACCI SEQUENCES:

THEOREM: 3 If F_n is as defined in (1.11), then for $n \geqslant 1$

$$F_n = \sum_{k=0}^{[\frac{n}{3}]} (-1)^k\, T_{n-k,k} \tag{3.1}$$

Proof: We note that the summation on the right of (3.1) is meaningful only for values of k for which k [n-k/2]. This implies that k ranges from 0 to [n/3].

Writing

$$a_n = \sum_{k=0}^{[\frac{n}{3}]} (-1)^k\, T_{n-k,k}$$

We obtain form (2.1)

$$a_n = \sum_{k=0}^{[\frac{n}{3}]} (-1)^k \left\{ 2\, T_{n-k-1} + T_{n-k-2,\, k-1} \right\}$$

Splitting the sum on the right side into two sums with appropriate values for summation index k or t, we have

$$a_n = 2 \sum_{k=0}^{J_1} (-1)^k\, T_{n-k-1,k} + \sum_{t=0}^{J_3} (-1)^{t+1}\, T_{n-t-3,t}$$

where $J_1 = [n-1/3]$ and $J_3 = [n-3/3]$

Therefore we obtain

$$(3.2) \qquad a_n = 2\, a_{n-1} - a_{n-3}$$

As $a_0 = a_1 = 1$, from (3.2), we conclude that $a_n = F_n$ for $n \geqslant 0$, and this proves (3.1).

THEOREM: 4. If $\{F_n\}$ is as defined in (1.11), then for $n \geqslant 1$

$$(3.3) \qquad f_n = 1 + \sum_{k=0}^{[\frac{n-2}{3}]} (-1)^k\, U_{n-k-1,k}$$

Proof: The summation on the right side of (3.3) is meaningful only for values of k from 0 to [n-2/3] as $U_{n-k-1,k}$ is defined only for $k \leq \left[\frac{n-2}{3}\right]$

Analogous to the proof of (3.3), one has to make use of (2.2) for arriving at (3.3). The details are omitted.

R E F E R E N C E S

1. John Riordan : Combinatorial Identities: John Wiley & Sons Inc (1968) Chapter 2, Section 2.4 54-65

2. George F. Simmons : Differential Equations with applications and historical notes Tata McGraw Hill Pub Co Ltd. New Delhi Fourteenth reprint (1992) Chapter 5 Appendix D 204-211

February 10, 1995.

Computer Fitness Test for Tertiary Students

by

Dr Rosalind Phang
Dr Teo Beng Chong

Division of Mathematics, School of Science
Nanyang Technological University
National Institute of Education
469 Bukit Timah Road
Singapore

Abstract

A statistical fitness test using the computer was conceptualise, designed and pilot-tested with some students at the National Institute of Education. This paper will discuss the rationale of conducting such a test, the design and implementation of the test.

Introduction

It is not uncommon to hear instructors lament that their students do not seem to have learnt or remember what they have been taught just the previous semester. Often instructors feel that it is necessary to go over substantial amounts of material that students are expected to have grasped earlier. In fact, many feel that some of the students are not 'fit' to be taking a particular course. We are proposing a fitness test for students who are taking a second or subsequent courses for a particular subject. We have chosen statistics as the subject of interest.

Rationale

This test is designed to help instructors and students in the following ways:

1. it will help 'push' students to revise concepts of the last course
2. students has the opportunity to re-test at his own time
3. the test is self-correcting and students have feedback (marks) immediately.
4. it will enable instructors to understand student's difficulties and student's possible misconceptions. This will enable the instructor to evaluate his approach to the subject.

Background of Subject chosen

This test is intended to serve as a fitness test for students doing their second course in statistics at the National Institute of Education, Singapore. The test is based on the course material in the first course of statistics.

The 5 main topics or categories are

1. Descriptive statistics
2. Probability
3. Discrete distributions (binomial, Poisson)
4. Continuous distributions (uniform, exponential, normal)
5. Use of MINITAB

Design of Test Questions

The test questions are of two types:

TYPE 1:
For each question on the above categories, the students will be shown two possible answers. Students will be shown one answer at a time and be required to answer *true* or *false* or *don't know* and to support their answer with a brief reason.

TYPE 2:
This is a multiple-choice question and students are also expected to supply a reason for the choice made.

An example of the two types of questions and its display format can be found in Appendix 1.

Students are required to supply a short reason to their choice and it is hoped that this will allow the instructor to spot any misconceptions that the student have but is still lucky enough to pick on the right answers.

Developing the test

The design of this fitness test includes a set-up program that will provide for easy input of questions by the instructors. Appendix 2 shows a sample of the screen for the test set-up. Adding and editing questions and answers can be done easily.
The test is scored automatically, 5 points for category 1 questions and 10 points for category 2 questions. The responses given by the students are logged by the computer for instructors reference. There is also a time limit of one hour for the test. This time is more than sufficient as we would like students to spend time thinking but not to hog the computer for too long.

Administering the Test

All students taking the second course in statistics must do this fitness test and clear it by the end of the second week into the semester. The student will be given three chances to retake the test. If the student fails after the third attempt, he will be referred to the instructor for advice. However, students are only allowed a re-test after 48 hours.

The student is expected to answer ten (True/False) questions and five multiple-choice questions in a maximum time of 1 hour. Questions are chosen from the pool of questions for the above topics.

As the questions are permuted and students are allowed to retake the test, they will not be able to obtain a copy of the questions. They will have to sign in to take the test and the laboratory supervisors have been instructed that there should be no discussion among the students when they are taking the test. They can only take the test with an access diskette from the laboratory supervisors.

Pilot Test

This was pilot tested with ten year 4 students. We asked students three main areas: presentation, test material and time constraint. These were some of the comments made by the students:

(i) Presentation: 95% of the students are comfortable with the presentation mode. Most would like the questions to be in a bigger font but most found the test very user-friendly. A small minority felt that it was difficult to think 'straight' in front of the computer. These students much prefer the pen-and-paper test mode.

(ii) Test material : About 50% of the students felt that it should be an open book test to reduce the stress of taking the test. They feel that in this way it will enable them to revise even further.

(iii) Time constraint: Most students (80%) felt that there should not be a time limit in taking such a test as the countdown clock makes them rather anxious. The time limit set for the pilot-test was 30 minutes. Most students feel that if there must be a time limit, then one hour is better.

Other issues were raised:

Availability of resources: As the number of computer terminals is rather limited on our campus, students voiced their concern on their being able to take the test at a convenient time.

Implementation and Future Plans

As continuous assessment in courses become more and more prominent in our courses in National Institute of Education, there is more written work to be graded by instructors. In places where teaching assistants are not easily available for the grading process, this type of assessment would be a great help in reducing the time spent on marking. This model would provide a relief to instructors who might want to test students some basic concepts as the course progresses. Structuring the questions and familiarity of students using the computers will be important factors in the design of the test.

APPENDIX 1

A sample of Type 1 Question

TYPE 1 Question - Category 2: Question Q2

We are given that $P(A) = 0.4$ and $P(B) = 0.5$.
If we are told that A and B are mutually exclusive, is the following probability statement correct? Give reasons.

1T	P(both A and B) = 0
2F	P(either A or B) = 0

A sample of Type 2 Question

TYPE 2 Question - Category 1: Question Q1

The following graphical display shows the verbal scores of students that were taught using a newly designed teaching format.

Stem	Leaves
8	90
9	50 60
10	30 50 60 80 90
11	30 40 50 80 90
12	10 10 20 70
13	20 40
14	50

Based on the information provided, the median verbal scores is

A	1150
B*	1145
C	1140
D	1210

APPENDIX 2

The Fitness Test Set-Up Program