A Taste Of Mathematics

Aime–T–On les Mathématiques

Volume / Tome II

ALGEBRA — INTERMEDIATE METHODS

Revised Edition

Bruce Shawyer

Memorial University of Newfoundland

Publisher: Canadian Mathematical Society
Managing Editor: Graham Wright
Editor: Richard Nowakowski
Cover Design: Bruce Shawyer and Graham Wright
Typesetting: CMS CRUX with MAYHEM Office
Printing and Binding: The University of Toronto Press Inc.

Canadian Cataloguing in Publication Data

Shawyer, Bruce

Algebra — Intermediate Methods

Revised Edition

(A Taste Of Mathematics = Aime-T-On les Mathématiques ; v. 2)
Prefatory material in English and French.
ISBN 0–919558–11–9

1. Algebra. I. Canadian Mathematical Society. II. Title.
III. Series: A Taste Of Mathematics ; v. 2.

QA152.2.S53 1999 512.9 C99-900614-2

ISBN 0–919558–11–9

This is a PDF version of the original printed volume. There are changes in fonts and format that could result in different line and page breaks.

Ceci est une version PDF de l'ouvrage imprim. Il est possible que des changements de polices ou de format modifient les coupures de ligne ou de page.

A Taste Of Mathematics

Aime–T–On les Mathématiques

Volume / Tome II

ALGEBRA — INTERMEDIATE METHODS

Revised Edition

Bruce Shawyer

Memorial University of Newfoundland

The ATOM series

The booklets in the series, **A Taste of Mathematics**, are published by the Canadian Mathematical Society (CMS). They are designed as enrichment materials for high school students with an interest in and aptitude for mathematics. Some booklets in the series will also cover the materials useful for mathematical competitions at national and international levels.

La collection ATOM

Publiés par la Société mathématique du Canada (SMC), les livrets de la collection Aime-t-on les mathématiques (ATOM) sont destinés au perfectionnement des étudiants du cycle secondaire qui manifestent un intérêt et des aptitudes pour les mathématiques. Certains livrets de la collection ATOM servent également de matériel de préparation aux concours de mathématiques sur l'échiquier national et international.

Table of Contents

Foreward

This volume contains a selection of some of the basic algebra that is useful in solving problems at the senior high school level. Many of the problems in the booklet admit several approaches. Some worked examples are shown, but most are left to the ingenuity of the reader.

While I have tried to make the text as correct as possible, some mathematical and typographical errors might remain, for which I accept full responsibility. I would be grateful to any reader drawing my attention to errors as well as to alternative solutions. Also, I should like to express my sincere appreciation for the help given by Ed Barbeau in the preparation of this material.

It is the hope of the Canadian Mathematical Society that this collection may find its way to high school students who may have the talent, ambition and mathematical expertise to represent Canada internationally. Those who wish more problems can find further examples in:

1. The International Mathematical Talent Search (problems can be obtained from the author, or from the magazine *Mathematics & Informatics Quarterly*, subscriptions for which can be obtained (in the USA) by writing to Professor Susan Schwartz Wildstrom, 10300 Parkwood Drive, Kensington, MD USA 20895 `ssw@ umd5.umd.edu`, or (in Canada) to Professor Ed Barbeau, Department of Mathematics, University of Toronto, Toronto, ON Canada M5S 3G3 `barbeau@ math.utoronto.ca`);
2. The Skoliad Corner in the journal *Crux Mathematicorum with Mathematical Mayhem* (subscriptions can be obtained from the Canadian Mathematical Society, 577 King Edward, PO Box 450, Station A, Ottawa, ON, Canada K1N 6N5);
3. The book *The Canadian Mathematical Olympiad 1969–1993 L'Olympiade mathématique du Canada*, which contains the problems and solutions of the first twenty five Olympiads held in Canada (published by the Canadian Mathematical Society, 577 King Edward, PO Box 450, Station A, Ottawa, ON, Canada K1N 6N5);
4. The book *Five Hundred Mathematical Challenges*, by E.J. Barbeau, M.S. Klamkin & W.O.J. Moser (published by the Mathematical Association of America, 1529 Eighteenth Street NW, Washington, DC 20036, USA).

Bruce L.R. Shawyer
Department of Mathematics and Statistics
Memorial University of Newfoundland
St. John's, NF
Canada A1C 5S7

1 Introduction

The purpose of this booklet is to gather together some of the miscellaneous mathematics that is useful in problem solving, but is often missing from high school mathematics textbooks.

The topics are grouped under five main headings

1. **Mathematical Induction**.
2. **Series**.
3. **Binomial Coefficients**.
4. **Solution of Polynomial Equations**.
5. **Vectors and Matrices**.

A little explanation is necessary.

1. Mathematical Induction.

   This method is very useful for proving results about whole numbers. It is, in fact, basic to the number system that we use.

2. Series.

   Although infinite series is really a university level topic, a gentle introduction leads to some very useful results.

3. Binomial Coefficients.

   Binomial coefficients are everywhere. Many of the properties can be obtained using the results of the previous two section. However, many of the properties can be obtained using counting arguments. See, for example, the forthcoming ATOM booklet on Combinatorics.

4. Solution of Polynomial Equations.

   Almost everyone knows how to solve a quadratic equation. But few people know about the methods for solving cubic and quartic equations. Also included are properties of the roots of polynomials. This used to be a well studied topic, but is now almost "lost" from the curriculum.

5. Vectors and Matrices.

   The basic properties are followed by properties of determinants and properties of conic section. Again, these are "lost" branches of mathematics.

2 Mathematical Induction

A very important method of proof is the *Principle of Mathematical Induction*. This is used to prove results about the natural numbers, such as the fact that the sum of the first n natural numbers is $\dfrac{n(n+1)}{2}$.

We use functional notation: $P(n)$ means a result (proposition) for each natural number n. For example, the statement that the sum of the first n natural numbers is equal to $\dfrac{n(n+1)}{2}$ is written:

$$P(n): \quad \sum_{k=0}^{n} k \;=\; \frac{n(n+1)}{2}\,.$$

2.1 Induction

Proof by induction consists of three parts:

TEST: Find an appropriate starting number for the result (if it is not already given in the statement). Usually 1 is appropriate (sometimes 0 is better). For the purposes of this explanation, we shall take 1 as the starting number.

Test (check) the result for this number: is $P(1)$ correct?

If "YES", we proceed: if "NO", the suggested result is false.

STEP: We *assume* that the result is indeed true for some particular (general) natural number, say k. This is the *inductive hypothesis*. With this assumed result $P(k)$, we deduce the result for the next natural number $k+1$. In other words, we try to prove the implication

$$P(k) \implies P(k+1)\,.$$

Make sure that your logic includes the case $k = 1$.
If "YES", we proceed: if "NO", then we are in trouble with this method of proof.

PROOF. The *formal* proof is now:

$P(1)$ is true.
$P(k) \implies P(k+1)$ for $k \geq 1$.
Hence $P(1) \implies P(2)$,
and so $P(2) \implies P(3)$,
and so $P(3) \implies P(4)$,
and so on.
Therefore $P(n)$ is true for all natural numbers n.

The last part is usually omitted in practice.

Prove the following using mathematical induction:

Problem 1 $1 + 2 + \cdots + n \;=\; \dfrac{n(n+1)}{2}\,.$

Problem 2 $1^2 + 2^2 + \cdots + n^2 \;=\; \dfrac{n(n+1)(2n+1)}{6}\,.$

Problem 3 $1^3 + 2^3 + \cdots + n^3 \;=\; \dfrac{n^2(n+1)^2}{4}\,.$

Problem 4 $\displaystyle\sum_{k=0}^{n} \binom{n}{k} \alpha^k \beta^{n-k} \;=\; (\alpha+\beta)^n\,.$

Problem 5 $\displaystyle\sum_{k=1}^{n} \frac{1}{\sqrt{n}} \;<\; 2\sqrt{n}\,.$

Problem 6 $\dfrac{n^5}{5} + \dfrac{n^4}{2} + \dfrac{n^3}{3} - \dfrac{n}{30}$ *is an integer for* $n = 0,\, 1,\, 2,\, \ldots$.

Problem 7 *If* $x \in [0, \pi]$, *prove that* $|\sin(nx)| \;\le\; n\,\sin(x)$ *for* $n = 0,\, 1,\, 2,\, \ldots$.

Problem 8 *If* $a_k \;\ge\; 1$ *for all* k, *prove that*

$$2^n \left(\prod_{k=1}^{n} a_k \;+\; 1 \right) \;\ge\; 2 \prod_{k=1}^{n} (1 + a_k)$$

for $n = 1,\, 2,\, \ldots$.

2.2 Horses

Just to check that you really understand the Principle of Mathematical Induction, find the fallacy in the following "proof".

Proposition. Let $P(n)$ mean "in any set of n horses, all the horses are the same colour".

TEST. Consider a set consisting of exactly one horse. All the horses in it are the same colour (since there is only one).

STEP. Assume that $P(k)$ is true for some natural number k, that is, in any set of k horses, all the horses are the same colour. This is the inductive hypothesis.

Now, consider a set consisting of $k+1$ horses. We place each horse in a stable, in a row, and number the stables, $1,\, 2,\, \ldots,\, k,\, k+1$.

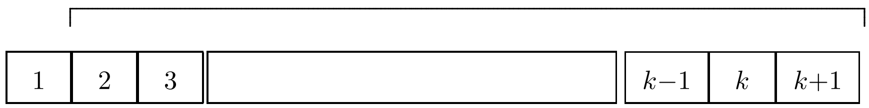
1	2	3		$k-1$	k	$k+1$

Consider the horses in stables numbered $1,\, 2,\, \ldots,\, k-1,\, k$.

1	2	3		$k-1$	k	$k+1$

This is a set of k horses, and so, by the inductive hypothesis, must consist of horses, all of the same colour.

Now, consider the horses in stables numbered 2, 3, ..., k, $k+1$.



This is a set of k horses, and so, by the inductive hypothesis, must consist of horses, all of the same colour.

By observing the overlap between the two sets, we see that all the horses in the set of $k+1$ of horses must be of the same colour. And so, we are done!

Clearly, this is nonsense! We know that all horses are not the same colour. The reader is asked to examine this "proof" and find out where it goes wrong. The explanation may be found on page 42.

2.3 Strong Induction

Strong induction (with is mathematically equivalent to induction) has an apparently stronger **STEP** condition:

STRONG STEP: We *assume* that the result is indeed true for all natural numbers, $1 \le j \le k$. This is the *strong inductive hypothesis*. With these assumed results $P(1)$, $P(2)$, ..., $P(k)$, we deduce the result for the next natural number $k+1$. In other words, we try to prove the implication

$$\{P(1), P(2), \ldots, P(k)\} \implies P(k+1)\ .$$

Again, make sure that your logic includes the case $k = 1$.

Problem 9 *Pick's theorem states that the area of a polygon, whose vertices have integer coordinates (that is, are lattice points) and whose sides do not cross, is given by*

$$I + \frac{B}{2} - 1\,,$$

where I and B are then numbers of interior and boundary lattice points respectively.

Prove Pick's Theorem for a triangle directly.

Use strong induction to prove Pick's theorem in general.

Problem 10 *Prove that every natural number may be written as the product of primes.*

Problem 11 *Assume Bertrand's theorem: for every $x > 1$, there is a prime number strictly between x and $2x$.*

Prove that every positive integer can be written as a sum of distinct primes. (You may take 1 to be a prime in this problem.)

Problem 12 *Show that every positive integer can be written as a sum of distinct Fibonacci numbers.* [1]

Problem 13 *For the Fibonacci numbers, show that* $F_{n+1}^2 + F_n^2 = F_{2n+1}$.

3 Series

We start with a given sequence $\{a_j\}$, and we use "sigma" notation for addition:

$$\sum_{j=k+1}^{n} a_j := a_{k+1} + a_{k+2} + \cdots + a_n .$$

We have "series" when we add up sequences. They may start with term 0 or with term 1 as is convenient. We define the partial sum of a series by

$$A_k := \sum_{j=0}^{k} a_j ,$$

so that

$$a_{k+1} + a_{k+2} + \cdots + a_n = \sum_{j=k+1}^{n} a_j = A_n - A_k .$$

We note that series may be added term by term, that is

$$\sum_{j=k+1}^{n} (\alpha a_j + \beta b_j) = \alpha \sum_{j=k+1}^{n} a_j + \beta \sum_{j=k+1}^{n} b_j .$$

3.1 Telescoping Series

A series is telescoping if the sequence of terms satisfies

$$a_j = f_j - f_{j-1} ,$$

and so we get

$$\sum_{j=1}^{n} a_j = \sum_{j=1}^{n} (f_j - f_{j-1}) = f_n - f_0 .$$

For example $\dfrac{1}{j(j+1)} = \dfrac{1}{j} - \dfrac{1}{j+1} = \dfrac{-1}{j+1} - \dfrac{-1}{j}$, so that $f_j = -\dfrac{1}{j+1}$, and

$$\sum_{j=1}^{n} \frac{1}{j(j+1)} = \sum_{j=1}^{n} \left(\frac{-1}{j+1} - \frac{-1}{j} \right) = \frac{-1}{n+1} - \frac{-1}{1} = 1 - \frac{1}{n+1} .$$

[1] Fibonacci numbers are given by

$$F_0 = 0 , F_1 = 1 , \text{for } n \geq 2 , \quad F_n = F_{n-1} + F_{n-2} .$$

Problem 14 *Evaluate* $\displaystyle\sum_{j=1}^{n} \frac{1}{j(j+1)(j+2)}$.

Problem 15 *Evaluate* $\displaystyle\sum_{j=1}^{n} \frac{1}{j(j+1)(j+2)(j+3)}$.

Generalize this!

Problem 16 *Evaluate* $\displaystyle\sum_{j=1}^{n} \frac{j}{j^4+j^2+1}$.

3.2 Sums of Powers of Natural Numbers

We are interested in evaluating $\displaystyle\sum_{j=1}^{n} j^{\alpha}$, where α is a natural number.

If $\alpha = 0$, then we have $\displaystyle\sum_{j=1}^{n} 1$. It is easy to see that the value of this is n.

The next case is $\alpha = 1$. To evaluate $\displaystyle\sum_{j=1}^{n} j$, we first note that $j^2 - (j-1)^2 \ = \ 2j - 1$. Therefore

$$\begin{array}{rcl} n^2 - (n-1)^2 & = & 2n - 1 \\ (n-1)^2 - (n-2)^2 & = & 2(n-1) - 1 \\ \vdots & & \\ 2^2 \ - \ 1^2 & = & 2(2) - 1 \\ 1^2 \ - \ 0^2 & = & 2(1) - 1\,. \end{array}$$

The sum of the left sides is $n^2 - 0 \ = \ n^2$ and the sum of the right sides is $\left(2\displaystyle\sum_{j=1}^{n} j\right) \ - \ n$ (did you get the last term?) This gives

$$\sum_{j=1}^{n} j \ = \ \frac{n(n+1)}{2}\,.$$

This technique can be used to generalize this result to larger values of α.

Problem 17 *Show that* $\displaystyle\sum_{j=1}^{n} j^2 \ = \ \frac{n(n+1)(2n+1)}{6}$.

Problem 18 *Show that* $\displaystyle\sum_{j=1}^{n} j^3 \;=\; \left(\frac{n(n+1)}{2}\right)^2.$

Problem 19 *Calculate a formula for* $\displaystyle\sum_{j=1}^{n} j^4$.

Problem 20 *Calculate a formula for* $\displaystyle\sum_{j=1}^{n} j^5$.

3.3 Arithmetic Series

A sequence $\{a_j\}$ is called *arithmetic* if $a_{j+1} - a_j = \delta$ (a fixed quantity, called the *common difference*) for all values of $j \geq 0$. Thus, we have

$$a_j \;=\; a_{j-1} + \delta \;=\; a_{j-2} + \delta + \delta \;=\; \cdots \;=\; a_0 + j\,\delta\,.$$

Thus

$$\begin{aligned}
\sum_{j=k+1}^{n} a_j \;&=\; \sum_{j=k+1}^{n} (a_0 + j\,\delta) \;=\; (n-k)a_0 + \delta \sum_{j=k+1}^{n} j \\
&=\; (n-k)a_0 + \delta \left(\sum_{j=0}^{n} j \;-\; \sum_{j=0}^{k} j \right) \\
&=\; (n-k)a_0 + \delta \left(\frac{n(n+1)}{2} \;-\; \frac{k(k+1)}{2} \right) . \qquad (1)
\end{aligned}$$

Another way to look at arithmetic series, is to write the sum down in both directions!

$$\begin{aligned}
\text{Sum} \;&=\; a_0 + a_1 + \cdots + a_{n-1} + a_n\,, \\
\text{Sum} \;&=\; a_n + a_{n-1} + \cdots + a_1 + a_0\,.
\end{aligned}$$

Adding gives

$$2 \times \;\text{Sum} \;=\; \big(a_0 + a_n\big) + \big(a_1 + a_{n-1}\big) + \cdots + \big(a_{n-1} + a_1\big) + \big(a_n + a_0\big)\;.$$

Note that

$$\begin{aligned}
\big(a_0 + a_n\big) \;&=\; \big(a_1 + a_{n-1}\big) \\
&=\; \cdots \;=\; \big(a_k + a_{n-k}\big) \\
&=\; \cdots \;=\; \big(a_n + a_0\big)\;.
\end{aligned}$$

Therefore

$$2 \times \;\text{Sum} \;=\; (n+1) \times \big(a_0 + a_n\big) \;=\; (n+1)\big(\text{First} \;+\; \text{Last}\big)\;,$$

so that

$$\text{Sum} \;=\; \frac{(n+1)\,(a_0+a_n)}{2} \;=\; \frac{(n+1)\,(\text{First}\,+\,\text{Last})}{2}\,.$$

Put into words, we see that the sum of an arithmetic series is the average of the first and last terms, multiplied by the number of terms.

3.4 Geometric Series

A sequence $\{a_j\}$ is called *geometric* if $a_{j+1} = a_j\, r$ (a fixed quantity, called the *common ratio*) for all values of $j \geq 0$. Thus, we have

$$a_j \;=\; a_{j-1}\cdot r \;=\; a_{j-2}\cdot r\cdot r \;=\; \cdots \;=\; a_0\cdot r^j\,.$$

For notational ease, we write a for a_0.

We start with the easiest case: $r = 1$. Here, each term is the same as the one before. So it is easy to add them up. In this case

$$\sum_{j=k+1}^{n} a_j \;=\; \sum_{j=k+1}^{n} a\,r^j \;=\; \sum_{j=k+1}^{n} a \;=\; a(n-k)\,.$$

In words, we have a times the number of terms.

For the remainder of this section, we shall assume that $r \neq 1$. Thus

$$\sum_{j=k+1}^{n} a_j \;=\; \sum_{j=k+1}^{n} a\,r^j \;=\; a\left(r^{k+1}+r^{k+2}+r^{k+3}+\cdots+r^{n-1}+r^n\right)\,.$$

Also

$$\begin{aligned} r\sum_{j=k+1}^{n} a\,r^j \;&=\; \sum_{j=k+1}^{n} a\,r^{j+1} \\ &=\; a\left(r^{k+2}+r^{k+3}+\cdots+r^{n-1}+r^n+r^{n+1}\right)\,. \end{aligned}$$

Subtracting these gives

$$(r-1)\sum_{j=k+1}^{n} a\,r^j \;=\; a\left(r^{n+1}-r^{k+1}\right)\,,$$

so that (remember that $r \neq 1$)

$$\sum_{j=k+1}^{n} a\,r^j \;=\; a\left(\frac{r^{n+1}-r^{k+1}}{r-1}\right)\,.$$

In the ATOM volume on trigonometry [2], you will see how to use geometric series to find

$$\sum_{j=k+1}^{n} \cos(j\theta) \qquad \text{and} \qquad \sum_{j=k+1}^{n} \sin(j\theta)\,.$$

3.5 Infinite Series

In order to do infinite series properly, you need a good grounding in the theory of limits. This is university material (usually second year). However, a few results are useful to know.

1. Infinite series either converge or diverge.
2. Those which diverge are in four categories:
   (a) those which diverge to $+\infty$ [3]; for example $\displaystyle\sum_{k=1}^{\infty} 2^k$;
   (b) those which diverge to $-\infty$; for example $\displaystyle\sum_{k=1}^{\infty} \left(-2^k\right)$;
   (c) those whose sequence of partial sums oscillates unboundedly; for example $\displaystyle\sum_{k=1}^{\infty} (-2)^k$; or $\displaystyle\sum_{k=1}^{\infty}(-1)^k 2^{\lfloor k/2 \rfloor}$, where $\lfloor x \rfloor$ means the greatest integer less that or equal to x ;
   (d) those whose sequence of partial sums oscillates boundedly; for example $\displaystyle\sum_{k=1}^{\infty} (-1)^k$.
3. Convergent series have a bounded sequence of partial sums.
4. A bounded sequence of partial sums does not guarantee convergence.
5. A series of positive terms either converges or diverges to $+\infty$.
6. For a series of positive terms, a bounded sequence of partial sums does guarantee convergence.
7. If $\sum a_k$ converges, then $\lim\limits_{k\to\infty} a_k \;=\; 0$.
8. If $\lim\limits_{k\to\infty} a_k \;=\; 0$, then $\sum a_k$ does not necessarily converge; for example $\sum \frac{1}{k}$.
9. An infinite arithmetic series cannot converge unless both $a_0 = 0$ and $\delta = 0$, in other words, we are adding up zeros! See (1).

[2] A forthcoming volume in this series.

[3] We emphasize that ∞ is **not** a number, but a symbol used to describe the behaviour of limits. Since the sum of an infinite series is in fact the limit of the sequence of partial sums, the symbol is also used to describe the sum of certain divergent infinite series.

Some examples of positive series are:

$$\begin{array}{llll}
\text{(See below)} & \displaystyle\sum_{k=1}^{\infty} r^k & \text{converges for} \quad |r| < 1\,; & \\
\text{(See below)} & \displaystyle\sum_{k=1}^{\infty} r^k & \text{diverges for} \quad |r| \geq 1\,; & \\
\text{(See below)} & \displaystyle\sum_{k=1}^{\infty} \frac{1}{k} & \text{diverges}\,; & (2) \\
\text{(See below)} & \displaystyle\sum_{k=1}^{\infty} \frac{1}{k^2} & \text{converges}\,; & (3) \\
 & \displaystyle\sum_{k=1}^{\infty} \frac{1}{k^p} & \text{converges for} \quad p > 1\,; & \\
 & \displaystyle\sum_{k=1}^{\infty} \frac{1}{k^p} & \text{diverges for} \quad p \leq 1\,; & \\
 & \displaystyle\sum_{k=1}^{\infty} \frac{1}{k\log(k)} & \text{diverges}\,; & \\
 & \displaystyle\sum_{k=1}^{\infty} \frac{1}{k\log(k)\log(\log(k))} & \text{diverges}\,. &
\end{array}$$

An infinite geometric series may converge or diverge. For it to converge, we require that $\lim\limits_{n\to\infty} r^n$ exists. This is true if $-1 < r \leq 1$.

If $r = 1$, the partial sum to the $n^{\mathbf{th}}$ term is an, which does not tend to a finite limit. Thus $\sum\limits_{k=1}^{\infty} a$ diverges.

If $|r| < 1$, then $\lim\limits_{n\to\infty} r^n \;=\; 0$, and this gives the sum to infinity

$$\sum_{k=0}^{\infty} r^n \;=\; \frac{1}{1-r}\,.$$

For (3), we note that, for each integer n, there is an integer m such that $2^m \leq n < 2^{m+1}$. Thus

$$
\begin{aligned}
\sum_{k=1}^{n} \frac{1}{k^2} &= 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} \\
&< 1 + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16}\right) \\
&\quad + \cdots + \left(\frac{1}{2^{2m}} + \frac{1}{2^{2m}} + \cdots + \frac{1}{2^{2m}}\right) \\
&= 1 + \frac{1}{2} + \frac{1}{4} + \cdots + \left(\frac{1}{2}\right)^m \;=\; 2\left(1 - \frac{1}{2^{m+1}}\right) \;<\; 2\,.
\end{aligned}
$$

Thus, the sequence of partial sums is bounded above. Since series of positive terms either converge or diverge to plus infinity, it follows that (3) converges. We do not know the value of the sum — only that it is less than or equal to 2. For those interested in numerical values, the actual sum is approximately 1.644934.

Problem 21 *Investigate the exact value of* $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^2}$.

For (2), using the same m as above, we have

$$
\begin{aligned}
\sum_{k=1}^{2^m} \frac{1}{k} &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \cdots + \frac{1}{2^m} \\
&> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \cdots + \left(\frac{1}{2^m} + \frac{1}{2^m} + \cdots + \frac{1}{2^m}\right) \\
&= 1 + \frac{1}{2} + \frac{1}{2} + \cdots + \frac{1}{2} \;=\; 1 + \frac{m}{2}\,.
\end{aligned}
$$

Thus, the sequence of partial sums increases without bound. Since series of positive terms either converge or diverge to plus infinity, it follows that (2) diverges.

We also note that the partial sums of $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k}$ grow slowly. If you try to sum this series on a computer, you may reach the conclusion that it converges, because the rate of growth will be too slow for the accuracy of the computer to detect. The partial sums grow like $\log(n)$. In fact

$$
\sum_{k=1}^{n} \frac{1}{n} \;=\; \gamma + \log(n) + \epsilon_n\,,
$$

where γ is Euler's constant, and ϵ_n is an error term that tends to zero and n tends to infinity. For those interested in numerical values, $\gamma \approx 0.577215665$.

Problem 22 *Investigate Euler's constant.*

With general series, the situation is a bit different. For example

$$\sum_{k=1}^{\infty} (-1)^n \, \frac{1}{n^p}$$

is a convergent series if $p > 0$. What we need here is the *alternating series test*, which states that

If $a_n \geq 0$ and $\lim_{n\to\infty} a_n \;=\; 0$, then

$$\sum_{k=1}^{\infty} (-1)^k \, a_k$$

is convergent.

Care is required about what is meant by the *sum* of such a series. For example $\displaystyle\sum_{k=1}^{\infty} (-1)^k \, \frac{1}{k} \;=\; \log(2)$. But this is only so if the terms are added up in the order given and not re-arranged.

4 Binomial Coefficients

4.1 Factorials

For every natural number k, we define "k factorial" [4], written as $k!$ as the running product of all the natural numbers from 1 up to k:

$$k! \;:=\; 1 \cdot 2 \cdot 3 \cdot \cdots \cdot (k-1) \cdot k \,.$$

This can also be defined inductively from

$$\begin{aligned} 1! \;&:=\; 1 \\ k! \;&:=\; k \,\times\, (k-1)! \end{aligned}$$

We also define $0! \;=\; 1$ for reasons of consistency that will become clear.

4.2 Binomial Coefficients

The Binomial Coefficient, $\binom{n}{k}$ usually read as "n choose k", is defined [5], initially for positive integers n, k with $n \geq k$, from:

$$\binom{n}{k} \;:=\; \frac{n!}{k!(n-k)!} \,.$$

This is equivalent to

$$\binom{n}{k} \;=\; \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} \,. \tag{4}$$

[4] Also read as "factorial k

[5] Older notation includes ${}^{n}C_k$.

We see that $\binom{n}{n} = \binom{n}{0} = 1$ and that $\binom{n}{k} > 0$.

We also note these coefficients in the Binomial Theorem:

If n is a positive integer, then

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k .$$

This can be proved by induction on n. The important property required in given in problem 25.

Because the expression (4) makes sense when n is not a positive integer, we can use it to extend the definitions of binomial coefficients as follows:

If k is an integer greater than (the integer) n, then $\binom{n}{k} = 0$. For example, $\binom{3}{5} = \dfrac{3 \cdot 2 \cdot 1 \cdot 0 \cdot (-1)}{5!} = 0\,.$

Let n be any real number. Note that $\binom{n}{k}$ is non-zero when n is not an integer and k is greater than n. For example,

$$\binom{-1/2}{2} = \frac{(-1/2)(-3/2)}{1 \cdot 2} = \frac{3}{8}\,.$$

4.3 Pascal's Triangle

One way to generate the binomial coefficients, is to use Pascal's Triangle.

$$\begin{array}{ccccccccccccccc}
 & & & & & & & 1 & & & & & & & \\
 & & & & & & 1 & & 1 & & & & & & \\
 & & & & & 1 & & 2 & & 1 & & & & & \\
 & & & & 1 & & 3 & & 3 & & 1 & & & & \\
 & & & 1 & & 4 & & 6 & & 4 & & 1 & & & \\
 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & \\
 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1 \\
. & . & . & . & . & . & . & . & . & . & . & . & . & . & .
\end{array}$$

We presume that you know how to generate this!

It is better to write this as an (infinite) matrix, putting in a 0 in every space to the right of the last 1. This matrix is considered to begin with row and column numbered 0.

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	.	.
$n=0$	1	0	.	.	.	.	.	.	.	.	.
$n=1$	1	1	0	.	.	.	.	.	.	.	.
$n=2$	1	2	1	0	.	.	.	.	.	.	.
$n=3$	1	3	3	1	0	.	.	.	.	.	.
$n=4$	1	4	6	4	1	0	.	.	.	.	.
$n=5$	1	5	10	10	5	1	0	.	.	.	.
$n=6$	1	6	15	20	15	6	1	0	.	.	.
$n=7$	1	7	21	35	35	21	7	1	0	.	.
.	.	.	.	.	.	.	.	.	.	.	.

The binomial coefficient $\binom{n}{k}$ occurs at position (n, k) of this matrix.

4.4 Properties

The following formulae should be proved:

Problem 23

$$k\binom{n}{k} = n\binom{n-1}{k-1}.$$

Problem 24

$$\binom{n}{k} = \binom{n}{n-k}.$$

Problem 25

$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}.$$

Problem 26

$$\binom{n}{k+1} = \binom{n}{k} \cdot \frac{n-k}{k+1}.$$

Problem 27

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}.$$

Problem 28

$$\binom{n+1}{k+1} = \binom{k}{k} + \binom{k+1}{k} + \cdots + \binom{n}{k}.$$

Problem 29

$$\binom{k+n+1}{n} = \binom{k}{0} + \binom{k+1}{1} + \binom{k+2}{2} + \cdots + \binom{k+n}{n}.$$

Problem 30

$$\binom{n}{1} = \binom{n}{n-1} = n.$$

Problem 31

$$2^n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n}.$$

Problem 32

$$0 = \binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \cdots + (-1)^n \binom{n}{n}.$$

Problem 33

$$2^{n-1} = \binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots.$$

Problem 34

$$2^{n-1} = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \cdots.$$

Problem 35

$$\binom{2n}{n} = \binom{n}{0}^2 + \binom{n}{1}^2 + \binom{n}{2}^2 + \cdots + \binom{n}{n}^2.$$

Problem 36

$$n\,2^{n-1} = \binom{n}{1} + 2\binom{n}{2} + 3\binom{n}{3} + \cdots + n\binom{n}{n}.$$

Problem 37

$$0 = \binom{n}{1} - 2\binom{n}{2} + 3\binom{n}{3} - \cdots + (-1)^{n-1} n \binom{n}{n}.$$

Problem 38

$$\binom{m+n}{k} = \sum_{j=0}^{k} \binom{n}{j}\binom{m}{k-j}.$$

There are several methods for proving these inequalities. For example, one could use mathematical induction (as developed earlier in this booklet), or the binomial theorem. They can also be obtained combinatorically. See, for example the ATOM booklet on Combinatorics [6].

[6] A planned ATOM booklet.

5 Solution of Polynomial Equations

5.1 Quadratic Equations

We begin with the quadratic equation $0 \;=\; a\,x^2 + b\,x + c$, where x, a, b and c are real numbers.

There are only certain conditions under which this equation has a solution: first, if $a = b = c = 0$, then any x satisfies the equation. If both $a = b = 0$ and $c \neq 0$, then the equation makes no sense. So we shall assume that at least one of a and b is a non-zero real number.

If $a = 0$, then $b \neq 0$ and so we can solve $0 \;=\; b\,x + c$ to get $x = \frac{c}{b}$. This is easy. So we shall assume henceforth that $a \neq 0$.

Since $a \neq 0$, we may divide throughout by a to get:

$$0 \;=\; x^2 + \frac{b}{a}\,x + \frac{c}{a}\,. \tag{5}$$

The standard technique for solution is to complete the square. Since
$(x+u)^2 \;=\; x^2 + 2xu + u^2$, we get

$$\begin{aligned} 0 \;&=\; x^2 + \frac{b}{a}\,x + \left(\frac{b}{2a}\right)^2 + \frac{c}{a} - \left(\frac{b}{2a}\right)^2 \\ &=\; \left(x + \frac{b}{2a}\right)^2 - \left(\frac{b^2 - 4ac}{4a^2}\right)\,. \end{aligned} \tag{6}$$

and this leads to the standard solution:

$$x \;=\; \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}\,.$$

The expression $b^2 - 4ac$, called the *Discriminant* and usually denoted by Δ[7], must be non-negative if the quadratic equation is to have a real number solution. If $\Delta = 0$, then the quadratic equation has only one solution: we have a perfect square. If $\Delta > 0$, then we have two distinct solutions.

If we allow x to be a complex number, in particular, when $\Delta < 0$, then the solutions occur in conjugate pairs, that is, of the form $p \pm i\,q$, where p, q are real numbers.

If we also allow the coefficients, a, b, c to be complex numbers, and allow complex solutions, then we always have two (distinct) solutions unless $\Delta = 0$, which we can interpret as *two identical solutions*.

The word **root** of an equation is another word for the solution of an equation. This is frequently used in the circumstances like those in the next paragraph.

[7]This is the upper case Greek letter, delta.

Suppose that α and β are the roots of (5). Then we have that

$$(x-\alpha)(x-\beta) \;=\; x^2 + \frac{b}{a}\,x + \frac{c}{a}\,.$$

By multiplying out the left side of this identity, we obtain that

$$\alpha+\beta \;=\; -\frac{b}{a} \qquad \text{and} \qquad \alpha\beta \;=\; \frac{c}{a}\,.$$

In words, this tells us that in the quadratic equation (5), the constant term is the **product** of the roots, and that the coefficient of x is **minus the sum** of the roots. For more on this, see the section entitled **symmetric functions**, later in this booklet.

Problem 39 *If α and β are the roots of $x^2+5x+7=0$, find a quadratic equation with roots $\frac{1}{\alpha}$ and $\frac{1}{\beta}$.*

Solution 39 We have $\alpha+\beta=-5$ and $\alpha\beta=7$. The equation we want is

$$\begin{aligned} 0 &= \left(x-\frac{1}{\alpha}\right)\left(x-\frac{1}{\beta}\right) \\ &= x^2-\left(\frac{1}{\alpha}+\frac{1}{\beta}\right)x+\frac{1}{\alpha\beta} \\ &= x^2-\left(\frac{\alpha+\beta}{\alpha\beta}\right)x+\frac{1}{\alpha\beta} \\ &= x^2-\left(\frac{-5}{7}\right)x+\frac{1}{7}\,, \end{aligned}$$

or $7x^2+5x+1=0$.

Problem 40 *If α and β are the roots of $x^2+5x+7=0$, find a quadratic equation with roots $\frac{1}{\alpha^2}$ and $\frac{1}{\beta^2}$.*

Solution 40 As before, we have $\alpha+\beta=-5$ and $\alpha\beta=7$. The equation we want is

$$\begin{aligned}
0 &= \left(x-\frac{1}{\alpha^2}\right)\left(x-\frac{1}{\beta^2}\right)\\
&= x^2-\left(\frac{1}{\alpha^2}+\frac{1}{\beta^2}\right)x+\frac{1}{(\alpha\beta)^2}\\
&= x^2-\left(\frac{\alpha^2+\beta^2}{(\alpha\beta)^2}\right)x+\frac{1}{(\alpha\beta)^2}\\
&= x^2-\left(\frac{\alpha^2+2\alpha\beta+\beta^2-2\alpha\beta}{(\alpha\beta)^2}\right)x+\frac{1}{(\alpha\beta)^2}\\
&= x^2-\left(\frac{(\alpha+\beta)^2-2\alpha\beta}{(\alpha\beta)^2}\right)x+\frac{1}{(\alpha\beta)^2}\\
&= x^2-\left(\frac{(-5)^2-2\times 7}{7^2}\right)x+\frac{1}{7^2}\\
&= x^2-\frac{11}{49}x+\frac{1}{49}\,,
\end{aligned}$$

or $49x^2-11x+1=0$.

5.2 Cubic Equations

We start with a consideration of the roots of $0 = A\,x^3+B\,x^2+C\,x+D$ with $A\neq 0$. First, as in the previous section, we write this as

$$0 = x^3+\frac{B}{A}\,x^2+\frac{C}{A}\,x+\frac{D}{A}. \tag{7}$$

Suppose that the roots are α, β and γ. Then we have that

$$(x-\alpha)(x-\beta)(x-\gamma) = x^3-(\alpha+\beta+\gamma)x^2+(\alpha\beta+\beta\gamma+\gamma\alpha)x-\alpha\beta\gamma\,.$$

Comparing this with the equation above, we obtain that

$$\alpha+\beta+\gamma=-\frac{B}{A}\,,\quad \alpha\beta+\beta\gamma+\gamma\alpha=\frac{C}{A}\quad\text{and}\quad \alpha\beta\gamma=-\frac{D}{A}\,.$$

In words, this tells us that in the cubic equation (7), the constant term is **minus the product** of the roots, that the coefficient of x is the **sum of the products** of the roots **taken in pairs**, and that the coefficient of x is **minus the product** of the roots. For more on this, see the section entitled **symmetric functions**, later in this booklet.

Problem 41 *If α, β and γ are the roots of $x^3-2x^2+3x-4=0$, find an equation with roots α^2, β^2 and γ^2.*

Solution 41 We have $\alpha+\beta+\gamma=2$, $\alpha\beta+\beta\gamma+\gamma\alpha=3$ and $\alpha\beta\gamma=4$. The equation we seek is

$$\begin{aligned}
0 &= (x-\alpha^2)(x-\beta^2)(x-\gamma^2)\\
&= x^3-(\alpha^2+\beta^2+\gamma^2)x^2+(\alpha^2\beta^2+\beta^2\gamma^2+\gamma^2\alpha^2)x-\alpha^2\beta^2\gamma^2\,.
\end{aligned}$$

Here, we will make use of

$$\begin{aligned}
(\alpha+\beta+\gamma)^2 &= \alpha^2+\beta^2+\gamma^2 \;+\; 2(\alpha\beta+\beta\gamma+\gamma\alpha)\,, \\
(\alpha\beta+\beta\gamma+\gamma\alpha)^2 &= \alpha^2\beta^2+\beta^2\gamma^2+\gamma^2\alpha^2 \;+\; 2\alpha^2\beta\gamma+2\alpha\beta^2\gamma+2\alpha\beta\gamma^2 \\
&= \alpha^2\beta^2+\beta^2\gamma^2+\gamma^2\alpha^2 \;+\; 2\alpha\beta\gamma(\alpha+\beta+\gamma)
\end{aligned}$$

to give us

$$\begin{aligned}
\alpha^2+\beta^2+\gamma^2 &= (\alpha+\beta+\gamma)^2-2(\alpha\beta+\beta\gamma+\gamma\alpha) \\
&= 2^2-2(3) \;=\; -\,2\,, \\
\alpha^2\beta^2+\beta^2\gamma^2+\gamma^2\alpha^2 &= (\alpha\beta+\beta\gamma+\gamma\alpha)^2-2\alpha\beta\gamma(\alpha+\beta+\gamma) \\
&= 3^2-2(4)(2) \;=\; -\,7\,, \\
\alpha^2\beta^2\gamma^2 &= (\alpha\beta\gamma)^2 \;=\; 4^2 \;=\; 16\,.
\end{aligned}$$

Hence, the required equation is

$$0 \;=\; x^3+2x^2-7x-16\,.$$

We now look for a general solution.

The general form $0 \;=\; A\,x^3+B\,x^2+C\,x+D$ with $A \neq 0$ is usually reduced to the *standard form*

$$0 \;=\; x^3+a\,x^2+b\,x+c\,.$$

The substitution

$$x = y-\frac{a}{3}$$

leads to the *reduced form*

$$0 \;=\; y^3+p\,y+q\,.$$

This can be solved using **Cardano's Formula**.

Here is the technique. Let $y = u+v$. (This may seem to be complicating matters by having two variables instead of one, but it actually give more room to manoeuvre.) The reduced equation becomes

$$0 \;=\; u^3+3uv(u+v)+v^3+py+q \;=\; u^3+v^3+(3uv+p)y+q\,.$$

We have one equation in two unknowns, so there is not a unique solution. We thus have the freedom to impose the condition: $3uv+p=0$. Now, we have a system of two equations in two unknowns:

$$\begin{aligned}
u^3+v^3 &= -q\,, \\
u^3v^3 &= -\frac{p^3}{27}\,.
\end{aligned}$$

Thinking now about the roots of quadratic equations, we see that u^3 and v^3 are the roots of the quadratic equation

$$0 \;=\; x^2+q\,x-\frac{p^3}{27}\,.$$

This leads to the solutions

$$u = \left(-\frac{q}{2} + \sqrt{\Delta}\right)^{\frac{1}{3}}, \tag{8}$$

$$v = \left(-\frac{q}{2} - \sqrt{\Delta}\right)^{\frac{1}{3}}, \tag{9}$$

where

$$\Delta = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3, \tag{10}$$

and the cube roots chosen so that $uv = -\frac{p}{3}$. Let

$$y_1 = u + v\,, \tag{11}$$

$$y_2 = -\left(\frac{u+v}{2}\right) + i\sqrt{3}\left(\frac{u-v}{2}\right), \tag{12}$$

$$y_3 = -\left(\frac{u+v}{2}\right) - i\sqrt{3}\left(\frac{u-v}{2}\right). \tag{13}$$

Now, we can check that y_1, y_2 and y_3 are the three solutions of the reduced equation.
There are three cases.

If $\Delta > 0$, then we get one real solution and two conjugate complex solutions.

If $\Delta = 0$, then we get three real solutions (including a double root).

If $\Delta < 0$, then we get three real solutions, which can be found from trigonometry (see the ATOM volume on trigonometry [8] if necessary):

$$y_1 = 2\cos\left(\frac{\phi}{3}\right)\sqrt{\frac{|p|}{3}}\,; \tag{14}$$

$$y_2 = 2\cos\left(\frac{\phi}{3} - \frac{2\pi}{3}\right)\sqrt{\frac{|p|}{3}}\,; \tag{15}$$

$$y_3 = 2\cos\left(\frac{\phi}{3} + \frac{2\pi}{3}\right)\sqrt{\frac{|p|}{3}}\,; \tag{16}$$

where $\phi = \cos^{-1}\left(\dfrac{-\frac{q}{2}}{\sqrt{\left(\frac{|p|}{3}\right)^3}}\right)$.

[8] A forthcoming volume in this series.

5.3 Quartic Equations

Again, we could start with a consideration of the roots, but we will leave that until the next section.

The general form $0 \;=\; A\,x^4 + B\,x^3 + C\,x^2 + D\,x + E$ with $A \neq 0$ is usually reduced to the *standard form*

$$0 \;=\; x^4 + a\,x^3 + b\,x^2 + c\,x + d\,.$$

The substitution

$$x = y - \frac{a}{4}$$

leads to the *reduced form*

$$0 \;=\; y^4 + p\,y^2 + q\,y + r\,.$$

but, for all u, this is equivalent to

$$\begin{aligned} 0 &= \bigl(y^4 + y^2\,u + u^2/4\bigr) \;+\; \bigl(p\,y^2 + q\,y + r - u^2/4 - y^2\,u\bigr) \\ &= \bigl(y^2 + u/2\bigr)^2 \;-\; \bigl((u-p)\,y^2 - q\,y + (u^2/4 - r)\bigr)\,. \end{aligned} \tag{17}$$

The first term is a perfect square, say P^2 with $P = y^2 + u/2$. The second term is also a perfect square Q^2 for those value of u such that

$$q^2 \;=\; 4(u-p)(u^2/4 - r)\,. \tag{18}$$

(This is obtained by setting the discriminant of the quadratic in y equal to zero.)

Consider (18) as a cubic in u; that is

$$H(u) \;=\; u^3 - pu^2 - 4ru + (4pr - q^2) \;=\; 0\,.$$

We note that there is a root of H that is greater than p. This follows since $H(p) = -q^2$, since $H(u) \to \infty$ and $u \to \infty$, and since H is continuous.

So, this cubic in u can be solved as in the previous section and has at least one root, say u_1, of (18) satisfies $u_1 \geq p$. Substituting back into (17) gives the form

$$\bigl(y^2 + u_1/2 + Q\bigr)\,\bigl(y^2 + u_1/2 - Q\bigr) \;=\; 0\,,$$

where

$$Q = \alpha y - \beta\,, \quad \text{where} \quad \alpha = \sqrt{u_1 - p}\,, \quad \beta = \frac{q}{2\alpha}\,. \tag{19}$$

Thus, the desired roots of the monic [9] quartic are the roots of the two quadratic factors of (19). These quadratics have real coefficients if the original quartic has real coefficients. Thus a quartic has an even number of real roots!

Historical Note

The solution of the quadratic was known to both ancient Hindu and Greek mathematicians. The solutions of the cubic and quartic are due to Italian mathematicians, Scipio del Ferro (1515) and Ferrari (1545).

[9] A polynomial is called **monic** if the coefficient of the highest power of the variable is 1.

5.4 Higher Degree Equations

There are no general algorithms for solving fifth and higher degree polynomial equations using the methods described above. This was shown in the nineteenth century by Abel and Galois. Despite this, there is the **Fundamental Theorem of Algebra** which states that every polynomial equation of degree n has a (complex) solution, and hence, including solutions according to their multiplicity, n solutions. The proof of this apparently simple theorem requires some techniques of advanced mathematics, and is usually not proved until second or third year university courses.

However, there are several methods for approximating the roots of polynomials.

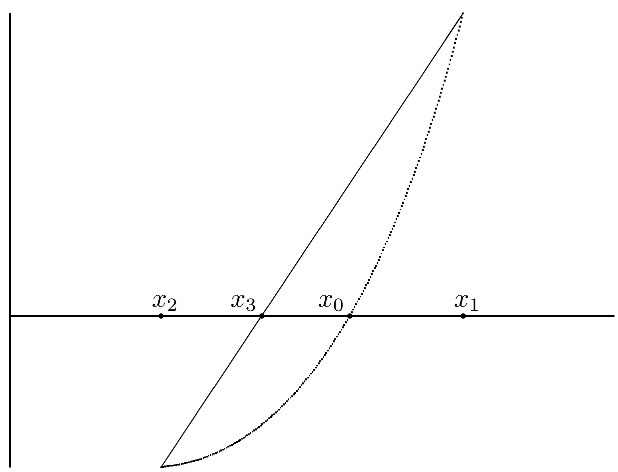
First, we note the following:

If the monic polynomial equation

$$p(x) \ = \ x^n + a_{n-1}\,x^{n-1} + a_{n-2}\,x^{n-2} + \cdots + a_1\,x + a_0 \ = \ 0$$

has a root between the values x_1 and x_2, that is, if $p(x_1)$ and $p(x_2)$ have opposite signs, then x_3 is an approximate value to a root of $p(x) = 0$, where

$$x_3 \ = \ x_1 \ - \ \frac{(x_2 - x_1)p(x_1)}{p(x_2) - p(x_1)}\,.$$



We have no guarantee that x_3 is a better approximation that either x_1 or x_2. This method is known as linear interpolation, and sometimes as **Regula falsi** (rule of false position).

However, all is not lost. If $p(x_3) = 0$, then we are done. Otherwise, $p(x_3)$ will have opposite sign to one of $p(x_1)$ or $p(x_2)$, so that we can proceed again.

Another procedure, known as **Newton's Method**, requires a knowledge of Calculus, and applies (as does Regula falsi) to more than polynomials.

If x_1 is an approximation to a root of the polynomial equation

$$p(x) \;=\; x^n + a_{n-1}\,x^{n-1} + a_{n-2}\,x^{n-2} + \cdots + a_1\,x + a_0 \;=\; 0\,,$$

then a better approximation may be obtained from (it usually works, but not always)

$$x_2 \;=\; x_1 \;-\; \frac{p(x_1)}{p'(x_1)}\,,$$

provided that $p'(x_1) \neq 0$.

As with Regula falsi, this may be repeated for better approximations.

Newton's method, usually works if you start close to a suspected root. For polynomials, and when there is no other root close by, the procedure works well, giving rapid convergence to the root. However, considerable care is necessary in general. Real difficulties arise when roots are close together.

Newton's method is especially useful for estimating $n^{\mathbf{th}}$ roots of numbers when no calculator is available (as in competitions like the IMO).

Let $p(x) \;=\; x^n - A$ where A is a positive real.

Then $p'(x) \;=\; nx^{n-1}$ so that the Newton approximation equation becomes

$$x_{k+1} \;=\; x_k \;-\; \frac{(x_k)^n - A}{n(x_k)^{n-1}}\,.$$

For example, if $n = 3$ and $A = 10$, this is

$$x_{k+1} \;=\; x_k \;-\; \frac{(x_k)^3 - 10}{3(x_k)^2}\,.$$

We now choose (almost arbitrarily) a first approximation, say $x_1 = 2$. Thus

$$x_2 \;=\; 2 - \frac{(2)^3 - 10}{3(2)^2} \;=\; 2 - \frac{-2}{12} \;=\; 2 + \frac{1}{6} \;=\; \frac{13}{6}\,.$$

We now repeat the process:

$$x_3 \;=\; \frac{13}{6} - \frac{(\frac{13}{6})^3 - 10}{3(\frac{13}{6})^2} \;=\; \frac{3277}{1521} \;\approx\; 2.1545\,.$$

Note that $10^{1/3} \approx 2.1544$

Also note that, if we start with a rational number, this process will always yield a rational approximation to a root of a polynomial equation with rational coefficients.

5.5 Symmetric Functions

We note that the monic polynomial can be written as

$$\begin{aligned} p(x) \;&=\; x^n + a_{n-1}\,x^{n-1} + a_{n-2}\,x^{n-2} + \cdots + a_1\,x + a_0 \\ &=\; (x - x_1)(x - x_2)(x - x_3)\cdots(x - x_n)\,, \end{aligned}$$

where $x_1,\, x_2,\, x_3, \ldots x_n$ are the roots of $p(x) = 0$.

The *elementary symmetric functions* S_k are defined from Vieta's Theorem

$$\begin{aligned} S_1 &= x_1 + x_2 + \cdots + x_n \\ &= -\,a_{n-1}\,; && (20) \\ S_2 &= x_1x_2 + x_1x_3 + \cdots + x_2x_3 + x_2x_4 + \cdots \\ &= \quad +x_3x_4 + x_3x_5 + \cdots + x_{n-1}x_n \\ &= +\,a_{n-2}\,; && (21) \\ S_3 &= x_1x_2x_3 + x_1x_2x_4 + \cdots + x_2x_3x_4 + x_2x_3x_5 + \cdots \\ &= \quad +x_3x_4x_5 + x_3x_4x_6 + \cdots + x_{n-2}x_{n-1}x_n \\ &= -\,a_{n-3}\,; && (22) \\ &\vdots \\ S_n &= x_1x_2x_3\cdots x_n \\ &= (-1)^n\, a_0\,. && (23) \end{aligned}$$

In the problems that follow, you will be making use of the known result that any symmetric polynomial in the roots can be expressed as a polynomial of the elementary symmetric functions.

Problem 42 *If $x_1,\, x_2,\, x_3$ are the roots of $x^3 + 9x^2 + 24x + 18 = 0$, prove that $\sum_{k=1}^{3} x_k^2 = 33$ and $\sum_{k=1}^{3} x_k^3 = -135$.*

Problem 43 *If $x_1,\, x_2,\, x_3$ are the roots of $x^3 - ax^2 + bx - c = 0$, prove that*

$$(x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_1)^2 \;=\; 2a^2 - 6b\,.$$

Problem 44 *If $x_1,\, x_2,\, x_3$ are the roots of $x^3 + ax^2 + bx + c = 0$, find the values of*

(i) $\sum_{k=1}^{3} x_1^2x_2^2 + x_2^2x_3^2 + x_3^2x_1^2,$

(ii) $\sum_{k=1}^{3} x_1(x_2^2 - x_3^2) + x_2(x_3^2 - x_1^2) + x_3(x_1^2 - x_2^2).$

Problem 45 *If $x_1,\, x_2,\, x_3$ are the roots of $x^3 + px + q = 0$, find equations whose roots are*

(i) $x_1^2,\, x_2^2,\, x_3^2,$

(ii) $\dfrac{1}{x_1},\, \dfrac{1}{x_2},\, \dfrac{1}{x_3},$

(iii) $\dfrac{1}{x_1}+\dfrac{1}{x_2},\ \dfrac{1}{x_2}+\dfrac{1}{x_3},\ \dfrac{1}{x_3}+\dfrac{1}{x_1},$

(iv) $\dfrac{x_1}{x_2x_3},\ \dfrac{x_2}{x_3x_1},\ \dfrac{x_3}{x_1x_2}.$

Problem 46 *If x_1, x_2, x_3, x_4 are the roots of $x^4+px^3+qx^2+rx+s=0$, find the values of*

(i) $\sum x_1^2x_2x_3$*, where is sum is over all terms of the given form.*

(ii) $\sum x_1^3x_2$*, where is sum is over all terms of the given form.*

Problem 47 *If x_1, x_2, x_3, the roots of $px^3+qx^2+rx+s=0$, are in geometric progression, prove that $pr^3=q^3s$.*

Problem 48 *If x_1, x_2, x_3, the roots of $x^3+px^2+qx+r=0$, are in arithmetic progression, prove that $2p^3=9(pq-3r)$.*

Problem 49 *If x_1, x_2, x_3 are the roots of $x^3+2x^2-36x-72=0$ and $\dfrac{1}{x_1}+\dfrac{1}{x_2}=\dfrac{2}{x_3}$, find the values of x_1, x_2 and x_3.*

5.6 Iterative Methods

We have already mentioned Newton's method. We conclude this section with an iterative method that was advertised in the 1970's when hand calculators became readily available.

> Suppose that $P(x) = x^n-a_{n-1}x^{n-1}-\cdots-a_1x-a_0$ with $a_0\neq 0$ and $n\geq 1$. This means that $x=0$ is not a solution. To find a solution, we must solve $0 = x^n-a_{n-1}x^{n-1}-\cdots-a_1x-a_0$.
>
> Write this as $x^n = a_{n-1}x^{n-1}+\cdots+a_1x+a_0$ and divide both sides of the equation by x^{n-1}.
>
> So we have $x = a_{n-1}+\dfrac{a_{n-2}}{x}+\cdots+\dfrac{a_1}{x^{n-2}}+\dfrac{a_0}{x^{n-1}}$.
>
> On the left side, replace "x" by "x_{k+1}", and on the right side, replace "x" by "x_k".
>
> This leads to the iteration
>
> $$x_{k+1} = a_{n-1}+\frac{a_{n-2}}{x_k}+\cdots+\frac{a_1}{{}_kx^{n-2}}+\frac{a_0}{x_k^{n-1}}\,. \tag{24}$$
>
> If we start with a "guess", x_0, we generate a sequence $\{x_k\}$.
>
> If we take limits as $k\to\infty$ on both sides, provided (24) converges, we will have generated a solution to $P(x)=0$.

This is all very appealing, but, unfortunately, in general, we cannot conclude that the sequence converges. It can lead to "strange attractors" or "chaos". It is suggested that the reader try the following example.

Problem 50 *Let* $P(x) \ = \ (x-1)(x-2)(x-3) \ = \ x^3 - 6x^2 + 11x - 6.$

Construct the iteration (24), and investigate the behaviour of the sequence $\{x_k\}$ *with the following choices of* x_0*:*

1.	$x_0 = 0.1.$	*2.*	$x_0 = 0.5.$	*3.*	$x_0 = 0.9.$
4.	$x_0 = 1.1.$	*5.*	$x_0 = 1.5.$	*6.*	$x_0 = 1.9.$
7.	$x_0 = 2.1.$	*8.*	$x_0 = 2.5.$	*7.*	$x_0 = 2.9.$
10.	$x_0 = 3.1.$	*11.*	$x_0 = 3.5.$	*12.*	$x_0 = 3.9.$

It will also be instructive to try this example too.

Problem 51 *Let* $P(x) \ = \ (x-\alpha)(x-\beta) \ = \ x^2 - 2ax - b$ *with* a *and* b *real.*

1. *Suppose that* α *and* β *are real and distinct with* $|\alpha| > |\beta| > 0$.

   Construct the iteration (24), and investigate the behaviour of the sequence $\{x_k\}$ *with the following choices of* x_0*:*

   1. $x_0 = \frac{\beta}{2}$. *2.* $x_0 = \frac{\alpha+\beta}{3}$. *3.* $x_0 = 2\alpha$.

2. *Suppose that* $\alpha = \beta$ *is real and non-zero.*

   Construct the iteration (24), and investigate the behaviour of the sequence $\{x_k\}$ *with the following choices of* x_0*:*

   1. $x_0 = \frac{\alpha}{2}$. *2.* $x_0 = 2\alpha$.

3. *Suppose that* α *and* β *are complex.*

   Construct the iteration (24), and investigate the behaviour of the sequence $\{x_k\}$ *with the following choices of* x_0*:*

   1. $x_0 = 1$. *2.* $x_0 = \frac{\alpha}{2}$. *3.* $x_0 = 2\beta$.

6 Vectors and Matrices

6.1 Vectors

There is a commonly held notion, especially in the physical sciences, that a *vector* is a thing that points! This is because vectors are often denoted by arrows! However, this is only a loose way of describing the important properties of a vector.

A vector is best defined as an ordered set of numbers (coordinates). For example, in 2-space (The Euclidean plane), we write

$$\overline{x} \ = \ (x_1, x_2)\,,$$

and in n-space,

$$\overline{x} \ = \ (x_1, x_2, \ldots, x_n)\,.$$

6.2 Properties of Vectors

6.3 Addition

The sum of two vectors $\overline{x}$, $\overline{y}$ is defined by

$$\overline{x} + \overline{y} \;:=\; (x_1 + y_1,\, x_2 + y_2,\, \ldots,\, x_n + y_n) \;=\; \overline{x + y}\,.$$

Note that it is necessary for both vectors to have the same number of coordinates.

6.4 Multiplication by a Scalar

The product of a real number (scalar), λ, with a vector, $\overline{x}$, is defined by

$$\lambda\,\overline{x} \;:=\; (\lambda\, x_1,\, \lambda\, x_2,\, \ldots,\, \lambda\, x_n)\,.$$

6.5 Scalar or Dot Multiplication

The scalar product of two vectors $\overline{x}, \overline{y}$ is defined by

$$\overline{x} \cdot \overline{y} \;:=\; \sum_{k=1}^{n} x_k y_k\,.$$

Note that it is necessary for both vectors to have the same number of coordinates.

This leads to the *length* or *norm* of a vector being defined by

$$|\overline{x}| \;:=\; \sqrt{\overline{x} \cdot \overline{x}}\,.$$

Geometrically, the scalar product is then

$$\overline{x} \cdot \overline{y} \;=\; |\overline{x}|\,|\overline{y}|\,\cos(\theta)\,,$$

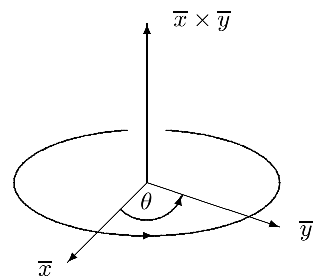
where θ is the angle between the vectors $\overline{x}$ and $\overline{y}$.

6.6 Vector or Cross Product

This is restricted to 3-dimensional space. The vector product of two vectors $\overline{x}$, $\overline{y}$ is defined by

$$\overline{x} \times \overline{y} \;:=\; (x_2 y_3 - x_3 y_2, x_3 y_1 - x_1 y_3, x_1 y_2 - x_2 y_1)\,.$$

Geometrically, the vector product is then a **vector**, orthogonal to both $\overline{x}$ and $\overline{y}$, with norm $|\overline{x}|\,|\overline{y}|\,\sin(\theta)$, directed according to the right hand screw convention, and where θ is the angle between the vectors $\overline{x}$ and $\overline{y}$. (The right hand screw convention is illustrated by the following diagram.)



Problem 52 *Prove that* $\overline{x} \times (\overline{y} \times \overline{z}) = (\overline{x}.\overline{z})\overline{y} - (\overline{x}.\overline{y})\overline{z}$.

Problem 53 *Prove that* $\overline{x} \times (\overline{y} \times \overline{z}) + \overline{y} \times (\overline{z} \times \overline{x}) + \overline{z} \times (\overline{x} \times \overline{y}) \; = \; \overline{0}$.

6.7 Triple Scalar Product

This is restricted to 3-dimensional space. The triple scalar product of $\overline{y} \times \overline{z}$ with $\overline{x}$ is called the triple scalar product, and is defined by $[\overline{x}, \overline{y}, \overline{z}] = \overline{x} \cdot (\overline{y} \times \overline{z})$ gives the (signed) volume of the parallelepiped with sides $\overline{x}$, $\overline{y}$, $\overline{z}$ at a vertex.

Problem 54 *Show that* $\overline{x} \cdot (\overline{y} \times \overline{x}) \; = \; 0$.

Problem 55 *Show that* $\overline{x} \cdot (\overline{y} \times \overline{z}) \; = \; \overline{y} \cdot (\overline{z} \times \overline{x})$.

Problem 56 *Show that* $(\overline{x} \times \overline{y}) \cdot (\overline{z} \times \overline{w}) = [\overline{x}, \overline{y}, \overline{z} \times \overline{w}] = [\overline{y}, \overline{z} \times \overline{w}, \overline{x}]$.

Problem 57 *Show that* $[\overline{x}, \overline{y}, \overline{z}]\, \overline{w} = [\overline{w}, \overline{y}, \overline{z}]\, \overline{x} \, + \, [\overline{x}, \overline{w}, \overline{z}]\, \overline{y} \, + \, [\overline{x}, \overline{y}, \overline{w}]\, \overline{z}$.

This means that any vector $\overline{w}$ may be expresses as a linear combination of any three given non-coplanar vectors, $\overline{x}$, $\overline{y}$ and $\overline{z}$.

6.8 Matrices

A *Matrix* is a rectangular array of elements. We shall refer to the *rows* and *columns* of this matrix.

Suppose that the matrix has n rows and k columns. We then write

$$A \; = \; (a_{i,j})_{1 \le i \le n \, , \, 1 \le j \le k} \; ,$$

where $a_{i,j}$ is the element in the $i^{\mathbf{th}}$ row and the $j^{\mathbf{th}}$ column. Written in full, this is

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} & \cdots & a_{1,k} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} & \cdots & a_{2,k} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} & \cdots & a_{3,k} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & a_{n,4} & \cdots & a_{n,k} \end{pmatrix} .$$

These give useful ways of describing other mathematical quantities.

Note that a matrix with one row is a vector. This is often called a row vector to distinguish it from a matrix with one column, which is known as a column vector.

The matrix obtained by interchanging the rows and columns of a matrix, A, is known as the **transpose** of the given matrix, and is written as A^T.

6.9 Properties of Matrices

6.10 Multiplication by a Scalar

The product of a scalar, λ, with a matrix, A, is given by

$$\lambda A \ := \ (\lambda a_{i,j})_{1\le i\le n\,,\,1\le j\le k}\ .$$

6.11 Multiplication of Vectors and Matrices

The scalar product of two vectors can be written as a product of a row vector with a column vector thus:

$$\overline{x}\cdot\overline{y} \ = \ (x_1, x_2, \cdots, x_n)\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \ = \ \sum_{k=1}^{n} x_k y_k\ .$$

This gives the basis for multiplying a row vector by a matrix. Consider the matrix a consisting of a set of column vectors (of course, the number of coordinates in the row vector must be the same as the number of columns in the matrix). This gives, for example

$$\overline{x}\,A \ = \ (x_1, x_2, \ldots, x_n)\begin{pmatrix} y_1 & z_1 \\ y_2 & z_2 \\ \vdots & \vdots \\ y_n & z_n \end{pmatrix} \ = \ \left(\sum_{i=1}^{n} x_i y_i\,,\ \sum_{i=1}^{n} x_i z_i\right)\ .$$

and, in general,

$$\begin{aligned} \overline{x}\,A \ &= \ (x_1, x_2, \ldots, x_n)\begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} & \cdots & a_{1,k} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} & \cdots & a_{2,k} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} & \cdots & a_{3,k} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & a_{n,4} & \cdots & a_{n,k} \end{pmatrix} \\ &= \ \left(\sum_{i=1}^{n} x_i a_{i,1},\ \sum_{i=1}^{n} x_i a_{i,2}, \ldots, \sum_{i=1}^{n} x_i a_{i,k}\right)\ . \end{aligned}$$

or, more compactly

$$\overline{x}\,A \;=\; \left(\sum_{i=1}^{n} x_i a_{i,j}\right)_{1\le j\le k}\,.$$

Similarly, we can multiply a matrix by a column vector (of course, the number of coordinates in the columns vector must be the same as the number of rows in the matrix). This gives,

$$A\,\overline{x} \;=\; \left(\sum_{j=1}^{k} a_{i,j} x_j\right)_{1\le i\le n}\,.$$

We are now ready to multiply two matrices! This first condition is that the number of columns in the left matrix must equal the number of rows in the right matrix. So, let

$$A \;=\; (a_{i,j})_{1\le i\le n,\;1\le j\le m} \qquad \text{and} \qquad B \;=\; (b_{j,k})_{1\le j\le m,\;1\le k\le p}\,.$$

and we get that

$$A\,B \;:=\; \left(\sum_{j=1}^{m} a_{i,j} b_{j,k}\right)_{1\le i\le n,\;1\le k\le p}\,.$$

Note that even if we can calculate $A\,B$, it is not necessary that $B\,A$ makes sense. And if it does, with A being $n\times k$ and so with B being $k\times n$, we have that $A\,B$ is $n\times n$ and $B\,A$ is $k\times k$. The only possibility for equality is if $n=k$, that is, for *square* matrices. But even then we shall find that equality is not necessary. For example

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \;=\; \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

whereas

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \;=\; \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

If a pair of matrices satisfy $A\,B = B\,A$, then we say that they *commute*. The pair of matrices given above do not commute.

6.12 Square Matrices

For square matrices, it is easy to show that

$$(A\,B)\,C \;=\; A\,(B\,C)\,.$$

A special square matrix is the one with 1 in each position where $i=j$ (the main diagonal), and 0 in every other position. For example

$$I \;:=\; \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This is known as the *Identity* matrix, for it is easy to show that, for any square matrix,

$$A\,I \;=\; I\,A \;=\; A\,.$$

We now ask if, given a square matrix A, there exists a square matrix, say B, such that $A\,B \;=\; I$. If such a matrix exists, we call it a *right inverse* of A. (Is it also a left inverse?) If a matrix is both a left inverse and a right inverse, we call it the *inverse* of A, and denote it by A^{-1}. Note that this does not mean $\dfrac{1}{A}$, because that expression has no meaning!

For example, if

$$A \;=\; \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

then

$$A^{-1} \;=\; \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

The condition for a square matrix to have an inverse is well known. It is that a quantity known as the determinant of the matrix should have a non-zero value.

6.13 Determinants

We shall start by considering a set of two equations in two unknowns:

$$\begin{aligned} ax + by &= c\,, && (25) \\ px + qy &= r\,. && (26) \end{aligned}$$

You can think of this algebraically, or, if you know some Cartesian geometry, geometrically, as two lines.

A unique solution of these equations will exist under certain conditions, and geometrically, this is if the lines are not parallel. If the lines are parallel, then there will either no solution (when the lines are distinct) or there will be an infinite number of solutions (when the lines coincide). In the cases for no solution, the slopes of the lines must be the same: that is (unless the lines are parallel to the y–axis, that is $b = q = 0$)

$$-\frac{a}{b} = -\frac{p}{q} \qquad \text{or} \qquad aq - bp = 0\,.$$

(Note that the second form is also true in the case when the lines are parallel to the y–axis.)

So, if we assume that $aq - bp \neq 0$, we can find the unique solution of (25) and (26). This is

$$\begin{aligned} x &= \frac{cq - br}{aq - bp}\,, && (27) \\ y &= \frac{ar - cp}{aq - bp}\,. && (28) \end{aligned}$$

If we write (25) and (26) in matrix form, we have

$$\begin{pmatrix} a & b \\ p & q \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} c \\ r \end{pmatrix} .$$

This leads to the idea of writing $aq - bp$ in the form

$$\begin{vmatrix} a & b \\ p & q \end{vmatrix} .$$

This is called the **determinant** of $\begin{pmatrix} a & b \\ p & q \end{pmatrix}$.

Thus we can write (27) and (28) in the form

$$x = \frac{\begin{vmatrix} c & b \\ r & q \end{vmatrix}}{\begin{vmatrix} a & b \\ p & q \end{vmatrix}} , \tag{29}$$

$$y = \frac{\begin{vmatrix} a & c \\ p & r \end{vmatrix}}{\begin{vmatrix} a & b \\ p & q \end{vmatrix}} . \tag{30}$$

So, we see that the system of two simultaneous linear equations, (25) and (26), has a solution if and only if

$$\begin{vmatrix} a & b \\ p & q \end{vmatrix} \neq 0 .$$

This is a very useful criterion.

We can now consider three equations in three unknowns:

$$ax + by + cz = p , \tag{31}$$

$$dx + ey + fx = q , \tag{32}$$

$$gx + hy + iz = r . \tag{33}$$

Three dimensional geometric considerations lead us to a similar result.

The algebraic condition for a unique solution is

$$aei + bfg + cdh - ahf - dbi - gec \neq 0 .$$

It is not easy to see how to relate this to the matrix equation

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} p \\ q \\ r \end{pmatrix} . \tag{34}$$

We shall define the determinant of $\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ to be

$$aei + bfg + cdh - ahf - dbi - gec\,.$$

In both cases, with the appropriate interpretation, we have a matrix equation $A\overline{x} = \overline{p}$, and we write the determinant of the square matrix A as $\det(A)$ or $|A|$.

6.14 Properties of determinants

Let A be a square matrix (for the moment, either 2×2 or 3×3).

1. If B is a matrix obtained from A by interchanging two rows, then $\det(B) = -\det(A)$.

2. If B is a matrix obtained from A by interchanging two columns, then $\det(B) = -\det(A)$.

   These can readily be checked by direct computation.

3. If two rows of a matrix (two columns of a matrix) are the same, then the determinant has value zero.

   Suppose the matrix is A and the matrix obtained by interchanging the two identical rows (or columns) is B.

   From the previous result $\det(A) = -\det(B)$. But A and B are in fact identical. Thus $\det(A) = \det(B)$. This means that $\det(A)$ must be zero.

4. The addition of a constant multiple of a row (resp. column) of a determinant to another row (resp. column) does not change the value of the determinant.

   We show this for a 2×2 determinant.

   Let $\det(A) = \begin{vmatrix} a & b \\ p & q \end{vmatrix}$ and $\det(B) = \begin{vmatrix} a + \lambda p & b + \lambda q \\ p & q \end{vmatrix}$.

   Then, $\det(B) = \begin{vmatrix} a + \lambda p & b + \lambda q \\ p & q \end{vmatrix} = (a+\lambda p)q-(b+\lambda q)p = aq-bp = \det(A)\,.$

   This enables us to devise methods for evaluating determinant efficiently. By using appropriate additions of constant multiples of rows to other rows (resp. columns to other columns), we can put zeros into specific entries. For example, with $\lambda = -\frac{b}{q}$ (assuming $q \neq 0$)

$$\begin{aligned} \begin{vmatrix} a + \lambda p & b + \lambda q \\ p & q \end{vmatrix} &= \begin{vmatrix} a - \frac{bp}{q} & 0 \\ p & q \end{vmatrix} \\ &= \begin{vmatrix} \frac{aq-bp}{q} & 0 \\ p & q \end{vmatrix} \\ &= \frac{aq - bp}{q} \times q - 0 \;=\; aq - bp \quad \text{of course!} \end{aligned}$$

Provided that $\det(A)$ is non-zero, it is possible to perform these operations and result in a determinant with non-zero terms, only on the main diagonal: for example, we can proceed in the above example and get

$$\begin{vmatrix} \frac{ap-bq}{q} & 0 \\ 0 & q \end{vmatrix} .$$

5. The determinant of the transposed matrix is the same at the determinant of the matrix; that is $|A^T| = |A|$,
6. The determinant of the product of two matrices is equal to the product of the determinants of the matrices; that is $|A\,B| = |A|\,|B|$.

6.15 Determinants and Inverses of Matrices

In the section on square matrices, we mentioned that the condition for a square matrix to have an inverse was that its determinant was non-zero. We will show this here for 2×2 matrices. We recall that the solution of

$$\begin{pmatrix} a & b \\ p & q \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} c \\ r \end{pmatrix}$$

is given by (29) and (30). These two equations can be written in matrix form as follows:

$$\frac{1}{aq-bp} \begin{pmatrix} q & -b \\ -p & a \end{pmatrix} \begin{pmatrix} c \\ r \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} .$$

In other words, if $A = \begin{pmatrix} a & b \\ p & q \end{pmatrix}$, then the matrix $\dfrac{1}{\det(A)} \begin{pmatrix} q & -b \\ -p & a \end{pmatrix}$ is the inverse of the matrix A, and a necessary condition for this is that the determinant $|A| = \begin{vmatrix} a & b \\ p & q \end{vmatrix}$ is non-zero.

Writing this in matrix form, we have that $A\overline{x} = \overline{p}$ has a solution $B\overline{p} = \overline{x}$, provided that $\det(A)$ is non-zero. Here, B is the inverse of A. Note that $\det(B) = \det(A)$.

6.16 Generalities

To bring all this together, we have defined the determinants for 2×2 and 3×3 matrices as follows: if

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} ,$$

then

$$\det(A) = |A| := (a_{1,1}a_{2,2} - a_{1,2}a_{2,1}) .$$

This is also written as

$$A = \begin{vmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{vmatrix} ,$$

For 3×3 matrices, the rule is in terms of 2×2 determinants, expanding along the top row:

$$
\begin{aligned}
|A| &= \begin{vmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{vmatrix} \\
&= a_{1,1} \begin{vmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{vmatrix} - a_{1,2} \begin{vmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{vmatrix} + a_{1,3} \begin{vmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{vmatrix} .
\end{aligned}
$$

To the element $a_{i,j}$ in the matrix A, the *minor* $A_{i,j}$ is the matrix obtained from A by **deleting** the row with index i and the column with index j. in other words, delete the row and column containing $a_{i,j}$. We also associate with the position (i, j), the number $\delta_{i,j} := 1$ if $(i + j)$ is even, and $\delta_{i,j} := -1$ if $(i + j)$ is odd. This gives an array of signs:

$$
\begin{array}{ccccccc}
+ & - & + & - & + & - & \cdots \\
- & + & - & + & - & + & \cdots \\
+ & - & + & - & + & - & \cdots \\
- & + & - & + & - & + & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}
$$

and now we get

$$
\begin{aligned}
|A| &= \begin{vmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{vmatrix} \\
&= \delta_{1,1} a_{1,1} \left| A_{1,1} \right| + \delta_{1,2} a_{1,2} \left| A_{1,2} \right| + \delta_{1,3} a_{1,3} \left| A_{1,3} \right| \\
&= \sum_{j=1}^{3} \delta_{1,j} a_{1,j} \left| A_{1,j} \right| .
\end{aligned}
$$

Note that taking the sum of the elements of one row multiplied by the minors corresponding to elements of a different row (or similarly for columns) will result in a value of 0 (why?).

In general, we get

$$
|A| = \sum_{j=1}^{k} \delta_{1,j} a_{1,j} \left| A_{1,j} \right| .
$$

However, it is not required to expand along any particular row (or column), for we have

$$
|A| = \sum_{j=1}^{k} \delta_{i,j} a_{i,j} \left| A_{i,j} \right| .
$$

Since we know that addition of multiples of rows (or columns) to a row (or column) does not change the value of a determinant, we note that it may be a good strategy to do such additions first before expanding a determinant.

Finally, for 3×3 matrices, there is a diagonal process that parallels the 2×2 case. But it must be emphasised that this does not extend to higher orders.

We consider diagonals that go down and right to be positive, and those that go up and right to be negative.

$$
\begin{aligned}
|A| &= \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \\
&= \begin{bmatrix} a & & (b) \\ & \searrow & \\ (c) & & d \end{bmatrix} - \begin{bmatrix} (a) & & b \\ & \nearrow & \\ c & & (d) \end{bmatrix};
\end{aligned}
$$

$$
\begin{aligned}
|A| &= \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} \\
&= (aei + bfg + cdh) - (gec + hfa + idb) \\
&= \left[\begin{array}{ccccc|cccc}
a & & b & & c & & & & \\
& \searrow & & \searrow & & \searrow & & & \\
(d) & & e & & f & & d & & \\
& & & \searrow & & \searrow & & \searrow & \\
(g) & & (h) & & i & & g & & h
\end{array}\right] \\
&\quad - \left[\begin{array}{ccccc|cccc}
(a) & & (b) & & c & & a & & b \\
& & & \nearrow & & \nearrow & & \nearrow & \\
(d) & & e & & f & & d & & \\
& \nearrow & & \nearrow & & \nearrow & & & \\
g & & h & & i & & & &
\end{array}\right].
\end{aligned}
$$

We also see that the triple scalar product $\overline{x} \cdot \overline{y} \times \overline{z}$ can be evaluated as a determinant

$$
\begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}.
$$

6.17 Cramer's Rule

Returning to the solution of the matrix equation:

$$
\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} p \\ q \\ r \end{pmatrix}. \tag{34}
$$

which we write as

$$A\overline{x} \;=\; \overline{p}\,.$$

This has a unique solution

$$x \;=\; \frac{|A_1|}{|A|}\,, \qquad y \;=\; \frac{|A_2|}{|A|}\,, \qquad z \;=\; \frac{|A_3|}{|A|}\,,$$

if and only if $|A| \neq 0$, where A_1 is the matrix obtained for A by replacing the elements in the first row by the elements of $\overline{p}$, A_2 is the matrix obtained for A by replacing the elements in the second row by the elements of $\overline{p}$, and A_3 is the matrix obtained for A by replacing the elements in the third row by the elements of $\overline{p}$.

This result (which is also true for a system of n linear equations in n variables) is called **Cramer's Rule**.

Problem 58 *Prove Cramer's Rule for a system of two linear equations in two variables.*

Problem 59 *Prove Cramer's Rule for a system of three linear equations in three variables.*

6.18 Underdetermined systems

If we have a system of n equations in m unknowns, with $m > n$, then we do not, in general, have enough information to obtain a unique solution. We shall illustrate this by considering the system of two equations in three unknowns:

$$ax + by + cz \;=\; 0\,, \tag{35}$$

$$px + qy + rz \;=\; 0\,. \tag{36}$$

These tell us that (x, y, z) is perpendicular to (a, b, c) and (p, q, r), so that (x, y, z) is proportional to the cross product of (a, b, c) and (p, q, r). This is system is satisfied by

$$(x, y, z) \;=\; (br - cq, cp - ar, aq - bp)\,.$$

Returning to 35, clearly, $(x, y, z) = (0, 0, 0)$ is a solution. But this is not very interesting! We seek non-zero solutions for z by dividing both equations by z. Writing $X = \frac{x}{z}$ and $Y = \frac{y}{z}$, we obtain

$$aX + bY \;=\; -c\,,$$

$$pX + qY \;=\; -r\,.$$

We write this as a matrix equation:

$$A\overline{w} \;=\; \overline{v}\,,$$

where $A \;=\; \begin{pmatrix} a & b \\ p & q \end{pmatrix}$, $\overline{w} \;=\; \begin{pmatrix} X \\ Y \end{pmatrix}$ and $\overline{v} \;=\; \begin{pmatrix} -c \\ -r \end{pmatrix}$.

If this matrix A has an inverse (so, if $|A| \neq 0$), then we get the solution $\overline{w} \;=\; A^{-1}\overline{v} \;=\; \overline{u}$, say, where $\overline{u} \;=\; \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. We then have the solutions of the original pair of equations as $x \;=\; \alpha z$, $y \;=\; \beta z$.

6.19 Conics

The general equation of a conic

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + c \;=\; 0\,, \tag{37}$$

can be written as

$$(x, y, 1) \begin{pmatrix} a & h & g \\ h & b & f \\ g & f & c \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \;=\; 0\,,$$

or

$$x^T A\, x \;=\; 0\,.$$

(Reminder: x^T means the transpose [interchange of rows and columns] of the [column] matrix x.)

There are two ways of remembering the positions of the coefficients of the terms in x and y. One is that they follow alphabetical order as one travels down the main diagonal, and returns to the starting position via the sides.

$$\begin{array}{ccccc} a & \leftarrow & h & \leftarrow & g \\ \uparrow & \searrow & & & \uparrow \\ h & & b & & f \\ \uparrow & & & \searrow & \uparrow \\ g & \leftarrow & f & \leftarrow & c \end{array}$$

The other way, is from a mnemonic (courtesy Ed Barbeau):

a*ll* **h***airy* **g***orillas* **h***ave* **b***ig* **f***eet* **g***ood* **f***or* **c***limbing.*

6.19.1 Shifting the origin

By a change of coordinates, we can transform (37) to one where the linear terms to not appear:

$$AX^2 + 2HXY + BY^2 + C \;=\; 0\,. \tag{38}$$

Let $X = x - u$ and $Y = y - v$ (so that $x - X + u$ and $y = Y + v$). This represents a translation of the origin of the coordinate system to the point (u, v). Substituting into (37) yields:

$$\begin{aligned} aX^2 + 2hXY + bY^2 + 2(au + hv + g)X + 2(hu + bv + f)Y \\ +(au^2 + 2huv + bv^2 + 2gu + 2fv + c) \;=\; 0\,. \end{aligned}$$

If $ab - h^2 \neq 0$, then the system

$$\begin{aligned} au + hv + g &= 0\,, \\ hu + bv + f &= 0\,, \end{aligned} \tag{39}$$

can be solved uniquely for u and v. With these values, we obtain (38) with

$$A = a\,,, \quad H = h\,, \quad B = b\,,$$

$$C = au^2 + 2huv + bv^2 + 2gu + 2fv + c = gu + fv + c\,.$$

The matrix for the conic is then

$$\begin{pmatrix} A & H & 0 \\ H & B & 0 \\ 0 & 0 & C \end{pmatrix} = \begin{pmatrix} a & h & 0 \\ h & b & 0 \\ 0 & 0 & gu + fv + c \end{pmatrix},$$

and its determinant is

$$\begin{aligned} &\left(ab - h^2\right)(gu + fv + c) \\ &= \left(ab - h^2\right) + (hu + bv)(af - gh) + (au + hv)(bg - fh) \\ &= abc - ch^2 - af^2 - bg^2 + 2fgh\,, \end{aligned}$$

which is the determinant of the original matrix of coefficients.

Thus, $\begin{vmatrix} a & h & g \\ h & b & f \\ g & f & c \end{vmatrix}$ remains **invariant** when we make a translation of coordinates to eliminate the linear terms. Note that another invariant of this translation is $ab - h^2$, since $AB - H^2 = ab - h^2$.

In the case when $ab - h^2 = 0$, the system (39) is not uniquely solvable, and then we have

$$ax^2 + 2hxy + by^2 = (ax + by)^2\,.$$

The system (37) becomes

$$(ax + by)^2 = -2(gx + fy) - c\,.$$

If $af \neq bg$, then we can make the change of coordinates

$$\begin{aligned} X &= -2(gx + fy)\,, \\ Y &= ax + by\,, \end{aligned}$$

to obtain the form

$$Y^2 = X + C\,,$$

while, if $af = bg$, then $gx + fy$ is a constant multiple of $ax + by$, and the equation has the form

$$Y^2 = C\,.$$

Thus, when $ab - h^2 = 0$, we find that (37) is the equation of a parabola or of a pair of parallel (possibly coincident) straight lines.

6.19.2 Rotating the axes

We suppose that we have selected a coordinate system to put (37) in the form

$$ax^2 + 2hxy + by^2 + c \;=\; 0\,, \tag{40}$$

and that $ab - h^2 \neq 0$. The next step is to rotate the coordinate axes to obtain a form without an "xy" term. So, we let

$$\begin{array}{rcl} X &=& x\cos\theta - y\sin\theta\,, \\ Y &=& x\sin\theta + y\cos\theta\,. \end{array} \tag{41}$$

This is equivalent to

$$\begin{array}{rcl} x &=& X\cos\theta + Y\sin\theta\,, \\ y &=& -X\sin\theta + Y\cos\theta\,. \end{array} \tag{42}$$

The axis $X = 0$ in the new system corresponds to the line $y = x\cot\theta$ in the old system, and the axis $Y = 0$ in the new system corresponds to the line $y = -x\tan\theta$ in the old system.

We can write this in vector–matrix from as

$$(X\;Y\;1) = (x\;y\;1)R \qquad \text{and} \qquad \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = R^{-1}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix},$$

where

$$R \;=\; \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Substituting (41) into $ax^2 + 2hxy + by^2 + c = 0$ yields

$$\left(a\cos^2\theta - 2h\sin\theta\cos\theta + b\sin^2\theta\right)X^2 + \left((a-b)\sin 2\theta + 2h\cos 2\theta\right)XY$$
$$+ \left(a\sin^2\theta + 2h\sin\theta\cos\theta + b\cos^2\theta\right)Y^2 + c \;=\; 0\,.$$

The trick now is to select θ such that $\tan 2\theta = -\frac{2h}{a-b}$. Then (40) has the form

$$AX^2 + BY^2 + C \;=\; 0\,, \tag{43}$$

or

$$(X\;Y\;1)\begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{pmatrix}\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \;=\; 0\,,$$

where

$$\begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} a & h & 0 \\ h & b & 0 \\ 0 & 0 & c \end{pmatrix}\begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Since the determinant of R is equal to 1, the determinant of the final matrix of coefficients is equal to the determinant of the original matrix, and so $AB = (ab - h^2)$,

The only case in which the rotation of axes is indeterminate occurs when $h = a - b = 0$. Here, equation (40) has the form $a(x^2 + y^2) + c = 0$, and this is the equation of a circle when $ac < 0$, and has a null locus when $ac > 0$ or $a = 0$. When $a \neq 0$ and $c = 0$, we have the point $(0, 0)$.

We now consider equation (43). If A, B and $C \neq 0$ all have the same sign, then the locus is null. If A and B have the same sign and C has the opposite sign, then the locus is an ellipse. If A and B have opposite signs and $C = 0$, then the locus is a pair of intersecting straight lines. If A and B have opposite signs and $C \neq 0$, then the locus is a hyperbola.

6.19.3 Classification of conics

Because of the invariance under translation and rotation of axes, we can describe the locus of (37) without actually having to effect the transformations.

If $ab - h^2 = 0$, then (37) can be changed to the form

$$y^2 + 2px + q \;=\; 0\,,$$

with coefficient matrix

$$\begin{pmatrix} 0 & 0 & p \\ 0 & 1 & 0 \\ p & 0 & q \end{pmatrix}.$$

The determinant of this matrix is $-p^2$. So, if $p = 0$, we have parallel lines and if $p \neq 0$, we have a parabola.

On the other hand, if $ab - h^2 \neq 0$, then we have a central conic (ellipse, circle, hyperbola, intersecting line pair).

In summary

$\begin{vmatrix} a & h & g \\ h & b & f \\ g & f & c \end{vmatrix}$	$\begin{vmatrix} a & h \\ h & b \end{vmatrix}$	Conic
Negative	Positive	Ellipse
Negative	Positive	Circle if $a = b$
Positive	Negative	Hyperbola
Zero	Negative	Intersecting Line Pair
Zero	Zero	Parabola or Parallel Line Pair
Negative	Zero	Parallel Line Pair

Any other combination is impossible.

Problem 60 *Investigate the inability of the above to distinguish the Parabola for the parallel line pair of lines equidistant from the origin by considering the co-factors of other terms in the matrix* $\begin{pmatrix} a & h & g \\ h & b & f \\ g & f & c \end{pmatrix}$.

Horses

The proof fails because the inductive step is not valid for the step from $k = 1$ to $k = 2$. The inductive step depends on there being at least three horses in $P(k+1)$.

ATOM

A Taste Of Mathematics / Aime-T-On les Mathématiques

1. Edward J. Barbeau *Mathematical Olympiads' Correspondence Program (1995-1996)*
2. Bruce L.R. Shawyer *Algebra — Intermediate Methods*
3. Peter I. Booth, John Grant McLoughlin, and Bruce L.R. Shawyer *Problems for Mathematics Leagues*
4. Edward J. Barbeau, and Bruce L.R. Shawyer *Inequalities*
5. Richard Hoshino, and John Grant McLoughlin *Combinatorial Explorations*

Cost per volume (including shipping and handling charges):
Regular Rate \$12.00 — CMS Member's Rate \$9.00 (excluding taxes).
For non-Canadian shipping addresses the rates quoted must be paid in US funds.

Coût par volume (frais y compris d'expédition et de manutention) :
Taux régulier 12,00 \$ — Tarif d'membre de la SMC 9,00 \$ (avant taxes).
Pour une adresse à l'étranger, paiement requis en devises américaines.

For more information and to order:
Pour plus d'informations et pour compléter une commande :

www.cms.math.ca/Publications/books

or contact:
ou contacter :

CMS/SMC Publications
577 King Edward, Ottawa (Ontario) Canada K1N 6N5

Email: publications@cms.math.ca
Courriel : publications@smc.math.ca

Canadian Mathematical Society / Société mathématique du Canada
577 King Edward, POB/CP 450-A
Ottawa, Ontario, Canada K1N 6N5
www.cms.math.ca