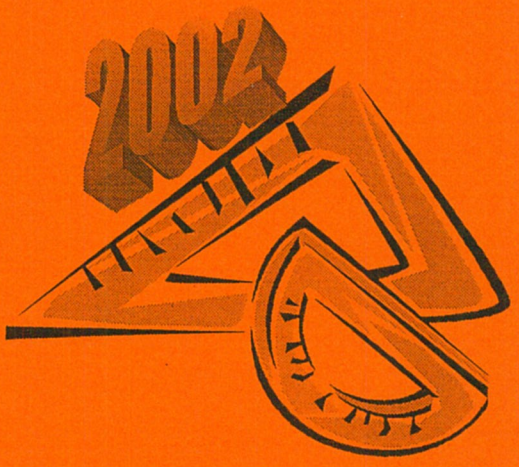
Collection VI



The Collection VI

Editors: Dr. I. Sciriha

and

Tanja Telenta

Department of Mathematics

Faculty of Science

University of Malta

Proceedings of Workshop held on the 30th October 2002

Contents

Foreword

"He who never made a mistake, never made a discovery."
Ted Hurley, National University of Ireland

The Collection VI workshop proved to be one of the most interesting. During the meeting, the lecturers in the mathematics department were very happy to discover that in their classes they have students who can discern and question established results. We were surprised that we managed to trigger off the interest of the students precisely in areas that they came across first at university. The motivation of some of them was so impelling that they somewhat shyly admitted that they spent the summer engaged in this new resaerch. The department of mathematics is committed to encourage such students and confesses that they are the driving force behind its endeavours.

Dr. Irene Sciriha.

Organizer.

The Collection VI

Date: 30th October 2002

Time: 15.00 - 16.15

Venue: Faculty of Science
Department of Mathematics
Room LC119

A seminar/workshop is being held on Wednesday 30th October 2002 at 3.00 p.m. Students and staff from the Department of Mathematics, Faculty of Science will present ideas from various fields of mathematics.

Keynote speakers:

Cheryl Zerafa & Pamela Cohen	Fibonacci Sequence & The Golden Ratio
Professor A. Buhagiar	Data Reduction and Eigenvalues
Anne Yancey	Geometry On a Triangular Grid
Andrew Duncan & Andrew Cortis	On Cayley Diagrams
Matthew Sant	On One of Hilbert's Problems
Alexander Farrugia	On Eigenvalues

We shall end with a brief session for spontaneous problem posing. You are cordially invited to attend.

Abstracts of possible proofs or conjectures which you wish to share with us in this meeting, or in a future one, may be sent to Dr. I. Sciriha or Ms. A. Attard, Department of Mathematics, (marked The Collection), at any time of the year.

Dr. I. Sciriha

Organiser

Fibonacci Sequence and The Golden Ratio

Pamela Cohen and Cheryl Zerafa

Fibonacci, or more correctly Leonardo da Pisa, was born in Pisa in 1175AD. He was the son of a Pisan merchant who also served as a customs officer in North Africa. He travelled widely in Barbary (Algeria) and was later sent on business trips to Egypt, Syria, Greece, Sicily and Provence. In 1200 he returned to Pisa and used the knowledge he had gained on his travels to write ***Liber abaci*** in which he introduced the Latin-speaking world to the decimal number system.

Fibonacci is perhaps best known for a simple series of numbers, introduced in ***Liber abaci*** and later named the ***Fibonacci numbers*** in his honour.

The resulting sequence is:

1, 1, 2, 3, 5, 8, 13, 21, 34, 55, ...

(Fibonacci omitted the first term in *Liber abaci*). This sequence, in which each number is the sum of the two preceding numbers, has proved extremely fruitful and appears in many different areas of mathematics and science.

A problem in the third section of *Liber abaci* which can be solved using this sequence is the following:

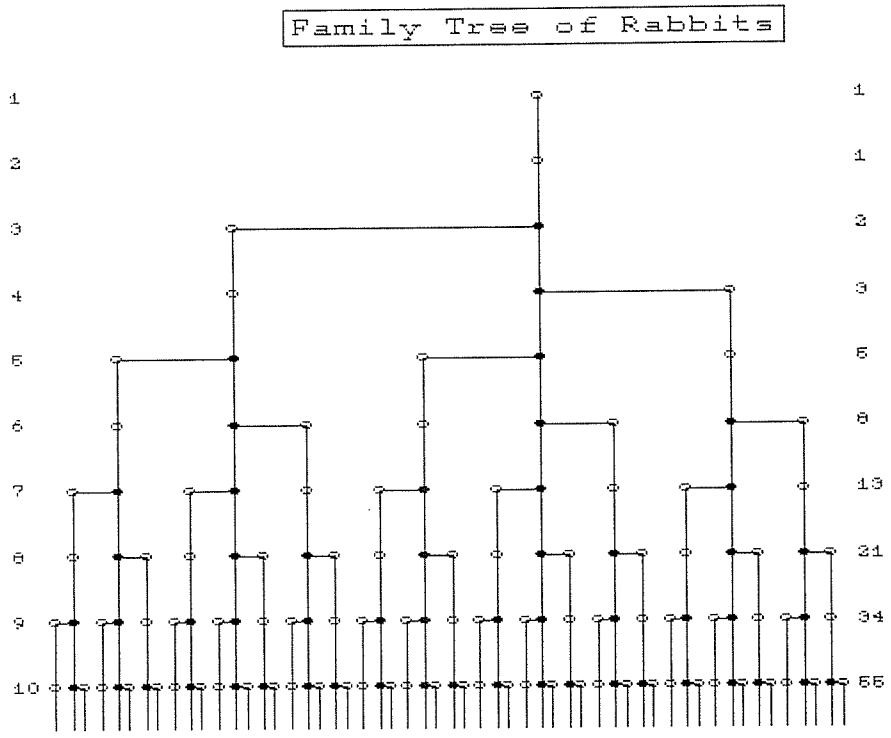
Suppose a newly-born pair of rabbits, one male, one female, are put in a field. Rabbits are able to mate at the age of one month so that at the end of its second month a female can produce another pair of rabbits. Suppose that our rabbits ***never die*** *and that the female* ***always*** *produces one new pair (one male, one female)* ***every month*** *from the second month on. The puzzle that Fibonacci posed was...*
How many pairs will there be in one year?

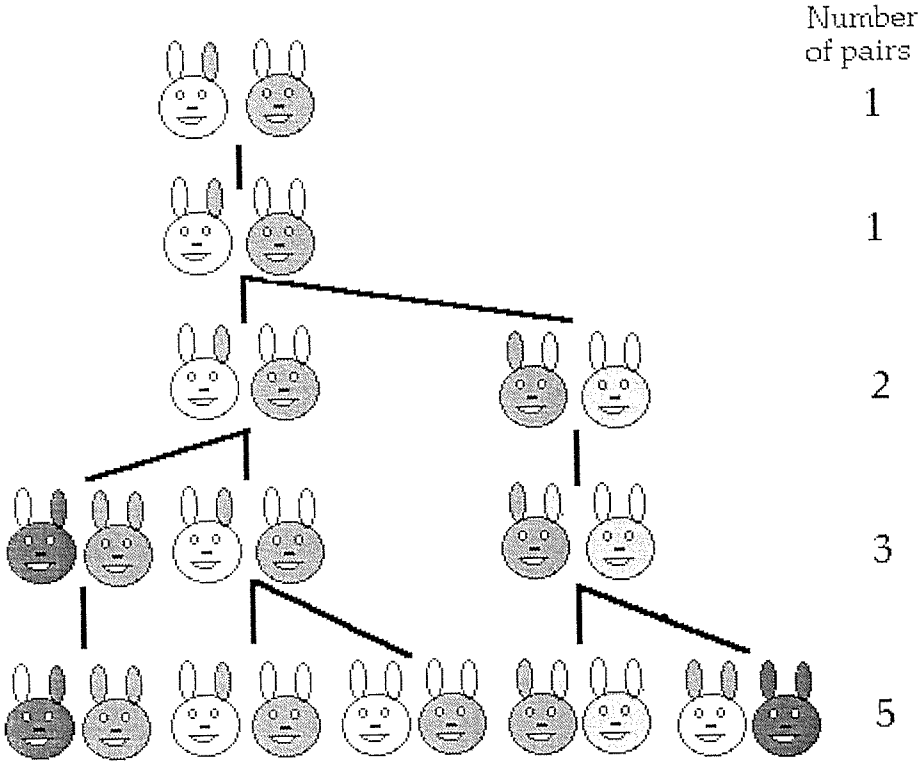
Imagine that there are x_n pairs *of rabbits after n months. The number of pairs in month n+1 will be* x_n *(in this problem, rabbits never die) plus the number of new pairs born. But new pairs are only born to pairs at least 1 month old, so there will be* x_{n-1} *new pairs.*

$$x_{n+1} = x_n + x_{n-1}$$

Which is simply the rule for generating the Fibonacci numbers!!!

The figure below is a view of the rabbit's family tree showing how the Fibonacci sequence is generate:


Family Tree of Rabbits
1
2
3
4
5
6
7
8
9
10
1
1
2
3
5
8
13
21
34
55


Number
of pairs
1
1
2
3
5

The Golden Section

Taking the ratio of successive terms in the Fibonacci series: (1, 1, 2, 3, 5, 8, 13, ..) and dividing each by the number before it, the following series of numbers is obtained:

$^{1}/_{1} = \mathbf{1}$

$^{2}/_{1} = \mathbf{2}$

$^{3}/_{2} = \mathbf{1{\cdot}5}$

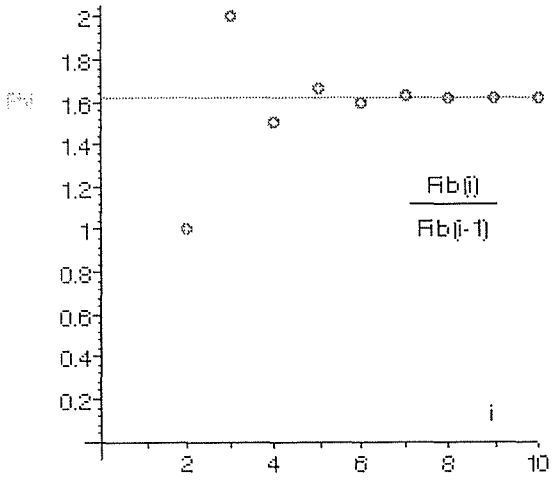
$^{5}/_{3} = \mathbf{1{\cdot}666...,}$

$^{8}/_{5} = \mathbf{1{\cdot}6}$

$^{13}/_{8} = \mathbf{1{\cdot}625}$

$^{21}/_{13} = \mathbf{1{\cdot}61538...}$

The graph shows that the values seem to be tending to a limit. This limit is actually the positive root of a quadratic equation and is called the *golden section*, *golden ratio* or sometimes the *golden mean*.



If you take two successive terms of the series, a, b, and $a+b$ then

$$\frac{b}{a} \cong \frac{a+b}{b}$$

$$\cong \frac{a}{b} + 1$$

We define the golden section, ϕ (*phi*), to be the limit of $\frac{b}{a}$, so:

$$\phi = \frac{1}{\phi} + 1$$

$$\phi^2 - \phi - 1 = 0$$

$$\phi = \frac{1+\sqrt{5}}{2} \approx 1.618$$

Note that the golden section is denoted by the Greek letter *phi*.

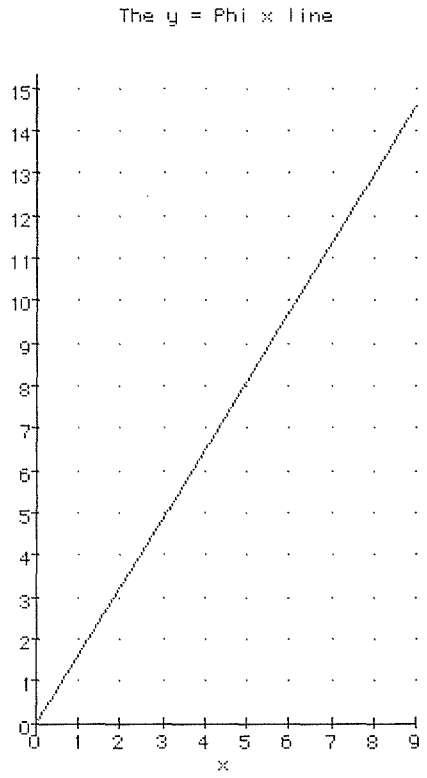
The table below shows properties of the solutions to the quadratic equation:

$$\phi^2 - \phi - 1 = 0$$

$\phi_1 . \phi_2 = 1$	$\phi_1 - \phi_2 = 1$	$\phi_1 + \phi_2 = \sqrt{5}$
$\phi_1 = 1.6180339..$	$\phi_2 = 0.6180339..$	
$\phi_1 = 1 + \phi_2$	$\phi_2 = \phi_1 - 1$	
$\phi_1 = 1/\phi_2$	$\phi_2 = 1/\phi_1$	
$\phi_1^{\ 2} = \phi_1 + 1$	$(-\phi_2)^2 = -\phi_2 + 1$ or $\phi_2^{\ 2} = 1 - \phi_2$	
$\phi_1 = (\sqrt{5} + 1)/2$	$\phi_2 = (\sqrt{5} - 1)/2$	

The Fibonacci numbers can also arise from the number ϕ. The graph below shows a line whose gradient is ϕ, that is the line :

$$y = \phi . x = (1{\cdot}6180339...)x$$



Since ϕ is irrational, the graph will never go through any points of the form (i,j) where are integers. The nearest integer-coordinate points to the ϕ - line are (0,1), (1,2), (2,3), (3,5)...

These coordinates are successive Fibonacci numbers. The ratio y/x for each Fibonacci point (x,y) approaches $\phi = 1.618...$

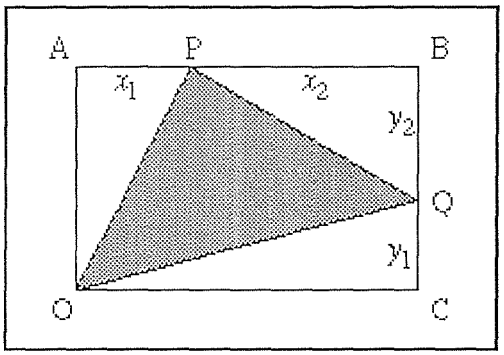
This graph also shows that the Fibonacci points are the closest points to the ϕ - line.

Problem involving the Golden Ratio:

Rectangle triangle problem

Consider a rectangle OABC from which you remove three right-angled triangles leaving a fourth triangle OPQ as shown in the diagram below.

How must you position the points P and Q so that the area of each of the three removed triangles is the same? That is, what are the ratios AP : PB and CQ : QB?



If we label the distances AP, PB, CQ and QB as shown above then we can write three equations for the areas of the triangles as follows:

$$\text{Area of OAP} = \frac{1}{2}x_1(y_1 + y_2) \qquad (1)$$

$$\text{Area of PBQ} = \frac{1}{2}x_2 y_2 \qquad (2)$$

$$\text{Area of OCQ} = \frac{1}{2}(x_1 + x_2)y_1 \qquad (3)$$

From (1) and (3):

$$x_1y_1 + x_1y_2 = x_1y_1 + x_2y_1$$

$$x_1y_2 = x_2y_1$$

$$\frac{x_2}{x_1} = \frac{y_2}{y_1}$$

Notice that the two ratios are the same! If we call this ratio r then we can calculate a polynomial for r as follows:

From (1) and (2):

$$x_1y_1 + x_1y_2 = x_2y_2$$

$$y_1 + y_2 = \frac{x_2}{x_1}y_2 = ry_2$$

$$1 + \frac{y_2}{y_1} = r\frac{y_2}{y_1}$$

$$1 + r = r^2$$

$$r^2 - r - 1 = 0$$

Taking the positive root gives us the golden ratio:

$$r = \frac{1}{2}(1 + \sqrt{5}) \approx 1.618$$

Binet's Formula for the nth Fibonacci number

The n-th Fibonacci number is the sum of the (n-1)th and the (n-2)th.

Theorem: A formula for the nth Fibonacci number Fib(n), which contains only n and does not need any other (earlier) Fibonacci values involves the golden section number ϕ_1 and its reciprocal ϕ_2;

$$\text{Fib}(n) = \frac{\phi_1^{\,n} - (-\phi_1)^{-n}}{\sqrt{5}} = \frac{\phi_1^{\,n} - (-\phi_2)^{n}}{\sqrt{5}}$$

Proof:

Since ϕ_1 and $-\phi_2$ are the two roots of $x^2 = x+1$, we get the following:

If $x^2 = x+1$ then,

$$\mathbf{x^n = fib(n)\ x + fib(n\text{-}1)}$$

..... (1) for n>0.

Proving (1) by induction:

RTP *True for n = 1*

$x^1 = (\text{fib} (1))x + \text{fib} (0)$

$x = 1 \cdot x + 0$

$x = x$

Suppose it is true for $n = k$

i.e. $x^k = \text{fib}(k) + \text{fib}(k-1)$

RTP *True for* $n = k + 1$

i.e. $x^{k+1} = (\text{fib}(k+1))x + \text{fib}(k)$

$$x^{k+1} = x^k \cdot x$$
$$= (\text{fib}(k) \cdot x + \text{fib}(k-1)) \cdot x$$
$$= \text{fib}(k) \cdot x^2 + \text{fib}(k-1) \cdot x$$

But $x^2 = x + 1$,

$$x^{k+1} = \text{fib}(k) \cdot (x+1) + \text{fib}(k-1) \cdot x$$
$$= \text{fib}(k) \cdot x + \text{fib}(k) + \text{fib}(k-1) \cdot x$$
$$= (\text{fib}(k) + \text{fib}(k-1)) \cdot x + \text{fib}(k)$$

But $\text{fib}(k) + \text{fib}(k-1) = \text{fib}(k+1)$

Therefore,

$$x^{k+1} = \text{fib}(k+1) \cdot x + \text{fib}(k)$$

Now the two roots of $x^2 = x+1$ are $\phi_1 = (1+\sqrt{5})/2 = 1{\cdot}6180339...$ and $-\phi_2 = (1-\sqrt{5})/2 = -0{\cdot}6180339....$ and thus that these are the only two values for which their powers can be expressed as Fibonacci multiples of themselves, as given in the formula.

So, from the formula above, we have:

$$\phi_1^{\,n} = fib(n)\,\phi_1 + fib(n-1) \qquad \textbf{(A)}$$

and also

$$(-\phi_2)^n = fib(n)\,(-\phi_2) + fib(n-1) \qquad \textbf{(B)}$$

Subtracting (B) from (A) gives:

$$\phi_1^{\,n} - (-\phi_2)^n = fib(n)(\,\phi_1 - (-\phi_2)) \qquad \textbf{(C)}$$

and from this we derive an initial formula for fib(n):

$$fib\,(n) = \frac{\phi_1^{\,n} - (-\phi_2)^n}{(\,\phi_1 - (-\phi_2))}$$

But $\phi_1 - (-\phi_2) = \sqrt{5}$, so we can write this as:

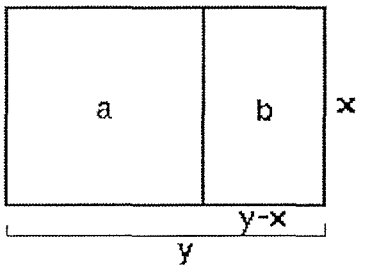
$$fib\,(n) = \frac{\phi_1^{\,n} - (-\phi_2)^n}{\sqrt{5}}$$

To get the form of the formula which involves only ϕ_1, we replace ϕ_2 by $1/\phi_1$ so that:

$$fib\,(n) = \frac{\phi_1^{\,n} - (-(1\,/\,\phi_1))^n}{\sqrt{5}}$$

$$fib\,(n) = \frac{\phi_1^{\,n} - (-\phi_1)^{-n}}{\sqrt{5}}$$

The Golden Rectangle

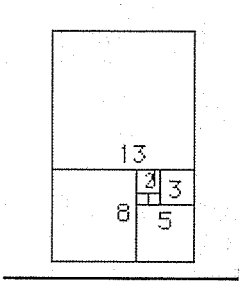


The rectangle shown here is a Golden Rectangle with proportions x/y. The section labeled "a" is a square drawn in the rectangle with proportions x/x. The section labeled "b" is another Golden Rectangle, this one with proportions (y-x)/x. In other words, the ratio of the lengths of the sides of section "b" is the same as the ratio of the length of the sides of the entire large rectangle. This is the characteristic of a Golden Rectangle. When you square it (inscribe a square with lengths the same as the length of the short side of the rectangle), you are left with another rectangle with the same proportions as the original.

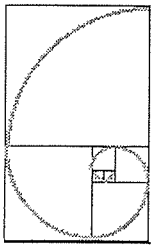
The Fibonacci Rectangles and Shell Spirals

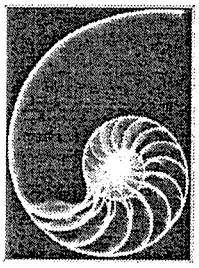
Another picture showing the Fibonacci numbers 1,1,2,3,5,8,13,21,.. can be achieved starting with two small squares of size 1 next to each other. On top of both of these a square of size 2 (=1+1) is drawn.

Now a new square - touching both a unit square and the latest square of side 2 - so having sides 3 units long is drawn; and then another touching both the 2-square and the 3-square (which has sides of 5 units). We can continue adding squares around the picture, *each new square having a side which is as long as the sum of the latest two square's sides*. This set of rectangles whose sides are two successive Fibonacci numbers in length and which are composed of squares with sides which are Fibonacci numbers, we will call this set the **Fibonacci Rectangles**.



The next diagram shows that we can draw a spiral by putting together quarter circles, one in each new square. This is a spiral (the **Fibonacci Spiral**). A similar curve to this occurs in nature as the shape of a snail shell or some sea shells.

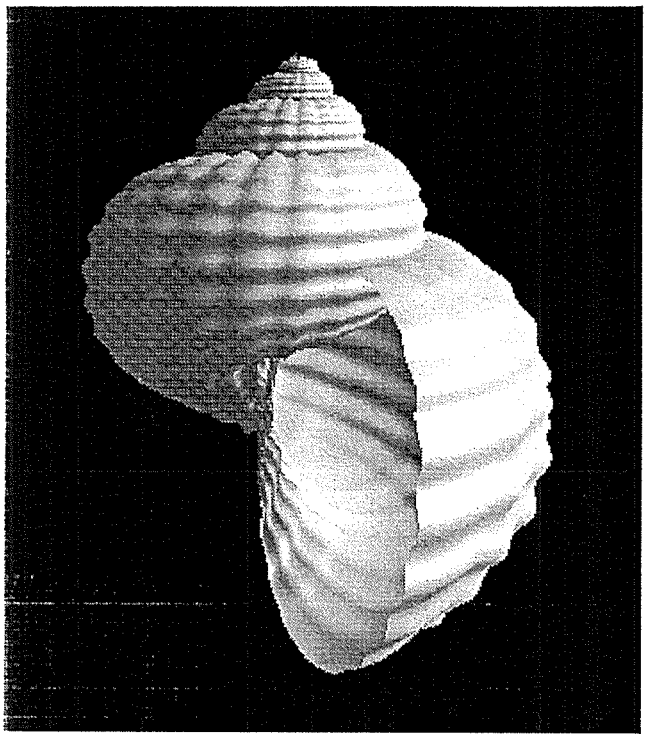


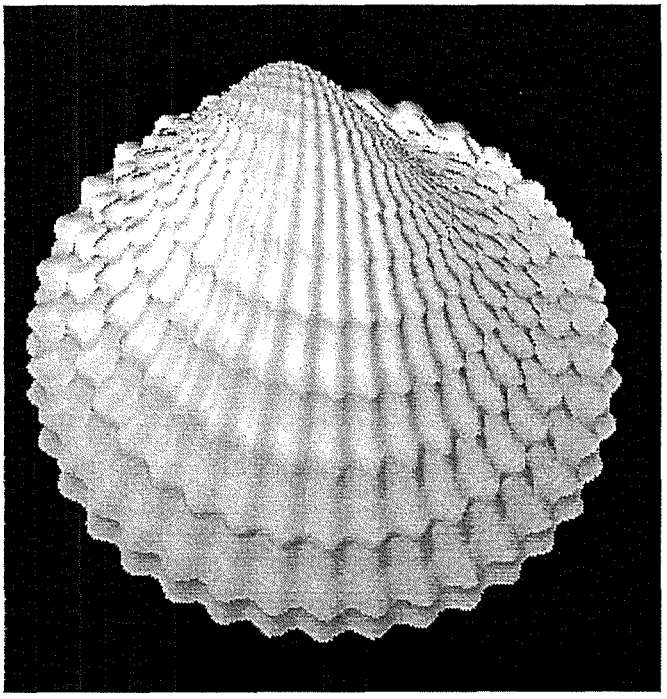


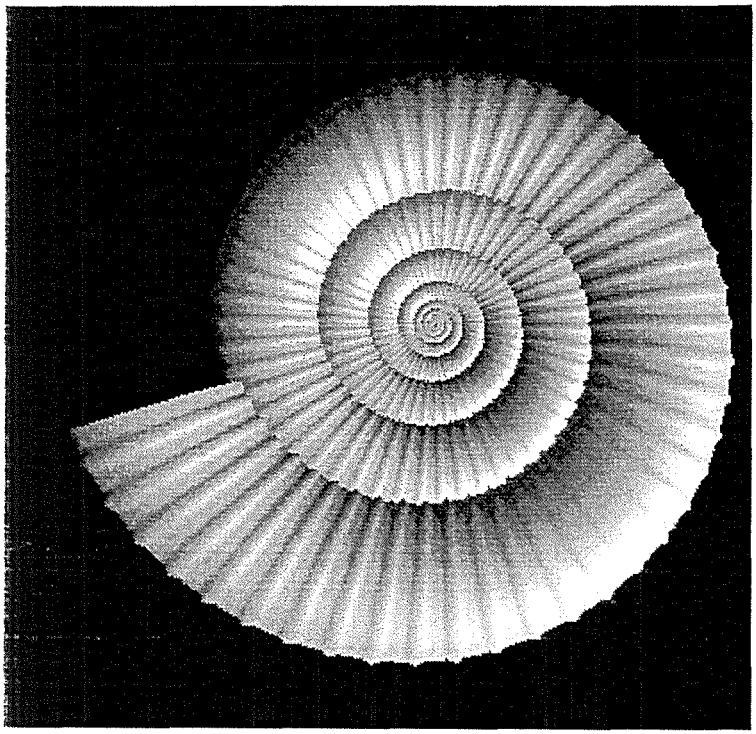
Whereas the Fibonacci Rectangles spiral increases in size by a factor of ϕ (1.618..) in a *quarter of a turn* (i.e. a point a further quarter of a turn round the curve is 1.618... times as far from the centre, and this applies to *all* points on the curve), the Nautilus spiral curve takes a *whole turn* before points move a factor of 1.618... from the centre.

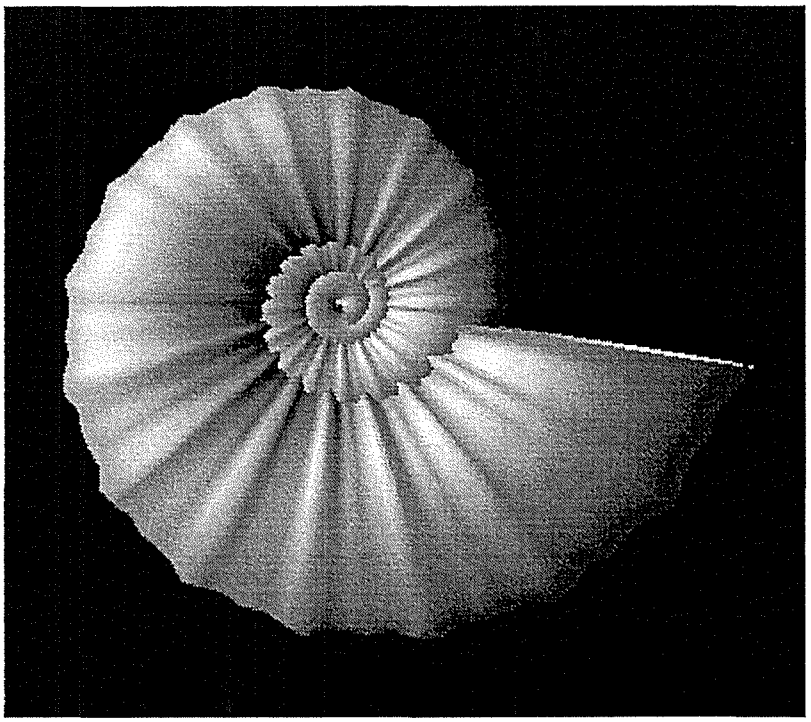
These spiral shapes are called Equiangular or Logarithmic spirals.

The following are some examples of fibonacci spirals in nature







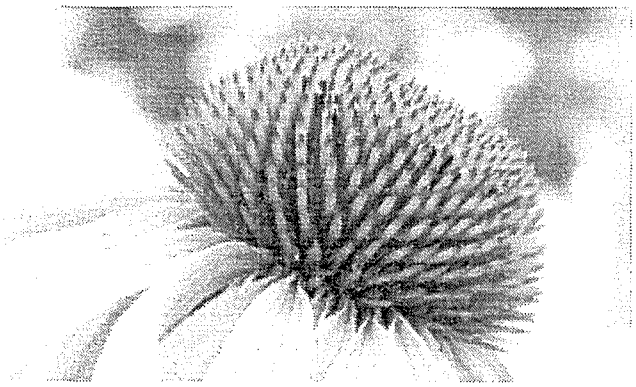


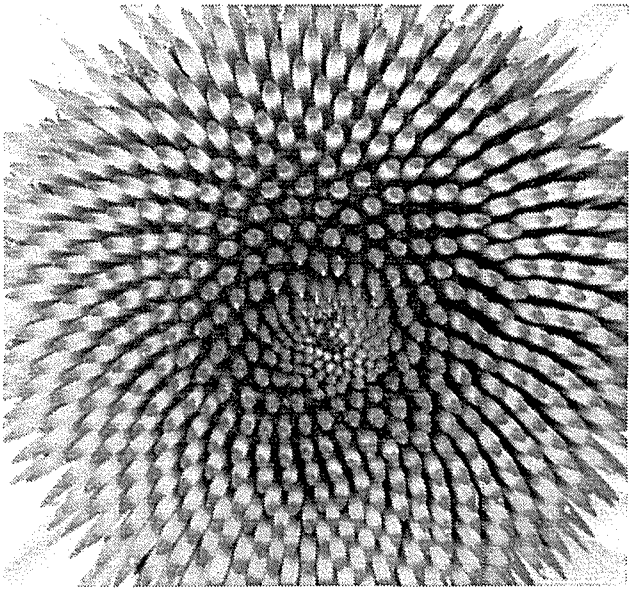
Petals on flowers

On many plants, the number of petals is a Fibonacci number: Buttercups have 5 petals; lilies and iris have 3 petals; some delphiniums have 8; corn marigolds have 13 petals; some asters have 21 whereas daisies can be found with 34, 55 or even 89 petals.

Some species are very precise about the number of petals they have - eg buttercups, but others have petals that are very near those above, with the average being a Fibonacci number.

Seed heads



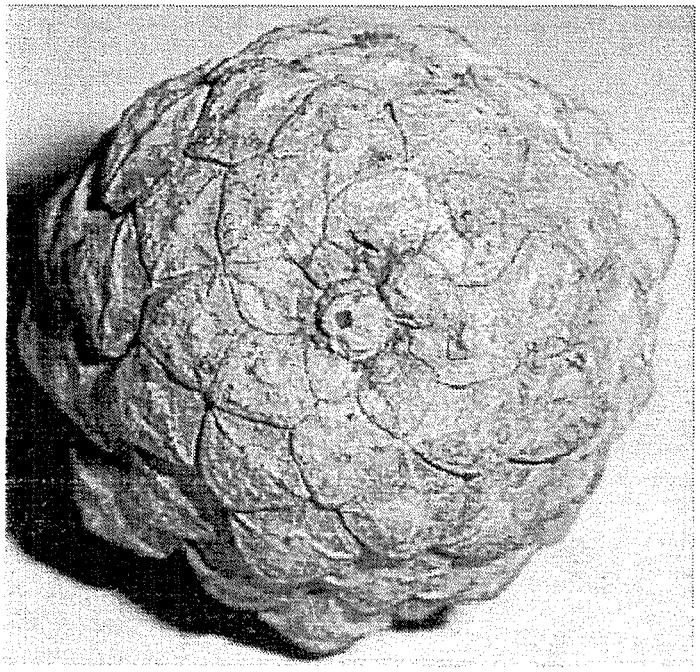


Fibonacci numbers can also be seen in the arrangement of seeds on flower heads. The picture here is a photograph of a Coneflower.

You can see that the orange "petals" seem to form spirals curving both to the left and to the right. At the edge of the picture, if you count those spiralling to the right as you go outwards, there are 55 spirals. A little further towards the centre and you can count 34 spirals. The pair of numbers are neighbours in the Fibonacci series.

Pine cones

Pine cones show the Fibonacci Spirals clearly. Here is a picture of an ordinary pinecone seen from its base where the stalk connects it to the tree.



Exploration and reduction of data using principal component analysis.

Anton Buhagiar

Department of Mathematics
University of Malta
Msida.

ABSTRACT: In a data set with two variables only, a scatterplot between the two variables can be easily plotted to represent the data visually. When the number of variables in the data set is large, however, it is more difficult to represent visually. The method of *principal component analysis* (*PCA*) can sometimes be used to represent the data faithfully in few dimensions (eg. three or less), with little or no loss of information. This reduction in dimensionality is best achieved when the original variables are highly correlated, positively or negatively. In this case, it is quite conceivable that 20 or 30 original variables can be adequately represented by two or three new variables, which are suitable combinations of the original ones, and which are called *principal components*. Principal components are uncorrelated between themselves, so that each component describes a different dimension of the data. The principal components can also be arranged in descending order of their variance. The first component has the largest variance, and is the most important, followed by the second component with the second largest variance, and so on. The first two components can then be evaluated for each case in the data set and plotted against each other in a scattergraph, the score for the first component being plotted along the horizontal axis, the score of the second component being plotted on the vertical axis. This scatterplot is a parsimonious two-dimensional picture of the variables and cases in the original data set. We illustrate the method by applying it to simulated datasets, and to a dataset containing national track record times for males and females in various countries.

Keywords: Matrix plots, correlation, correlation matrix; spheres, ellipsoids; rotation of coordinates, principal components, factors, eigenvalues, scree plot, factor loadings for variables, factor scores of cases, uses of principal component analysis such as exploration of data, dimension reduction, regrouping of variables and ordering of data; application to a data set containing national track record times for males and females in various countries.

Introduction

Suppose we have a very detailed database on a random sample of 1000 sixth form students, containing information on their academic performance, medical details, body measurements etc. We also assume that there are no missing values. For simplicity's sake we will take different subsets of variables at a time to illustrate how the scatterplots of the variables are affected by the correlation between them.

We will start off at first with a data set of 1000 cases (rows) each containing four variables (or columns):

id, identification number of student (1-1000),
V1, mark attained by student in Mathematics test (0 to 100),
V2, height of student in metres,
V3, systolic blood pressure of student in mm Hg.

Since the interval variables V1 to V3 have widely disparate means and standard deviations, it is often convenient to standardize the variables. So if the mean mark of Mathematics is 58.2 and its standard deviation is 8.5, the standardised variable z1 corresponding to V1 is defined as z1 = (V1 – 58.2)/8.5, ie. z1 = (original variable – its mean)/ its standard deviation. The other interval variables can be standardised in a similar manner. All standardised variables are dimensionless (ie. do not carry any units) and have mean = 0 and standard deviation = 1. Besides, one does not have to worry about the scale of standardized variables in scatterplots: they are all centred about the origin and most of the readings lie between 2.0 and –2. 0 if the original variables are normally distributed. Very often, variables are used in their standardized form in multivariate statistics.

The variables mentioned above, namely V1, V2 and V3 are examples of *uncorrelated* variables. (The height of a student does not usually affect his performance in the Mathematics test! etc). Knowledge of one variable does not help to predict the other variables. The degree of association or correlation between two variables in a data set is measured by Pearson's coefficient of correlation, r . [1 – 5]. When two variables are uncorrelated, the coefficient of correlation r is near zero, and a scatterplot of the two variables in their standardized version will be roughly *circular* in shape. One cannot predict the value of one variable from values of the other. In this case the variables are said to be *independent* or *uncorrelated.* There is no relation between them.

To illustrate this we simulated [6, 7] a data file of 1000 cases each having three variables, which had little or no correlation between them, just like the variables V1, V2 and V3 above. The correlation between these three variables, which we shall also call V1, V2, V3, can be summarized in a correlation matrix[2] , $\mathbf{R}_1$, as follows:

$$\mathbf{R}_1 = \begin{array}{c|ccc} & \mathbf{V1} & \mathbf{V2} & \mathbf{V3} \\ \mathbf{V1} & 1 & -0.031 & -0.018 \\ \mathbf{V2} & -0.031 & 1 & -0.063 \\ \mathbf{V3} & -0.018 & -0.063 & 1 \end{array}$$

Like all correlation matrices, $\mathbf{R}_1$ has 1's down the diagonal signifying perfect correlation between a variable and itself! (in this case, V1 with V1, V2 with V2, and V3 with V3). The matrix is also symmetric ie. the correlation between V1 and V2 is – 0.031, which is identical to the correlation between V2 and V1; and so on for the other variables. As expected from the way we simulated the data, all the off diagonal correlations are very

small or near to zero (– 0.031 for the correlation between V1and V2, – 0.018 for V1-V3, and – 0.063 for V2-V3). It is now instructive to examine the geometric scatter of these uncorrelated variables.

The bivariate scatterplots of each pair of variables can be succinctly summarised by a *matrix plot* [8,9]. Essentially this plot is analogous to the correlation matrix, except that each correlation is replaced by the corresponding scatterplot. For example, the plot in the 1'st row and 2nd column is the scatter plot between V1 (in its standardised version) on the vertical axis against V2 (again standardized) on the horizontal axis. The matrix plot can be considered to be a pictorial representation of the correlation matrix itself. The plots on the main diagonal are perfect straight lines since here, a given variable is plotted against itself. For this reason the plots on the main diagonal are often left empty in such matrix plots.

The matrix plot corresponding to the correlation matrix $\mathbf{R}_1$ for our simulated data set with the variables V1, V2 and V3 are shown in **Figure 1a**. These three variables are uncorrelated to one other so that the scatterplot for the 1000 cases between each pair of variables looks circular in shape as explained above. When the three variables V1, V2 and V3 are plotted simultaneously in a three dimensional plot, the scatterplot assumes a spherical shape. Please refer to **Figure 1b**. Whatever the angle at which one chooses to look at the scatter, it always looks like a three dimensional sphere. This is the standard geometry for uncorrelated variables with equal variances.

The effect of high correlation on the geometry of the scatter

So far we have discussed the case when there was little or no correlation between the variables. We now discuss the scatter when there is a high correlation between the variables.

The coefficient of correlation coefficient r between two variables is always in the range -1 to $+1$, and cannot lie outside this range. The coefficient attains the value $+1$ or -1 if a *perfect* straight line is obtained in a scatterplot of the two relevant variables (or equivalently their standardised versions). The $+1$ is obtained when the slope of the line is positive, whilst -1 is obtained when the slope is negative. In these cases, one has a perfect linear relationship between the two variables. One variable can be perfectly predicted from the other and vice-versa.

For intermediate values of the correlation r, one gets an *elliptical* scatterplot for two variables in their standardised form. If we imagine the correlation between the two variables increasing gradually from 0 to 1, the scatterplot will gradually change from the circular shape when r is zero to elliptical for intermediate values of r. As the correlation increases, the eccentricity of the elliptical scatter will increase, i.e. the ellipse becomes thinner and longer, until r attains the maximum value of $+1$, when the ellipse flattens out to a perfect straight line as we have already mentioned above. The same thing happens when the correlation decreases from 0 to the minimum value of -1.

To see the effect of correlation on the geometry of the scatter, we now consider the set of variables W1, W2 and W3, where W1 and W2 are identical to V1 and V2 above, (ie. mark in Mathematics test and height respectively), whilst W3 is now the mark in the Statistics test. Logically, one would expect that W1 and W3 are substantially correlated to each other, whilst W2, the height, is not correlated to either. As before, therefore, we simulated a data file of 1000 cases each having three variables with correlation structure similar to W1, W2 and W3. The correlation between these three variables, which we shall also call W1, W2, W3, were then calculated from the simulated data set and are summarized in the correlation matrix $\mathbf{R}_2$ as follows:

$$\mathbf{R}_2 = \begin{array}{c|ccc} & \mathbf{W1} & \mathbf{W2} & \mathbf{W3} \\ \hline \mathbf{W1} & 1 & -0.031 & 0.948 \\ \mathbf{W2} & -0.031 & 1 & -0.010 \\ \mathbf{W3} & 0.948 & -0.010 & 1 \end{array}$$

As can be observed, the correlation between W1 and W3 is now over 0.9, whilst the correlations between W1 and W2, and W2 and W3 are very near 0. It is now instructive to examine the scatter of the variables W1, W2 and W3.

The matrix plot corresponding to the correlation matrix $\mathbf{R}_2$ for W1, W2 and W3 is shown in **Figure 2a**. In this case there is a strong correlation (0.948) between W1 and W3. This can be observed both from the top left or bottom right entries in matrix $\mathbf{R}_2$, as also from the corresponding scatterplots in Figure 2a. The scatterplot between W1 and W3 is not circular, but has the shape of a very long, thin ellipse with high eccentricity, and is practically a straight line. Because of the strong relationship between W1 and W3, the corresponding scatterplot is practically one dimensional in nature. On the other hand the correlations for the other pairs of variables (W1 - W2 and W3 - W2) are all small, and so their corresponding scatterplots look circular in nature. When the three variables are plotted simultaneously in three dimensions, it can be noticed that the scatter is no longer spherical, but is ellipsoidal in nature. Please refer to **Figure 2b** and **Figure 2c**. It is important to note however, that although the scatter is still three dimensional, the points are disposed mostly on a two dimensional plane which contains the W2 axis and makes about 45^0 with both the W1 and W3 axes. This plane can be seen end on in Figure 2b. There is very little variation normal to this plane, ie. in a direction going from left to right in Figure 2b. If in this Figure, one looks at the scatter from a different angle, say from the right or the left, rather than from the front, one can observe the scatter observed in Figure 2c. It is clear that most of the scatter (or *variation*) occurs in this plane, whilst little variation occurs in a direction perpendicular to this plane, as was shown in Figure 2b. The three dimensional spherical scatter of the uncorrelated variables V1, V2 and V3 in Figure 1b has now changed to the very flat ellipsoidal scatter of W1, W2 and W3 in Figures 2b and 2c, because of the large correlation between W1 and W3. This ellipsoid looks like a very flat rugby ball and is essentially two-dimensional!

An ellipsoid [10] is characterised by its three principal axes and their corresponding direction in space, the principal directions. Analogously to the two-dimensional ellipse, the ellipsoid has a major axis, where it is longest, an intermediate axis, and a minor axis,

where it is thinnest. These axes occur at right angles to one another, and their respective orientations in three dimensional space are called the principal directions. In Figures 2b and 2c, one is looking at different projections of the ellipsoidal scatter representing W1, W2 and W3. Figure 2b clearly shows the smallest axis of the ellipsoid, whilst Figure 2c displays the intermediate and major axes of the ellipsoid. The orientations of the largest, intermediate and smallest axes relative to the (W1, W2, W3) coordinate axes are called P1, P2 and P3 respectively. In mathematical jargon, these three orthogonal directions are called *principal components*, or *factors*, or even *eigenvectors*, whilst the square of the lengths of their corresponding axes are referred to as *eigenvalues*, usually denoted by λ_1, λ_2, and λ_3 respectively. The eigenvalue of a principal component is equal to the variance explained by that component. Consequently, the larger the eigenvalue, the more important is the associated principal component. As above, the principal components are usually arranged in descending order of eigenvalue, that is $\lambda_1 > \lambda_2 > \lambda_3$.

For our data set of the variables W1, W2 and W3, principal component analysis can be readily summarized in the following table:

Principal Component	**Axis of ellipsoid**	**Principal Components** *Orientation relative to W1, W2, W3.*	**Eigenvalues** *Variance explained or length squared of axis.*
P1:	along major axis:	P1 = .987*W1 + .987*W3,	$\lambda_1 = 1.95$;
P2:	intermediate axis:	P2 = 1.000*W2 ,	$\lambda_2 = 1.00$;
P3:	along minor axis:	P3 = .160*W1 – .160*W3,	$\lambda_3 = 0.05$.

Looking at the right hand side of the table, one can note that the sum of the variances (eigenvalues) of the three components 1.95+1.00+0.05 add up to 3.00, which is exactly equal to the total variance of the 3 variables W1, W2, W3 in their standardised form. (Note that each standardised variable has a variance of 1). Further, it is clear that the first two eigenvalues (1.95 and 1.00) are considerably larger than the eigenvalue of the last component (0.05), showing that our ellipsoid is very flat and that the third dimension can be ignored. In fact the first two components P1 and P2 explain (1.95+1.00)/3 or 98% of the total variation in the data.

The orientations of the principal components with respect to the W1, W2, W3 axes are given in the penultimate column. Thus the relation P1 = .987*W1 + .987*W3 shows that P1 is a line (or direction) lying in the W2=0 plane and making 45° with the positive directions of the W1 and W3 axes. The number 0.987 is called the *loading* [1, 11, 12] of W1 (or even W3 in this case) on P1. The loading can be considered to be the correlation between a given variable and the component, whilst the square of the loading, 0.987^2 or .974, implies that the component P1 explains 97.4% of the variation in W1. In fact, if one sums the squares of the loadings of a given component, one obtains the eigenvalue of that component, which stands for the total variation explained by the component. Thus for example, for P1, $0.987^2 + 0.987^2$ is equal to 1.95, the eigenvalue of P1. The fact that the

loadings of P1 are close to 1, imply that there is near perfect correlation of this factor with W1 and W3.

Similarly for the smallest axis, the relation P3 = 0.16*W1 – 0.16*W3 shows that P3 is a line (or direction) lying in the W2=0 plane and making 45° with the positive direction of the W1 axis and the negative direction of W3 axis. As stated previously, however, this is a weak component, with small loadings, and a small eigenvalue.

The second component satisfies P2 = 1.000*W2, making the second component practically identical to W2. This is not at all surprising since W2 did not have any loadings on P1 and P3. It was completely 'overlooked' by these two components which in turn explained *all* the variation in W1 and W3. The factor P2 would therefore have to account solely for all the variation in W2, making it identical to this variable. This phenomenon happened because originally, W2 was constructed to be uncorrelated to W1 or W3.

The principal components P1, P2 and P3 are orthogonal (perpendicular) to each other, and form a set of rectangular Cartesian axes just like W1, W2 and W3. In fact, principal component analysis can be considered to be a rotation from the W1, W2, W3 coordinate system to the system of principal components P1, P2 and P3. In our case, since P2 is identical to W2, the rotation takes place in the W1-W3 plane with the W2 (equivalently P2) axis fixed. The rotation from the set of original variables (W1, W3) to the principal components (P1, P3) is illustrated in **Figure 3**. This is essentially a replot of the top left graph in the matrix plot of Figure 2a, ie a plot between W3 and W1. The axis W2 (equivalently P2) is perpendicular to the plane of the diagram in Figure 3. The axes corresponding to W1 and W3 are rotated anticlockwise in the plane of the diagram itself through an angle of 45^0, so that they now point in the directions of the principal axes of the scatter. These directions, shown as dashed lines in Figure 3, are the two principal components P1 and P3, pointing respectively along the longest axis and the smallest axis respectively. The intermediate axis of the ellipsoid is normal to the diagram, along P2 (equivalently W2), and is not shown. One can visualise the geometry of this situation by imagining a flattened rugby-ball with the longest side pointing along P1, its width pointing along P2, and its flattened thickness pointing along P3.

It is clear that since the axis corresponding to P3 is so small compared to the others, one can effectively ignore P3 and consider the ellipsoid to be a two dimensional ellipse in the P1, P2 plane. Since the equations relating P1and P2 in terms of W1, W2 and W3 are known, the values of P1 and P2 can be computed for every case in the data-set. These *factor scores* [1, 11, 12] can then be plotted on a two-dimensional scatter-plot with P1 and P2 as axes. The old variables W1, W2, W3 have been adequately represented by a two-dimensional scatterplot in P1 and P2. The dimensionality of the system has been reduced from 3 to 2 with little or no loss of information. We have thus achieved a parsimonious description of our data set by means of a suitable rotation of coordinates. In this case, this was possible because the high correlation between the variables W1 and W3 makes one of them practically redundant.

Principal component analysis works best when there is substantial correlation between the variables. When variables are uncorrelated, like the set of variables V1, V2 and V3 given above, the spherical structure of the scatter (please refer to Figure 1b) ensures that no reduction in dimensionality can occur in this case, whatever rotation is performed. Principal component analysis would not be appropriate here. Since the variables are uncorrelated, there are no redundancies in the variables, and all dimensions (variables) have to be retained in this case.

A practical example of principal component analysis.

As a practical example of principal component analysis, we now present a data set containing the national track records for females and males in 55 countries [11, 12, 13].
The data set consists of 55 rows, each containing the following variables:

country	:	name of country;
f100	:	female record in seconds in the 100 metres event;
f200	:	female record in seconds in the 200 metres event;
f400	:	female record in seconds in the 400 metres event;
f800	:	female record in minutes in the 800 metres event;
f1500	:	female record in minutes in the 1500 metres event;
f3000	:	female record in minutes in the 3000 metres event;
fmar	:	female record in minutes in the marathon;
m100	:	male record in seconds in the 100 metres event;
m200	:	male record in seconds in the 200 metres event;
m400	:	male record in seconds in the 400 metres event;
m800	:	male record in minutes in the 800 metres event;
m1500	:	male record in minutes in the 1500 metres event;
m5000	:	male record in minutes in the 5000 metres event;
m10000	:	male record in minutes in the 10000 metres event;
mmar	:	male record in minutes in the marathon.

The numerical data therefore consists of a matrix of 55 rows and 15 columns, ie. 55 cases of 15 variables each. We would like to discover relationships between the various countries and the various events and to represent these relationships on suitable plots. For this end, principal component analysis will be used to elucidate the structure in the data.

The first step in a principal component analysis is the calculation of the correlation matrix. For this data matrix, the correlation matrix is given by:

Correlation Matrix.

	£100	£200	£400	£800	£1500	£3000	£mar	m100	m200	m400	m800	m1500	m5000	m10000	mmar
£100	1														
£200	.95	1													
£400	.83	.86	1												
£800	.73	.72	.90	1											
£1500	.73	.70	.79	.90	1										
£3000	.74	.71	.78	.86	.97	1									
£mar	.69	.69	.71	.78	.88	.90	1								
m100	.67	.73	.67	.63	.55	.60	.62	1							
m200	.77	.81	.73	.72	.66	.70	.71	.92	1						
m400	.80	.83	.81	.76	.70	.71	.66	.84	.85	1					
m800	.81	.82	.78	.79	.85	.86	.82	.76	.81	.87	1				
m1500	.79	.77	.77	.84	.88	.89	.83	.70	.78	.84	.92	1			
m5000	.73	.71	.74	.82	.86	.87	.81	.62	.70	.78	.86	.93	1		
m10000	.72	.72	.74	.82	.87	.87	.82	.63	.70	.79	.87	.93	.97	1	
mmar	.66	.63	.69	.78	.82	.82	.77	.52	.60	.71	.81	.87	.93	.94	1

The correlation matrix could be very awkward and cumbersome to present when there are many variables. For this reason, many programs present also *a sorted and shaded* correlation matrix [6, 7], which represents the correlation matrix succinctly in little space. The variables are sorted so that those with higher correlations are grouped together. The correlations are then represented by symbols: the denser the symbol, eg, the multiplication sign (X), closely followed by the addition sign (+), the larger is the magnitude of the correlation between two variables. Conversely, sparser symbols like the dash (–), the dot (.) and the space ( ) represent progressively smaller correlations. In our case one representation of the above correlation matrix is given by:

ABSOLUTE VALUES OF CORRELATIONS IN SORTED AND SHADED FORM

```
------------------------------------------------------------

      mmar      X
      £1500     XX
      m10000    XXX
      £3000     XXXX
      m5000     XXXXX
      £mar      +XXXXX
      m1500     XXXXXXX
      £800      +XXXX+XX
      m800      XXXXXXXXX
      m100      --++++++X
      m200      ++++++++XXX
      £200      ++++++++X+XX
      m400      ++X+++X+XXXXX
      £100      ++++++X+X++XXX
      £400      +X+++++XX++XXXX
```

THE ABSOLUTE VALUES OF THE MATRIX ENTRIES HAVE BEEN PRINTED ABOVE IN SHADED FORM ACCORDING TO THE FOLLOWING SCHEME:

```
                LESS THAN OR EQUAL TO        0.195
.         0.195  TO AND INCLUDING          0.390
-         0.390  TO AND INCLUDING          0.585
+         0.585  TO AND INCLUDING          0.780
X                    GREATER THAN          0.780
```

One can note here that all the correlations are quite high (>0.5) as evidenced by the dense symbols (X and +). The highest correlations however are observed between times for similar distances for males and females together. Thus for events longer than or equal to 800 metres (mmar, f1500, m10000, f3000, m5000, fmar, m1500, f800, m800) most of the correlations between them are represented by X, and are greater than 0.78. Similarly for the shorter events (m100, m200, f200, m400, f100, f400) most of the correlations exceed 0.78 and are therefore again represented by (X). These two groups of variables are still appreciably correlated together, as evidenced by the many plus signs (+) in the lower left part of the sorted and shaded correlation matrix.

The eigenvalues of the correlation matrix.

A very important part of the output of any factor analysis is the list of eigenvalues of the correlation matrix, since these give the square of the lengths of the principal axes of the ellipsoidal scatter. A 'histogram' of the eigenvalues, also called a *scree plot* [11, 14], is usually also given. This is a plot of the i'th largest eigenvalue λ_i against i. This plot is given so that one can visualise the relative sizes of the eigenvalues. The eigenvalues in this case, along with the associated scree plot, are given next:

```
Histogram of eigenvalues. (Scree Plot).

Eigenvalue      Histogram

1   11.9394     ************************************************************
2    1.1458     ******
3    0.5431     ***
4    0.4133     **
5    0.3195     **
6    0.1599     *
7    0.1130     *
8    0.0764     This and remaining eigenvalues are too small to appear.
9    0.0671
10   0.0634
11   0.0501
12   0.0434
13   0.0309
14   0.0197
15   0.0150
```

There are as many eigenvalues as there are variables originally, 15 in this case. Since each standardized variable has by definition a variance of 1.0, the total variance of these 15 variables (in standardised form) is exactly equal to 15. It can be shown mathematically that the sum of the eigenvalues is exactly equal to the total variance of the system, which is unaffected by any rotation of coordinates. In fact, the 15 eigenvalues given above have a sum of 15.0000 as predicted by the theory.

The eigenvalue of a given principal component is equal to the variance explained by that component. In this case, the largest eigenvalue, that is the eigenvalue of the first principal component, is 11.94 and therefore explains 11.94/15 or nearly 80% of the variation in the original data. Similarly the eigenvalue of the second principal component explains 1.15/15 or about 8% of the variation, whilst the third eigenvalue explains 0.54/15 or about 3% of the total variation. The first three factors together therefore account for 91% of the variance in the original data. This information is usually summarized in a table :

FACTOR	VARIANCE EXPLAINED	CUMULATIVE PROPORTION OF VARIANCE IN DATA SPACE	IN FACTOR SPACE	CARMINES THETA
1	11.9394	0.7960	0.8761	0.9817
2	1.1458	0.8724	0.9601	
3	0.5431	0.9086	1.0000	
4	0.4133	0.9361		
5	0.3195	0.9574		
6	0.1599	0.9681		
7	0.1130	0.9756		
8	0.0764	0.9807		
9	0.0671	0.9852		
10	0.0634	0.9894		
11	0.0501	0.9927		
12	0.0434	0.9956		
13	0.0309	0.9977		
14	0.0197	0.9990		
15	0.0150	1.0000		

THE VARIANCE EXPLAINED BY EACH FACTOR IS THE EIGENVALUE FOR THAT FACTOR. TOTAL VARIANCE IS DEFINED AS THE SUM OF THE POSITIVE EIGENVALUES OF THE CORRELATION MATRIX.

In the second column of the table above, the cumulative proportion of the total variance is the sum of the variance explained (eigenvalues) up to and including the factor, divided by the sum of all the eigenvalues. Thus the first factor explains 80% of the variance, the first two factors explain 87%, the first three explain 91% as shown above, and so on. The third column is similarly obtained by dividing by the cumulative sum of the eigenvalues by the sum of the first three eigenvalues only, rather than by the sum of all the eigenvalues (15.0). This is because we had requested the program to give us the first *three* principal components. (This will be discussed further on). The last column gives Carmines' theta, a parameter ranging from 0 to 1. The fact that its value of 0.9817 is very near to 1 implies that the factor analysis on our data set was successful, as a large proportion of the variance was explained by very few factors. A similar parameter is Cronbach's alpha, a measure of internal consistency of the variables in a data set. In our case alpha is found to be 0.9815 which is very near the maximum value of 1. This again implies that there are high correlations between our variables, and that parsimony is successfully achieved with the first few components.

The principal components or factors.

A principal component is described by the *factor loadings* [1, 11, 14] or, equivalently, *loadings* of the original variables on it. As explained above, a principal component can be uniquely specified by these loadings. The loading of an original variable on a principal component can be interpreted as the correlation between them. The loadings for the first three principal components are given in the following table:

UNROTATED FACTOR LOADINGS FOR PRINCIPAL COMPONENTS

VARIABLE	FACTOR 1	FACTOR 2	FACTOR 3
f100	0.868	0.253	0.283
f200	0.869	0.335	0.242
f400	0.881	0.138	0.343
f800	0.903	-0.111	0.191
f1500	0.913	-0.294	0.129
f3000	0.920	-0.252	0.073
fmar	0.874	-0.199	-0.029
m100	0.777	0.491	-0.311
m200	0.853	0.403	-0.177
m400	0.890	0.291	-0.122
m800	0.945	0.030	-0.094
m1500	0.954	-0.109	-0.110
m5000	0.925	-0.247	-0.138
m10000	0.930	-0.249	-0.158
mmar	0.865	-0.353	-0.139
VP	11.939	1.146	0.543

THE VP IS THE VARIANCE EXPLAINED BY THE FACTOR.
IT IS COMPUTED AS THE SUM OF SQUARES FOR THE
ELEMENTS OF THE FACTOR'S COLUMN IN THE FACTOR
LOADING MATRIX.

As pointed out above, the sum of the squares of the loadings of a given component is equal to the square of the length of the associated principal axis, that is, its eigenvalue or the variance explained by that component. Thus, for the first component under the heading 'Factor 1', we have that $.868^2+.869^2+....+.865^2$ is equal to 11.939, the eigenvalue of the first principal component. This component is defined by the equation

$$P1 = 0.868*f100+ 0.869*f200+ ..... + 0.865*mmar,$$

where all the variables on the right hand side are in their standardised version. The second and third components, P2 and P3, are similarly defined from the two columns on the right.

Since the correlations (loadings) of P1 with the 15 variables are all positive, one can interpret P1 to be a measure of the overall athletic prowess of a country. Countries with above average times on the majority of events tend to have high positive scores on this component. Conversely, countries which are strong in track events, and have shorter times, tend to have high negative scores on this component.

The second principal component under the heading 'Factor 2' has smaller loadings on the variables, but is still readily interpretable. It has positive loadings on the short distance events (m100, m200, m400 f100, f200 and f400) and negative loadings on the longer events. It therefore contrasts sprints with the longer distance times. Countries which are poor in sprint but do better in the longer distances tend to have high positive scores on this factor, whilst countries which do better in the sprint than in the long distances will have high negative scores.

The third component under the heading 'Factor 3' has positive loadings on most female events, and negative loadings on the male events. This factor therefore differentiates between those countries where females do worse than males from those countries where females do better. Countries where females fare worse than males have a high positive score on this factor, whilst countries where females do relatively better than males have a high negative score.

In the above we decided to retain only the first three components. In general, how does one decide how many components are needed to provide an adequate summary of the given data set? There are various *ad hoc* rules for this, the most common being: [11, 14] :

i) Retain only the components with eigenvalues larger than 1: components with eigenvalues less than one account for less variation than an original standardised variable. This is the default method in most computer programs.
ii) Include just enough components to explain some relatively large percentage of the total variation. Figures between 70% and 90% have been suggested although this will become smaller as the number of variables increases.
iii) The scree plot of the eigenvalues is inspected for a possible 'elbow' in the curve. Eigenvalues above this elbow are considered large and their principal components are retained.
In our case, we decided to retain three components because they account for 91% of the variation, and all three components can be readily interpreted.

The factor scores.

The three most important principal components P1, P2 and P3 have now been extracted from our data, and we have explicit equations for them in terms of the original standardised variables. It is therefore possible to find the values of P1, P2 and P3 for every case (country) in the data set. The values of P1, P2 and P3 obtained for each case are known as *factor scores*[1, 11, 14]. The factor scores of each country for the first three principal components are listed in the following table:

Country	Symbol	Factor Scores 1	2	3
The United States	usa	-1.38	-1.22	-0.12
East Germany	eg	-1.24	-0.83	-1.64
Russia	rus	-1.24	-0.66	-1.00
Great Britain	gb	-1.19	-0.53	-0.23
West Germany	wg	-1.12	-0.55	-0.75
Italy	it	-1.00	-0.55	0.75
Poland	po	-0.95	-0.68	-1.08
Australia	aus	-0.93	-0.50	-0.13
Czechoslovakia	cz	-0.89	-0.51	-2.32
Canada	ca	-0.88	-0.70	-0.99
France	fra	-0.83	-0.63	0.05
Finland	fin	-0.79	-0.05	-1.02
Belgium	bel	-0.73	0.19	0.23
Sweden	swe	-0.70	-0.10	-0.29
Netherlands	net	-0.69	0.48	-0.55
Rumania	rum	-0.65	0.81	-0.86
New Zealand	nze	-0.64	0.87	-0.07
Switzerland	swi	-0.61	0.42	0.42
Hungary	hun	-0.55	-0.07	-0.22
Kenya	ke	-0.55	0.57	1.08
Norway	nor	-0.47	1.33	-0.30
Denmark	de	-0.46	0.58	0.15
Austria	aut	-0.44	0.26	-0.84
Ireland	ire	-0.41	1.01	-0.15
Spain	spa	-0.39	0.83	1.16
Brazil	bra	-0.35	-1.09	0.88
Japan	jap	-0.27	0.57	1.43
Portugal	por	-0.24	1.63	1.09
Mexico	mex	-0.16	0.86	0.76
Colombia	co	-0.05	0.41	0.61
Chile	chl	0.03	0.44	1.39
Israel	isr	0.06	0.66	-0.39
Greece	gre	0.08	-0.36	1.31
Taiwan	tai	0.10	-0.57	-1.77
Argentina	arg	0.16	-0.64	0.27
India	ind	0.17	0.55	0.71
China	chi	0.21	0.80	0.77
Bermuda	ber	0.24	-1.88	-0.23
South Korea	skor	0.29	0.16	1.64
Luxemburg	lux	0.30	0.42	1.71
Turkey	tur	0.37	1.36	1.19
North Korea	nkor	0.45	2.16	-1.00
Burma	bur	0.75	0.56	0.15
Philippines	phi	0.76	-0.64	-0.79
Dominican Republic	dom	0.82	-2.18	1.01
Malaysia	mal	0.83	-1.39	0.99
Thailand	tha	0.97	-1.30	0.04
Costa Rica	cri	1.00	1.39	0.76
Indonesia	ndo	1.00	-0.60	-0.33
Singapore	sin	1.05	-0.78	0.05
Guatemala	gua	1.20	0.45	0.07
Papua & New Guinea	pap	1.61	-0.12	0.01
Mauritius	mau	1.74	-0.07	-0.91
Samoa	sam	3.18	-2.77	0.53
Cook Islands	cook	3.43	2.21	-3.19

==========

Although the original data set was in alphabetical order of country, we decided to list the countries in ascending order of the first factor score. As pointed out above, the first principal component measures the overall athletic prowess of a country with weaker countries having higher scores on this component. The above list is therefore *ordered*, with the stronger nations like the United States, East Germany, Russia, Great Britain etc. at the top of the list, right down to the weaker nations. We have therefore achieved a ranking of the countries in the original data set.

Similarly, the second factor score differentiates countries of similar ability according to whether they are better in the sprints than in the long distance events. Countries who are relatively stronger in sprints have negative scores on the second factor, whilst those who do relatively better in long distances have positive scores on this factor. To make this point clearer, one can plot the first two factor scores (given above) against each other. This scatterplot is given in **Figure 4**. Here, the factor score of the second principal component (Factor 2) is plotted on the vertical axis against the factor score of the first principal component (Factor 1) on the horizontal axis. In this figure, therefore, stronger nations appear to the left whilst the weaker nations appear to the right. Countries who do relatively better in sprints appear in the lower half of the plot, whilst those who do relatively better in the longer distances appear in the upper half.

Thus for example on the right hand side of the plot in Figure 4, the Cook Islands (labeled as *cook*) and Samoa (labeled as *sam*) both have high positive scores for Factor 1, indicating the relatively low overall athletic standard in these two countries. These are however differentiated by the second factor score. The Cook Islands have a high positive score on the second component, so that they do better in the long distance events. Conversely, Samoa has a high negative score, so it does relatively better in the sprints than in the long distances. Similarly looking at the central third of the table from bottom to top, one has the Dominican Republic, Bermuda, Malaysia and Thailand (*dom, ber, mal, tha*) who have very similar record profiles with a higher standard of sprinting , right up to North Korea, Turkey and Costa Rica (*nkor, tur, cri*) who are of comparable strength as the previous group, but are relatively better in the longer distances. In a similar way, the stronger nations in the left hand third of the table are separated by the second component into countries where the sprints predominate as in the United States, East Germany, Russia, Canada and Brazil (*usa, eg, rus, ca, bra*) and those who are stronger in the longer distances like Portugal, Norway, Ireland and New Zealand (*por, nor, ire, nze*). The factor scores of the first two principal components, which have been given above and plotted in Figure 4, therefore provide an effective two-dimensional summary of the original data set.

Similar plots can be given for Factor 3 scores versus Factors scores 1 or 2. They could even be plotted simultaneously in a three dimensional scatterplot. The third factor score, Factor 3, is the last column in the ranking list of nations given above. As pointed out above, the third factor distinguishes countries of comparable standards by the relative performance of males and females. Considering the stronger nations in the above list, females do relatively better than males in countries with negative third factor scores, such

as East Germany, Russia, Poland, Czechoslovakia and Canada. In the USA, Australia and France, this score is nearly zero, so the two genders have comparable performances in these countries. Conversely, males do better than females in countries like Italy with a high positive score on the third factor.

As can be seen from the above considerations, our original data set with 15 variables has been successfully summarized by just three components and has been effectively represented by a scatterplot between the scores on the first two components. Data reduction has been efficiently performed on this data set.

Rotation of the principal components.

In principal component analyses, rotations are sometimes performed also on the principal components themselves in a bid to obtain simpler factors. The goal is to make the loadings for each factor either large or small, not intermediate. There are two types of rotations [6, 8]. In *orthogonal* rotations, the resulting factors are still perpendicular (orthogonal) to one another and the factors are not correlated together. Alternatively, one could allow these rotations to be *oblique*, rather than orthogonal, so that the factors are allowed to be correlated between themselves. With oblique rotation there is a greater tendency for each variable to be associated with a single factor, thereby simplifying interpretation of the factors. Plots are sometimes drawn of the rotated factor loadings: the loadings of the variables for one factor are plotted against those of another factor in a two-dimensional scatterplot.

As an example, an oblique rotation was performed on the first two principal components of the track data. One factor is made up of the long distance events whilst the second factor comprises the shorter distances, as in the sorted and shaded correlation matrix which was presented above. Since the correlation between these two factors is quite high (0.746), a scatterplot for the factor scores of these oblique factors would be not be as informative or easily interpreted as the analogous plot in Figure 4, where the scores were for the necessarily orthogonal (and hence uncorrelated) principal components. Conversely, if the correlation between the two oblique factors was low, the plot for the factor scores would be very similar to the plot given in Figure 4 .

Oblique rotation of factors is very popular in applications of social science and psychology[14], where the emphasis is on the correlational structure of the *variables* rather than on the distinction between the cases.

Principal component analysis, factor analysis and other multivariate techniques.

The two terms *principal component analysis* (PCA) and *factor analysis* (FA) are sometimes used interchangeably, but this is not exactly correct. In fact, principal component analysis is the simplest type of factor analysis. A default run of the factor analysis option in most computer packages is usually a principal component analysis. For this reason principal components are often referred to as *factors*, but one should not

forget that factor analysis embraces a whole range of techniques for extracting factors from data. Of these techniques, principal component analysis is the simplest and the most intuitive. In all techniques of factor analysis, however, suitable rotations of coordinates are performed from the old variables to the final extracted factors.

Many multivariate techniques in statistics, like multidimensional scaling [8, 11], cluster analysis for variables[7] and cases [7, 8, 11], and correspondence analysis [7, 11] for frequency tables, resemble principal component analysis [6, 8, 11] and factor analysis [6, 8, 11] in that they try to achieve a parsimonious and faithful description of the underlying data. Data reduction is a common goal to most procedures in multivariate statistics. These multivariate techniques are described in many standard references, some of which are cited below.

Statistical Analyses.

The above statistical analyses were performed with BMDP, the Bio-Medical Data Package [6, 7]. In particular we used program 1D to obtain the simulated data sets, and program 4M for principal component analysis [6]. The sorted and shaded correlation matrix can also be obtained from the program 1M for cluster analysis of variables [7]. The above analyses can also be easily performed with other programs such as SPSS [8, 9, 14], which was used to plot the graphs in Figures 1, 2 and 4.

Suggestions for further reading.

The subject of correlation is treated in many elementary textbooks of statistics [3], as also in the biostatistical texts [2, 4, 5] and in the excellent archaeological text by Shennan[1]. This text also has a very readable exposition of principal component analysis and other multivariate techniques, and is strongly recommended for the non-mathematical reader. The reference manuals of statistical software packages like BMDP [6, 7] or SPSS [8] also describe most multivariate statistical techniques, of which they give many practical examples, along with clear, annotated output. These manuals are an excellent sourcebook of such techniques, and are strongly recommended to the general reader.

The book by Bryman [9] describes how to use SPSS for Windows to perform numerous statistical techniques, and clearly explains the SPSS output. The book by Tacq [14] also gives very good accounts of many multivariate techniques, how to execute them with the programming language of SPSS, and how to interpret the output of the program. This book gives many interesting examples from the social sciences, and also gives some mathematical and numerical details for the more mathematically oriented reader.

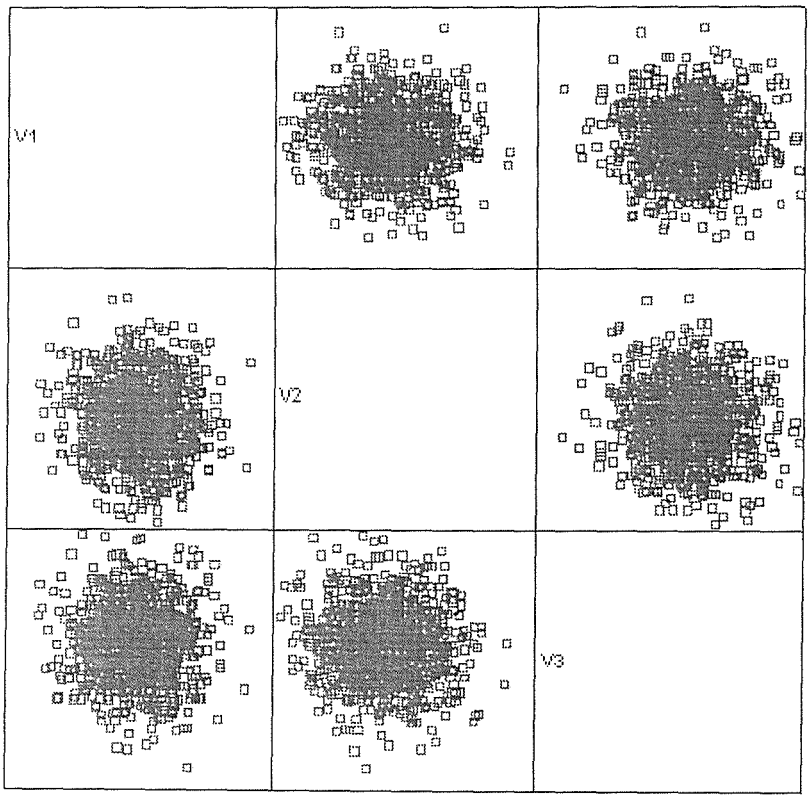
There are many good books on multivariate statistics, of which we cite a few [11, 12, 15, 16, 17]. These texts are more mathematically disposed, those by Flury[16] and Manley[17] being slightly easier than those by Everitt[11], Johnson[12] and Morrison[15]. These five books contain numerous applications of principal component analysis and other multivariate

techniques to different disciplines, and should therefore be of interest even to non-mathematical readers.

References.

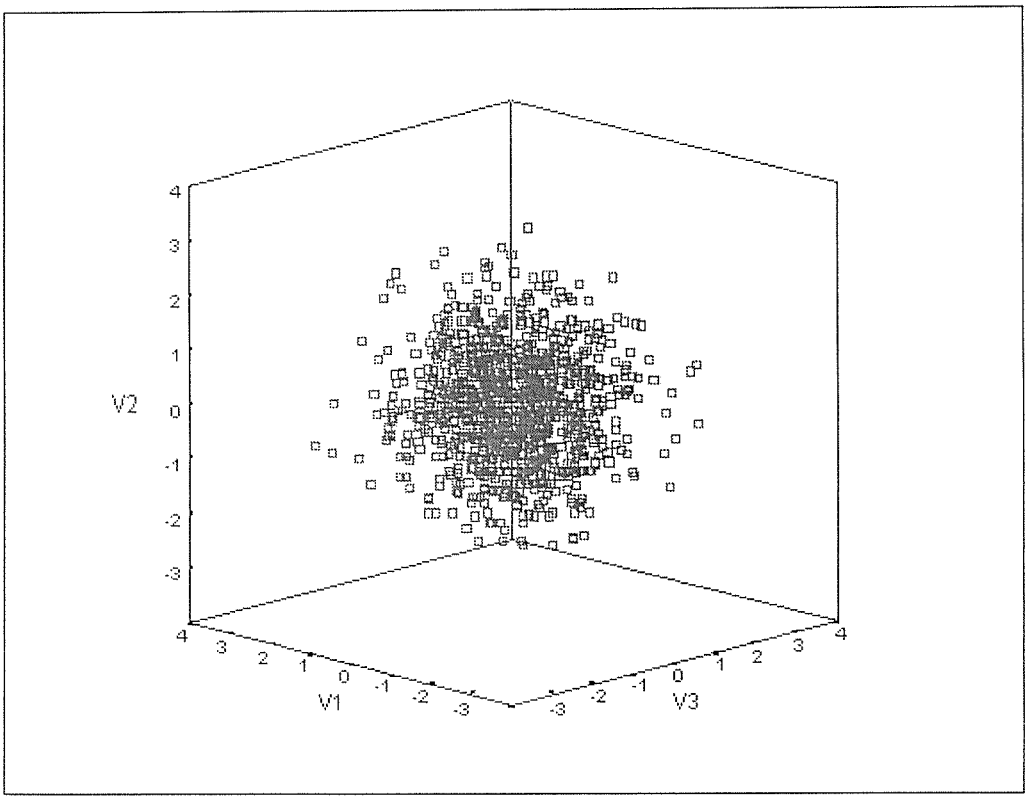
1. Shennan S. *Quantifying archaeology.* Edinburgh, Edinburgh University Press. 1988.
2. Sokal R.R. and Rohlf F.J. *Biometry.* 3rd Edition. New York, W.H. Freeman and Company. 1995.
3. Freund J.E. & Simon G.A. *Modern elementary statistics.* 9th Edition. New York, Prentice-Hall. 1997.
4. Zar J.H. *Biostatistical Analysis.* 2nd Edition. New Jersey, Prentice-Hall. 1984.
5. Sokal R.R and Rohlf F.J. *Introduction to biostatistics.* 2nd Edition. New York, W.H. Freeman and Company. 1987.
6. Dixon W.J. *BMDP Statistical Software Manual, Volume 1.* Berkeley, University of California Press. 1992. See in particular program 4M, mostly pages 329 to 362.
7. Dixon W.J. *BMDP Statistical Software Manual, Volume 2.* Berkeley, University of California Press. 1992. See in particular program 1M, mostly pages 867 to 878.
8. Norusis M.J. *SPSS Base System Use's Guide.* Chicago, SPSS Inc. 1990.
9. Bryman A. & Cramer D. *Quantitative Data Analysis with SPSS for Windows.* London, Routledge. 1997.
10. Edwards C.H. and Penney D.E. *Calculus with analytical geometry.* 4th edition. London, Prentice-Hall International. 1994.
11. Everitt B.S. and Dunn G. *Applied Multivariate Data Analysis.* London, Edward Arnold. 1991.
12. Johnson R.A. and Wichern D.W. *Applied Multivariate Statistical Analysis.* London, Prentice-Hall International. 1992.
13. Dawkins B. 'Multivariate Analysis of National Track Records', *The American Statistician*, **43** 1989, 110-115.
14. Tacq J. *Multivariate Analysis Techniques in Social Science Research.* London, Sage Publications. 1997.
15. Morrison D.F. *Multivariate Statistical Methods.* 2nd edition. New York, Mc Graw-Hill. 1976.
16. Flury B. and Riedwyl H. *Multivariate Statistics.* London, Chapman and Hall. 1988.
17. Manley B.F.J. *Multivariate Statistical Methods: A Primer.* London, Chapman and Hall. 1986.

Figure 1a : Matrix plot of the variables V1, V2 and V3. Their correlation matrix $\mathbf{R}_1$ is given in the text. Since the correlation between the variables is low, the scatter plot for each pair of variables assumes a circular shape. When plotted together in a three dimensional plot, the scatter assumes a spherical shape as shown in Figure 1b.



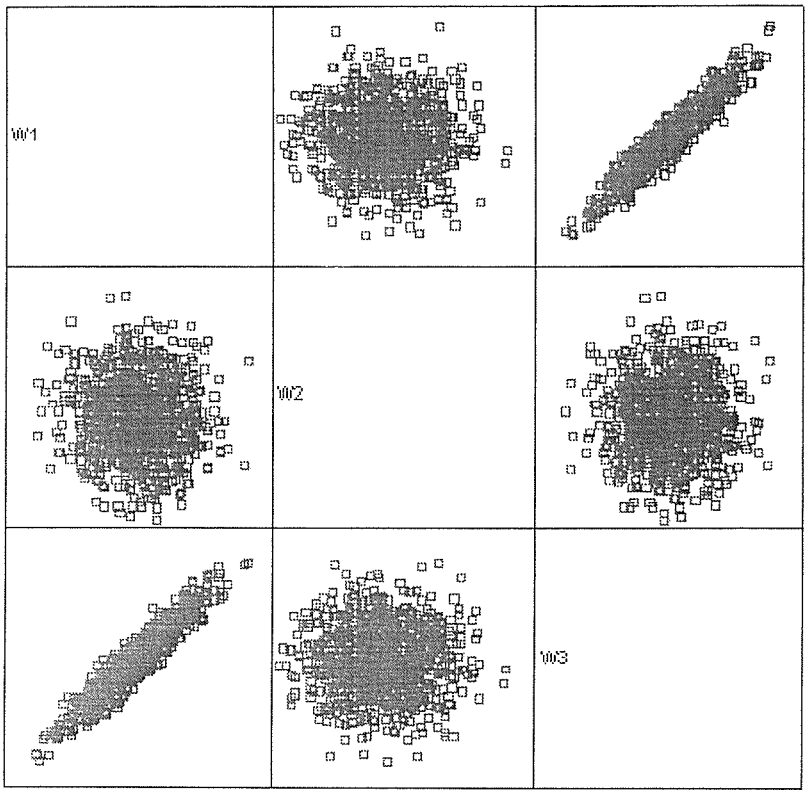
====

Figure 1b: Three dimensional scatterplot of the variables V1, V2 and V3. Since the variables are not correlated the scatter assumes a spherical shape.



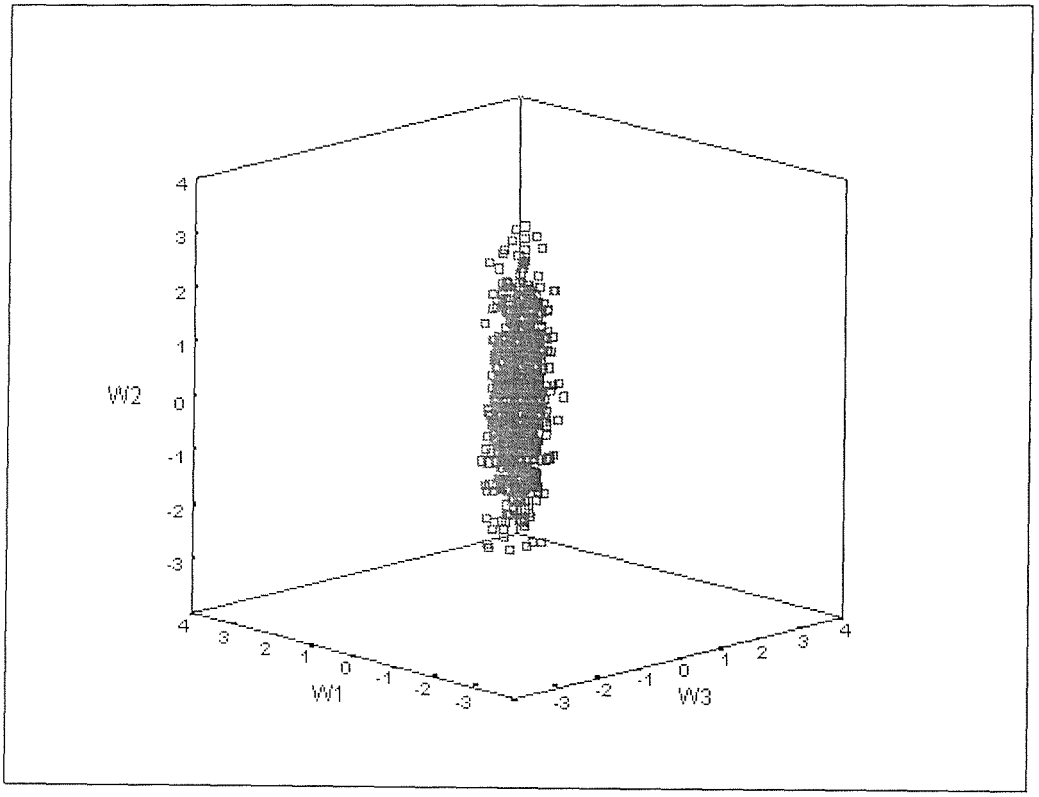
=====

Figure 2a: Matrix plot of the variables W1, W2 and W3. Their correlation matrix $\mathbf{R}_2$ is given in the text. Since the correlation between the variables W1 and W3 is high, the scatterplot for these two variables assumes the shape of a long, thin ellipse as shown in the top right or bottom left scatter plots. When plotted together in a three dimensional plot, the scatter assumes the form of a flat ellipsoid which is nearly two-dimensional. Two views of this ellipsoid are shown in Figures 2b and 2c.



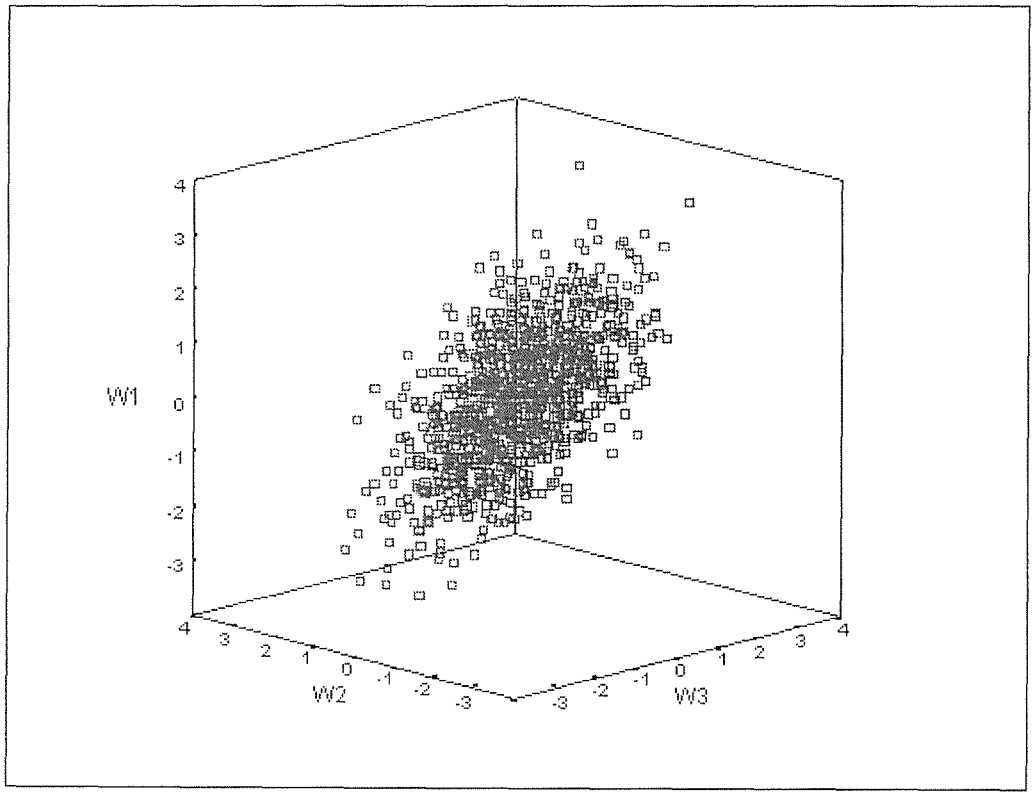
=====

Figure 2b: One view of the ellipsoidal scatter for the three variables W1, W2 and W3. From this view one can appreciate the flat nature of this ellipsoid. The scatter can be safely considered to be two dimensional as there is little scatter or variation in going from left to right in this Figure. Most of the scatter can be observed if one looks from the right or the left. The scatter from this point of view is shown in Figure 2c.



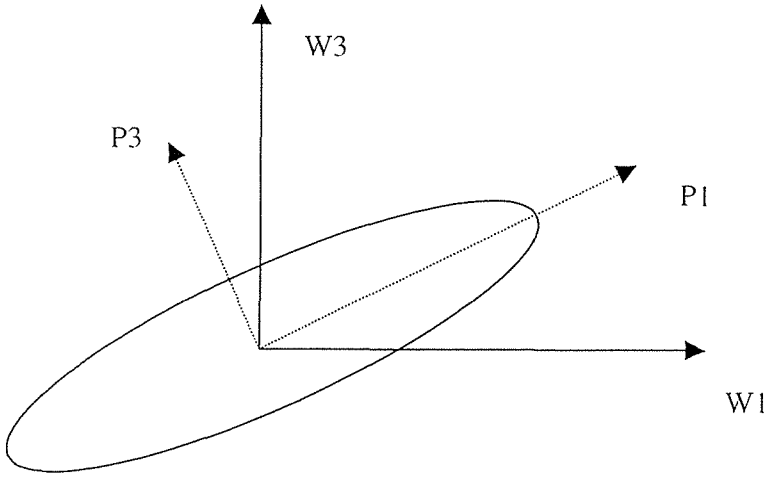
====

Figure 2c: A glimpse of the scatterplot between W1, W2 and W3 from another point of view, showing more clearly the two-dimensional plane containing most of the scatter. The scatter in this plane is elliptical with the major axis pointing from bottom left to top right, and with the other axis parallel to W2. These directions are called the principal directions or *principal components*. A third principal component exists normal to these two, but this is not important as little variation occurs in this direction, as was shown in Figure 2b.



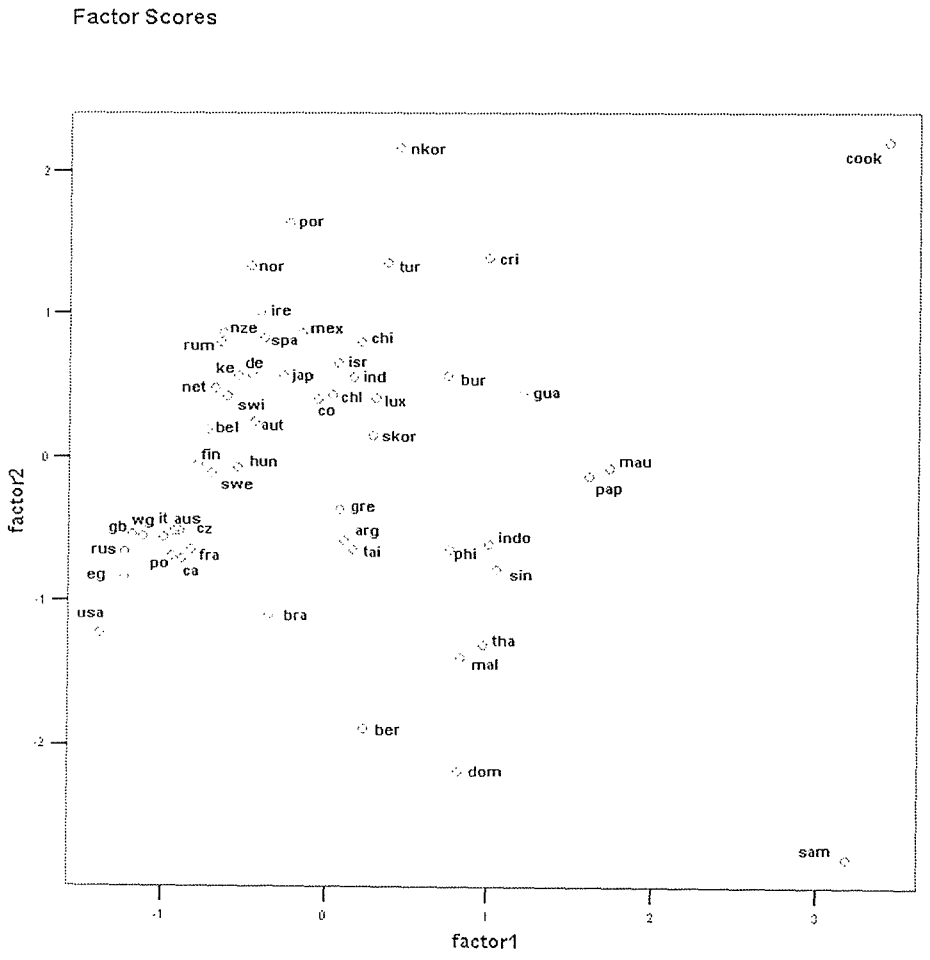
====

Figure 3: An illustration of the rotation of coordinates in principal component analysis. The scatterplot for W1 and W3 is shown as an ellipse. Please see the corresponding scatterplot in the matrix plot of Figure 2a. This ellipse is inclined at an angle of about 45^0 to the W1 and W3 axis. The principal components are defined to be the directions defined by the major and minor axes of the ellipse. These are shown as dashed lines and are labeled P1 and P3. The direction W2 is normal to the plane of the paper. Since W2 is not correlated with either W1 or W3, the principal component P2 is identical to W2, and is not affected by the rotation in the W1, W3 plane. Positions of points in the scatter can then be conveniently referred to the P1, P2 and P3 system of coordinates rather than to the old coordinates W1, W2 and W3. As shown in the diagram, little variation occurs along the P3 axis (the direction where the ellipse is very narrow). This component can therefore be safely ignored, retaining only the two components P1 and P2. This leads to a simpler description of the data.



======

Figure 4: Scatterplot of the first two factor scores for the athletic records data. This is a plot of a set of 15-dimensional observations in the space of the first two principal components. The first principal component represents a measure of the overall athletic prowess of a country: countries with above average times on the majority of events tend to have high scores on this component. So weak nations (eg the Cook Islands and Samoa) appear on the right of the scatterplot, whereas the stronger nations (eg the United States, Russia and East Germany) appear to the left. The second component mainly contrasts performance in sprints with the performance in the middle and long distances. Countries particularly good at the longer distances (eg Norway and Portugal) tend to have high scores on this component, and appear towards the top of the plot. Countries with good sprinting times (eg. the United States, Bermuda, the Dominican Republic) appear towards the bottom of the plot. The original data set with 15 variables has therefore been effectively and parsimoniously represented by a scatterplot of the factor scores derived from its first two principal components.



=====

Iso-Taxi Geometry: A New Approach

Anne C. Yancey

Math Anxiety

"*Let no one ignorant of geometry enter here*".

These words were inscribed over the doors of Plato's Academy, and such a forbidding perception of the field of Mathematics has permeated society since those days. For more than 5500 years, from Ancient Greece to the beginning of the third millennium, ordinary people have passed on this fear of Mathematics to their children. In ancient times, the public could have been informed clearly about the field of mathematics, the methods of mathematical thoughts, mathematical objects and their properties, and how these relate to nature and society. Unfortunately, however, Mathematics was conveyed as being both difficult and abstract, and thus, it became generally accepted that Mathematics is not for the average mind. As a result, instead of well-understood strategies of investigation, something quite awkward and unattractive appeared. The common widespread assumption that has continued through present day is that people are either good with words or with numbers, not both.

Generally once a negative attitude and an anxiety are formed, it becomes quite difficult to change. These feelings often persist into adult-life with far-reaching consequences in the form of avoidance of mathematics, distress, and interference with conceptual thinking and memory processes. Although mathematics aims at right answers, these answers can be reached through open-ended problems - mathematics being experienced as a series of discoveries to be made by the learner. Rather than mathematical methods and rules, learners need to acquire abilities to analyze, question, test and find solutions. Thus, developing knowledge and skills relating to the processes which can later be applied in any situation. If such a different approach could be taken in the early stages of Mathematics education, an approach that could be built upon in later years, then great steps could be taken at relieving math anxiety and reducing the common fear. One such approach to this teaching style is examined in *Iso-Taxi,* or *Chinese Checker* geometry.

A New Geometry

There are only three regular polygons that will tessellate a plane - the square, equilateral triangle, and regular pentagon. Of these, only two form a regular grid. Extensive work has been done with the square grid in *Taxi-Cab* geometry, where the placement of points and the movement between points is restricted to the lines of the grid as if the lines represent the city streets and the points represent taxi cabs. If we instead consider the plane of equilateral triangles, and restrict the points and movements similarly on this new isometric grid, a geometry quite different than either Euclidean or *Taxi-Cab* can be developed. In previous work this new geometry has been referred to as *Iso-Taxi* geometry, however, if we examine the physical appearance of our grid, it is noticed to resemble the

playing surface of a popular children's game. Hence, for our purposes, we will refer to it as *Chinese Checker* geometry.

Before looking at any geometric properties or geometric sets, we must first understand the isometric *Chinese Checker* plane and the naming of *Chinese Checker* points. The three axes that exist in our plane, *x, y* , and *y'*, are each separated by 60 degrees and divide the plane into hextants which are numbered I-VI in a counter-clockwise manner. (Figure 1) *Chinese Checker* points are then named according to their position relative to the *x*- and *y*-axes (Figure 2), and the slanted *y*-axis creates a situation such that points with the same *x* coordinate do not lie in a vertical line. Instead the geometric solution to an equation such as *x = -2* is a slanted line parallel to the *y*-axis. Also, the *y'*-axis is used only in determining the orientation of points which, in essence, chooses of the appropriate distance equation. (This idea will be further explained.)

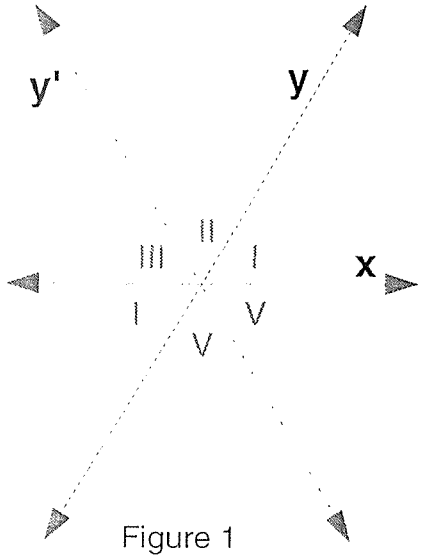


Figure 1

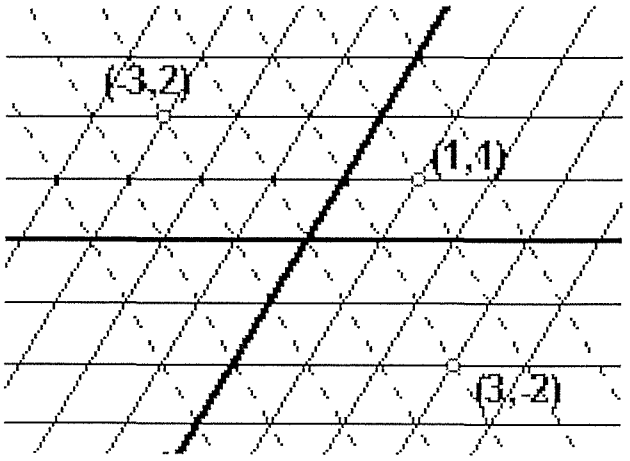


Figure 2

When comparing Euclidean geometry, *Taxi-Cab* geometry, and *Chinese Checker* geometry, points, lines, and angles are the same. The significant difference in the three is the distance equation. The restriction of movements between two arbitrary points to the lines of the playing surface leads to a situation such that the distance equation is dependent upon the relative position of the two points. From this concept, three separate distance equations have been developed:

I. If the points have a *I-IV* orientation, then $d_c = |x_1 - x_2| + |y_1 - y_2|$.

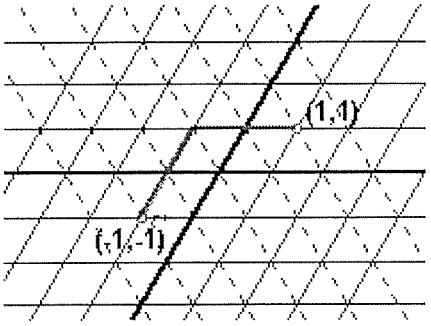
II. If the points have a *II-V* orientation or lie on a line parallel to the *y*- or *y'* axis, then $d_c = |y_1 - y_2|$.

III. If the points have a *III-VI* orientation or lie on a line parallel to the *x*-axis, then $d_c = |x_1 - x_2|$.

In each of the following examples, the orientation of points A and B were determined, and the appropriate distance equation was used.

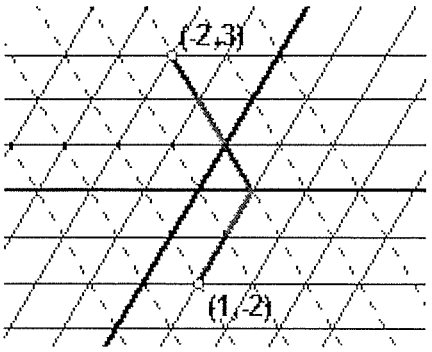
Point $A=(1,1)$ and point $B=(-1,-1)$ have a *I-IV* orientation.

$$\begin{aligned} d_C &= |x_1 - x_2| + |y_1 - y_2| \\ &= |1-(-1)| + |1-(-1)| \\ &= |2| + |2| \\ &= 4 \end{aligned}$$



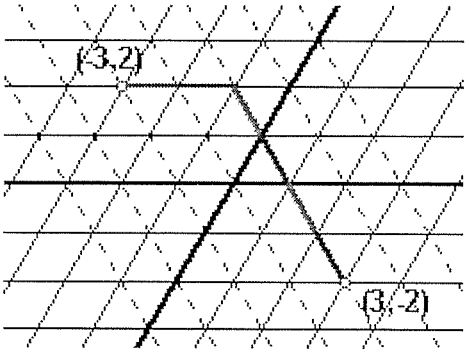
Point $C=(-2,3)$ and point $D=(1,-2)$ have a *II-V* orientation.

$$\begin{aligned} d_C &= |y_1 - y_2| \\ &= |3-(-2)| \\ &= |5| \\ &= 5 \end{aligned}$$



Point $E=(-3,2)$ and point $F=(3,-2)$ have a *II-V* orientation.

$$\begin{aligned} d_C &= |x_1 - x_2| \\ &= |(-3)-3| \\ &= |-6| \\ &= 6 \end{aligned}$$



As previously mentioned, the concepts of points, lines, and angles are unchanged. This is because they are independent of the distance formula. Certain geometric properties and geometric sets, however, are dependent upon the distance formula and are therefore affected by any change in the metric. Consider the circle.

A circle consists of all the points equidistant from a fixed point called the center. When such points are found on the isometric grid, the resulting figure is what we commonly call a hexagon. Furthermore, if we define *pi to* be the ratio of circumference to diameter the value of *pi* can be calculated to be 3. (Figure 3) The area of the circle is also changed when it exists on the isometric grid. Using the unit equilateral triangle that tessellates the plane as a unit of area, we can count the number of such units it takes to tessellate the interior of the figure - 24. Subsequently, we can derive a formula for *Chinese Checker*-area, $A_{C.} = 2\pi r^2$.

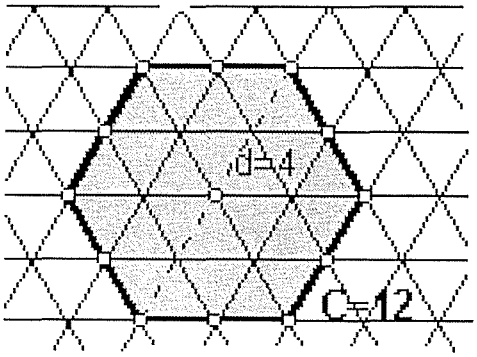


Figure 3

Further Consideration

While there are an endless number of geometric properties and geometric sets that can be examined in this manner, and while this is only one possible educational model, this glance into *Iso-Taxi* geometry, when considered as *Chinese Checker* geometry, allows us to see the impact that teaching style can have on the interest of students. If during instruction on this new geometry students are allowed a hands-on approach with the use of a Chinese Checker board and playing marbles, the subject can be presented in a manner that is much less foreign to the students. Hence, the common fear that is associated with new mathematical topics can be lessened, and the first steps at relieving math anxiety can be taken.

Basic Properties of Cayley Graphs

Andrew Duncan and Andrew Cortis

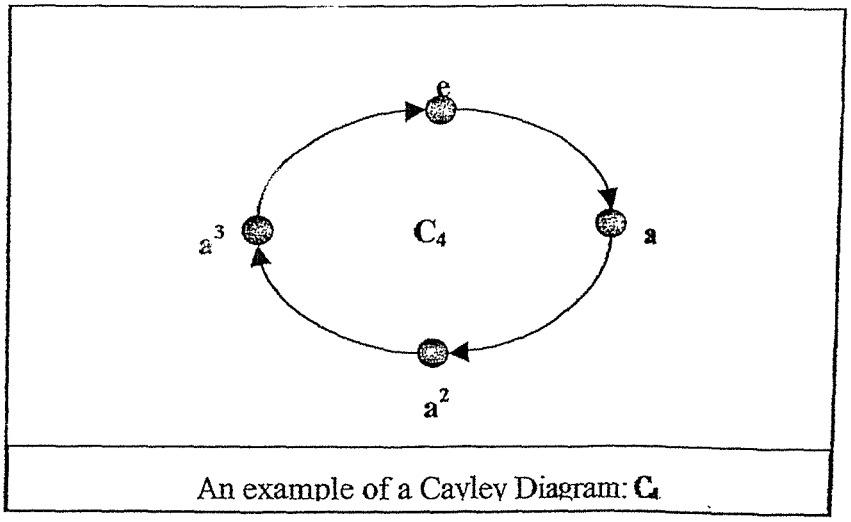
Introduction

Cayley Diagrams are one of many representations of finite groups. They provide a means of representing a group diagrammatically and various properties of groups including commutativity can be extracted from the graph. The Cayley diagram also provides sufficient information to test for isomorphism between groups, and thus is a useful tool for recognizing the type of a given group. This technique of representing groups as graphs was introduced by Cayley in 1878. The Cayley diagrams described below are a variant of the actual Cayley diagram and are referred to as Cayley Digraphs.

Let $S = \{g_1, g_2, \ldots, g_n\}$ be a set of distinct elements and let $G = < g_1, g_2, \ldots, g_n >$, i.e. G is the group generated by the set S.

We can define a relation ~ on G such that $a \sim b$ iff $b = g_i a$, where $g_i \in S$. Then the Cayley Digraph Cay(G,S) is the digraph formed from the relation ~, where the vertex set of the graph is the group G.

Informally, the Cayley Digraph is a digraph with the elements of G as vertices, and there is an edge from a to b if $b = g_i a$, where g_i is some generator of G.



An example of a Cayley Diagram: C_4

Here are various simple properties of Cayley Digraphs:

1. Let a be a vertex in Cay(G, S), and $|S| = n$, then $\deg^+(a) = \deg^-(a) = n$.

   Proof:

   Let $a \in G$, then $g_1^{-1}a, g_2^{-1}a, g_3^{-1}a, \ldots, g_n^{-1}a$ are n distinct elements in G (by closure), since suppose that:

   $$g_i^{-1}a = g_j^{-1}a, \quad g_i, g_j \in S,$$
   $$\Rightarrow g_i^{-1} = g_j^{-1} \Rightarrow g_i = g_j,$$

   since inverse elements are unique, contradicting the premise that elements of S are distinct.

   $$g_i^{-1}a \sim a, \; i = 1 \ldots n. \quad \therefore \deg^-(a) = n$$

   Similarly,

   $g_1a, g_2a, \ldots, g_na$ are n distinct elements in G (by closure), and $a \sim g_ia$, for all $i = 1 \ldots n$.

   $$\therefore \deg^+(a) = n$$

2. Cay(G,S) is strongly connected (ie. *there is a path from a to b and from b to a whenever a and b are vertices in the graph*)

   Proof:

   Let $a, b \in G$.

   $b = (ba^{-1})a$ and $ba^{-1} \in G$ (closure in G)

   Now since S is the generator set for G, ba^{-1} can be expressed as :

   $ba^{-1} = a^{1`}a^{2`}a^{3`}\ldots a^{k`}$ such that $a^{1`}, \ldots, a^{k`} \in S$ (not necc. distinct).

   Therefore, one can follow the edges defined by the sequence of generators $a^{1`}a^{2`}a^{3`}\ldots a^{k`}$, starting from vertex **a** to form a path from a to **b**.

   Using the same approach one can also find a path from b to a

3. If $G \neq \{e\}$ then $\sim$ is irreflexive i.e. Cay(G,S) has no loops

Proof:

If $G \neq \{e\}$, then $\{e\} \neq S$, and so $a \neq g_i a$ for any $g_i \in S$.

Thus, a is not related to itself, and so there are no reflexive loops in Cay(G,S).

4. If for all $g \in S$, $g^{-1} \in S$, then ~ is a symmetric relation, ie Cay(G,S) is an undirected graph.

The following proposition is stated without proof. This proposition can be possibly used as a test to check whether two groups are not isomorphic from their Cayley Digraph.

5. Let $G_1 = \langle S_1 \rangle$, $G_2 = \langle S_2 \rangle$ be isomorphic groups, $|S_1| \leq |S_2|$, then $Cay(G_1,S_1)$ is isomorphic to a sub graph of $Cay(G_2,S_2)$.

Paths in Cayley Digraphs

Definition: A Hamiltonian Path in a directed graph is a path passing through every vertex exactly once.

It was noted that there appeared to be Hamiltonian paths in all the Cayley diagrams sketched. As of yet, however, no proof or disproof exists that every Cayley diagraph has a Hamiltonian path. To make matters worse, searching for Hamiltonian paths in a graph is intractable, thus making testing of large Cayley diagrams for Hamiltonian paths prohibitive.

This is known as *Lovàsz conjecture* and is stated as follows:

Conjecture: (*Lovàsz*) All Cayley Diagrams have a Hamiltonian path.

ie. For any group $G = \langle S \rangle$, Cay(G,S) has a Hamiltonian path.

By exhaustively testing various Cayley digraphs, Willis [3] found a group ($C_2 x A_4$) which doesn't have a Hamiltonian path for a **particular generating set**. However, choosing a different generating set for the above will yield a Cayley Diagram with a Hamiltonian path.

The conjecture should thus be revised to the following form:

Conjecture: Let G be a finite group, then Cay(G,S) has a Hamiltonian path for some appropriate choice of the generating set S.

The following propositions prove the conjecture for various classes of groups. Many stronger proofs are available, usually using complex graph theory. For more information see [2]

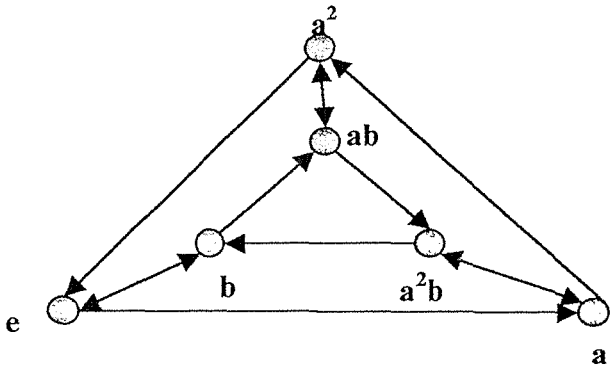
Prop 1: $Cay(C_n,\{a\})$ has a Hamiltonian cycle, for any $n \in \dot{U}^+$.

Proof:

Basis : $Cay(C_1,\{a\})$ is Hamiltonian since it consists of a single vertex.

Assume Cay(Ck, {a}) is Hamiltonian where $Ck = \{e, a, a^2, a^3,\ldots, a^{k-1}\}$

We can construct $Cay(C_{k+1},\{a\})$ from Cay(Ck, {a}) by inserting a vertex a^k between a^{k-1} and e. Inserting this element does not affect Hamiltonicity, hence by inductive hypothesis, $Cay(C_{k+1},\{a\})$ is also Hamiltonian.



The Dihedral group D_3

Dn can be regarded as two n-polygons inset in each other

Prop 2: $Cay(D_n,S)$, S= {a,b} is Hamiltonian

Proof:

Dn = <a,b>, st o(a) = n

o(b) = 2

$o(Dn) = 2n$

From the definition of D_n, we can also derive:

$$(ab)^2 = e$$
$$(ba^i) = a^{n-i}b \quad *$$

We can identify two cyles in Cay(Dn, S), namely

$$e \sim a \sim a^2 \sim a^3 \sim, \ldots, \sim a^{n-1}$$

and

$$b \sim ab \sim a^2b \sim, \ldots, \sim a^{n-1}b$$

The number of elements in these two cycles totals to 2n, so every element must be a member of one of the cycles. In fact, every element must be a member of exactly one of these cycles.

Since suppose $\exists\ d \in D_n$ such that $d = a^i = a^jb,\ 0 < i \neq j < n$

$\Rightarrow a^j = a^ib$ using *

$\Rightarrow a^j = a^ib^2$

$\Rightarrow a^i = a^j$ but this only holds when $i,j = 0$ or n, which is a contradiction.

In the case where $i = j$, then

$a^i = a^jb \Rightarrow e = b$, which is also a contradiction

Also, we have $a^{n-1} \sim ab$ since ($ba^{n-1} = ab$ by *), and $b \sim e$ (since $b^2 = e$).

We can now define the following path:

$e \rightarrow a \rightarrow a^2 \rightarrow \ldots \rightarrow a^{n-1} \rightarrow ab \rightarrow a^2b \rightarrow a^3b \rightarrow \ldots \rightarrow a^{n-1}b \rightarrow b \rightarrow e$, which is a cycle containing every vertex in $Cay(D_n,S)$ exactly once, hence it is a Hamiltonian Path.

Prop 3: Cay(Cm x Cn, S), S = {(a,e) , (e,b)} is Hamiltonian

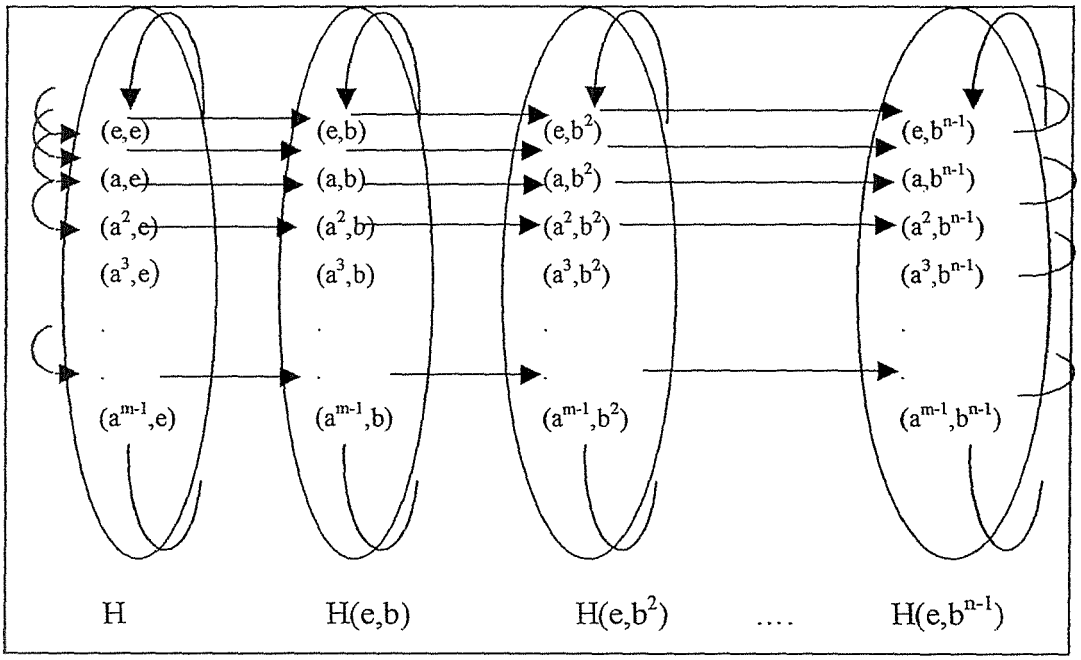
Proof:

Let $H = \langle a \rangle$. Then $H < Cm \times Cn$ and $o(H) = m$.

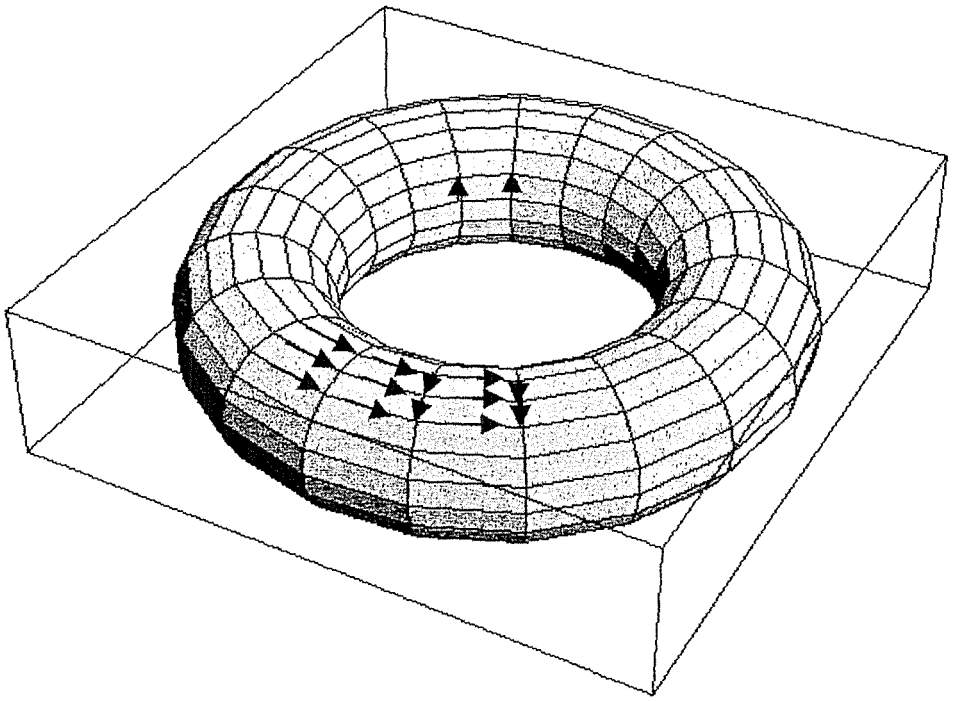
Therefore we can partition Cm x Cn into n cosets: H, H(e,b), …, $H(e,b^{n-1})$

Clearly in each coset we have $(a^i,b^j) \rightarrow (a^{i+1},b^j)$. Thus each coset forms a cycle within itself as shown in the diagram below. Between consecutive cosets we

also have that $(a^i,b^j) \rightarrow (a^i, b^{j+1})$. These relations impose a directed grid on the set set of elements in Cm x Cn.



In fact, the graph Cay(Cm x Cn, S) can be regarded as a directed grid on a torus in 3D space. Each ring in the torus represents a coset in the group. Finding a Hamiltonian in Cay(Cn x Cm, S) is reduced to finding a Hamiltonian path along the surface of this torus.

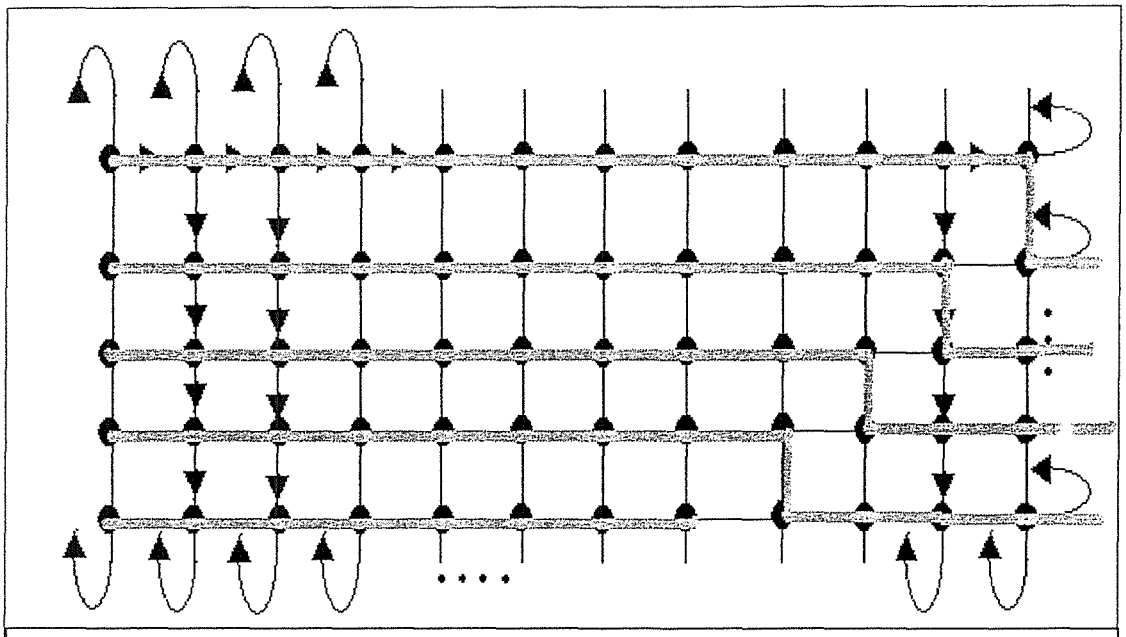


For each $i \in \mathbf{N}$, $0 < i < m$, we can define path Pi as:

$$Pi = (a^i,b^{n-i}) \rightarrow (a^i,b^{n-i+1}) \rightarrow \ldots \rightarrow (a^i,b^{n-1})$$
$$\rightarrow (a^i,e) \rightarrow (a^i,b) \rightarrow (a^i,b^2) \rightarrow \ldots \rightarrow (a^i,b^{n-i-1})$$

P_i represents a Hamiltonian path of the ith row of the Cayley digraph and contains n vertices

Also for all $i,j \in \mathbf{N}$, $0 \leq i < m, 0 \leq j < n$, $(a^i,b^j) \rightarrow (a^{i+1},b^j)$
Using these edges we can join each Pi, i = 1...m-1 to form a Hamiltonian path for the entire graph as illustrated in the diagram below:



An illustration of the path taken to form a Hamiltonian path on the surface of a torus

The path $(e,e) \rightarrow (e,b) \rightarrow (e,b^2) \rightarrow \ldots \rightarrow (e,b^{n-1}) \rightarrow P_1 \rightarrow P_2 \rightarrow P_3 \rightarrow \ldots \rightarrow P_{m-1}$ contains mn = o(Cm x Cn) elements, each element exactly once, therefore it is a Hamiltonian path as required.

While this proof only applies for the direct products of 2 cyclic groups, it could be used as a basis for a proof for the direct product of an arbitrary number of elements. Instead of considering a grid of vertices on a torus in 3D space, one must search for a Hamiltonian path in a grid of vertices on an n-dimensional torus. Since every Abelian group a direct product of cyclic groups, proving Hamiltonicity for an arbitrary direct

product will automatically imply the Hamiltonicity of all Abelian groups.

Applications

One application of Cayley Diagrams is in that on binary representation of data, namely Gray Codes. A gray code of length n is a sequence of n bit binary strings, with the property that consecutive words differ by at most one element. Gray codes are useful in mechanical encoders since a slight change in location only affects one bit. Using a typical binary code, up to n bits could change, and slight misalignments between reading elements could cause high levels of error since flipping a bit will increase/ decrease the value by a power of two. For example, an error flipping the MSB of an 8-bit word will change the value by 2^7.

Gray Codes can be represented by the direct product $(C_2)^n$. The difference between Gray Code and normal binary code is the ordering of the elements. In Gray code the "greater than" relation $\geq$ is defined as follows:

$$\text{For } a, b \in (C_2)^n, \quad a \geq b \text{ iff } a \rightarrow^* b$$

The fact that $((C_2)^n, \geq)$ is a totally-ordered set follows from that we can always find a Hamiltonian path in $Cay((C_2)^n,S)$ (since every two elements in the path are comparable) Thus $((C_2)^n, \geq)$ is a well ordered set by choosing the starting vertex (element) of the Hamiltonian path as the least element.

References

[1] The Fascination of Groups, F. J. Budden, 1972

[2] Hamiltonian Paths in Cayley Graphs, I. Pak; R. Radocic, Department of Mathematics, MIT, June 2002

[3] Hamiltonian Paths and Cayley Digraphs of Algebraic Groups, S. Willis, UCSD Honors Thesis 2001

Gödel's Theorem

Matthew Sant

At the beginning of the 20th century the mathematician David Hilbert posed a set of problems to the mathematical community that should have been the so-called road map of tasks to accomplish during the following hundred years. Among them was a problem which he posed in collaboration with Ackermann dealing with the question of whether a formal system of mathematical logic can be considered complete – where completeness implies that every true statement can be expressed within the system, possibly without a paradox.

This was probably inspired by the recent discovery of a series of paradoxes in Russell and Whitehead's *Principia Mathematica* which is now a *de facto* standard for defining and proving mathematical statements. The well-known Russell's paradox – formulated in a hundred different ways – has been catered for by denying the possibility of having a set being a member of itself. However, other forms of paradoxes are not that easy to eliminate. Epimenides' paradox falls into this category: "I am a liar" or in logic-speak: "This statement is false".

Gödel's seminal work in 1931 not only managed to show that the PM system was inconsistent, but that any sufficiently powerful formal system is bound to be littered with paradoxes. It is worth stating how series this matter is: practically speaking he stated that there might exist theorems that cannot be proved or disproved – theorems about number theory itself, for instance.

The approach to Gödel's proof I am going to use is a simplified version based on the work of Douglas R. Hofstadter, "Gödel, Escher, Bach: an Eternal Golden Braid". A book which I thoroughly recommend to anyone interested in the question of how animate matter can result out of combinations of inanimate matter.

INTRODUCING TNT

TNT stands for typographical number theory and it is basically an arbitrary formal system that is sufficiently complete for our purposes. The reason for which we shall be working in TNT rather than PM or any other established formal system is to emphasise the point that Gödel's theorem can be applied to ANY system whatsoever and still result in a paradox – or contradiction in our case.

The tools:

Logical operators:

A	-	for all
E	-	there exists
V	-	or
Λ	-	and
~	-	not
:	-	such that

Mathematical operators:
+, *, =

Variables:
a, a', a'', a''', ...

Number system:
0, S0, SS0, SSS0, ...

The axioms:

Axiom 1: Aa:~Sa=0
Axiom 2: Aa:(a+0)=a
Axiom 3: Aa:Aa':(a+Sa')=S(a+a')
Axiom 4: Aa:(a*0)=0
Axiom 5: Aa:Aa':(a*Sa')=((a*a')+a)

Without going into exceptional detail it is just sufficient to know that with these axioms it is possible to express any statement in number theory. In fact we are going to **assume** that fact in order to prove the theorem.

The task at hand is to import the statement "This is not a theorem of TNT" into TNT, possibly in a universal manner that is applicable to all formal systems.

GÖDEL NUMBERING

A first step in this direction is to introduce the revolutionary idea of replacing each symbol of the formal system by a number. You might ask yourself what is so revolutionary about such a change in notation. On one hand it represents an interesting way of making statements about numbers by using numbers that is going to be vital in order to prove this theorem, on the other it paved the way for the of Alan Turing and subsequently the invention of computers.

To give you an example, if we choose '123','666','434' and '000' to represent a, =, S and 0 respectively then the statement a=S0 becomes 123666434000.

Now we can use mathematical operations rather than string manipulations in order to express a new theorem.

THE CONCEPT OF THEOREMHOOD

I am now going to give two definitions of theoremhood that we shall use to come up with our final G sentence:

Definition: A number has *theoremhood* if it corresponds to a valid theorem of TNT—or, in other words, to a true statement about numbers.

Alternative definition: A number has *theoremhood* if it is possible to create that number from our small set of axiom-numbers, by the application of our small set of function-rules.

ARITHMOQUINING

This will be our next and final tool for the job. This is basically a way of expressing a theorem as part of itself – again a form of recursion. The method is quite simple: replace every occurrence of the variable 'a' by the Gödel number of the entire sentence.

Example: a statement like a=S0 has Gödel number 123666434000 (using our previous notation) hence the arithmoquined version would be 123666434000=S0 and the Gödelised form would be 123666434000666434000.

PROVING GÖDEL

All we need now is a statement about the impossibility of expressing the statement in TNT whose Gödel number happens to be the number of the sentence. Without further ado I'm going to give you this statement and allow you to ponder upon it on your own.

The arithmoquine of "The arithmoquine of a is not a valid TNT theorem-number" is not a valid TNT theorem-number.

SO WHAT?

I have already explained the implications of this theorem when it comes to logic and provability of theorems. However it is worth noticing that this theorem has some very interesting philosophical implications dealing with the way we think and the possibility / impossibility of artificial intelligence.

Bibliography:

Gödel, Escher, Bach - Douglas R. Hofstadter

Kurt Gödel – Collected Works Volume I (1929-1936)

Kenny's Overview of Hofstadter's Explanation of Gödel's Theorem – Kenny Felder

The Emperor's New Mind - Roger Penrose

The Duplex of a Graph

Alexander Farrugia

1 What is the Duplex?

Definition 1.1 *Consider a graph G with the vertex set $\mathcal{V}(G) = \{v_1, \ldots, v_n\}$ and edge set $\mathcal{E}(G) = \{e_1, \ldots, e_m\}$ as usual. The* **Duplex** *of G (denoted here by $\mathcal{D}G$) is the graph with the following vertex and edge set:*

- $\mathcal{V}(\mathcal{D}G) = \{v_1, \ldots, v_n, v_{1'}, \ldots, v_{n'}\}$
- $\mathcal{E}(\mathcal{D}G) = \{e_{1^*}, \ldots, e_{m^*}, e_{1'}, \ldots, e_{m'}\}$ *such that, for all $e_i \in \mathcal{E}(G)$, if $e_i = \langle v_i, v_j \rangle$, then $e_{i^*} = \langle v_i, v_{j'} \rangle$ and $e_{i'} = \langle v_{i'}, v_j \rangle$*

Clearly, $m(\mathcal{D}G) = 2m(G)$ and $n(\mathcal{D}G) = 2n(G)$ (hence the name 'duplex'). From the definition of the Duplex of G, it is definitely bipartite with vertex partitions $\{v_1, \ldots, v_n\}$ and $\{v_{1'}, \ldots, v_{n'}\}$. This means that $\mathcal{D}G$ always contains no odd circuits. Indeed, any odd circuit in G with v vertices will result in a circuit with $2v$ vertices in $\mathcal{D}G$. Even circuits in G with w vertices result in two circuits, each with w vertices, in DG. Figure 1 clarifies this point.

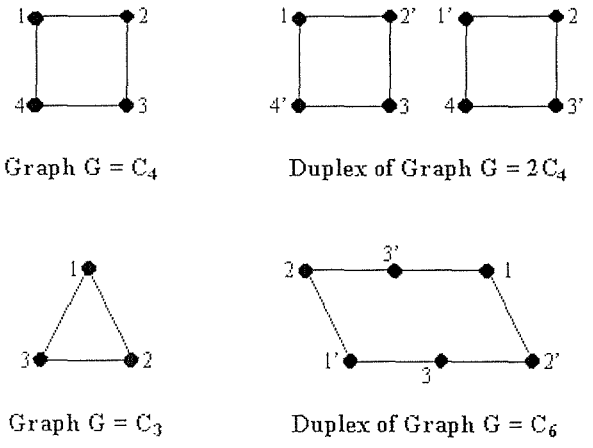


Figure 1: The Duplex of even and odd circuits

Thus we have the following lemma:

Lemma 1.2 *The Duplex is always bipartite.*

2 The Spectrum of the Duplex

A very interesting property that the Duplex has is that its spectrum contains the spectrum of the original graph G. To prove this, however, we first need a lemma from matrix theory, which will be quoted without proof.

Lemma 2.1 *If M is a non-singular square matrix, then*

$$\det \begin{pmatrix} M & N \\ P & Q \end{pmatrix} = \det(M) \cdot \det(Q - PM^{-1}N)$$

Theorem 2.2 $\phi(G)$ / $\phi(\mathcal{D}G)$

Proof: Consider the adjacency matrix of G. We can write it as:

$$\mathrm{A} = \begin{pmatrix} 0 & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & 0 & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \dots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & 0 \end{pmatrix} = \begin{pmatrix} 0 & a_{12} & a_{13} & \dots & a_{1n} \\ a_{12} & 0 & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \dots & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \dots & 0 \end{pmatrix}$$

(remembering that $\mathbf{A}$ is symmetric.) >From the definition of the duplex of graph G, the adjacency matrix of $\mathcal{D}G$ is:

$$\mathcal{D}\mathrm{A} = \begin{pmatrix} 0 & 0 & \dots & \dots & 0 & 0 & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & 0 & \dots & \dots & 0 & a_{12} & 0 & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & & & \vdots & \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & \dots & 0 & a_{1n} & a_{2n} & a_{3n} & \dots & 0 \\ 0 & a_{12} & a_{13} & \dots & a_{1n} & 0 & 0 & \dots & \dots & 0 \\ a_{12} & 0 & a_{23} & \dots & a_{2n} & 0 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & & \vdots & \vdots & \vdots & & & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \dots & 0 & 0 & 0 & \dots & \dots & 0 \end{pmatrix}$$

This can be written briefly as:

$$\mathcal{D}\mathrm{A} = \begin{pmatrix} \mathrm{0_{n\times n}} & \mathrm{A} \\ \mathrm{A} & \mathrm{0_{n\times n}} \end{pmatrix}$$

(Here, the fact that the duplex of any graph is bipartite is very apparent.)

Now it is well known that, for the case when $\mathbf{A}$ is an adjacency matrix,

$$\lambda^2\mathrm{I} - \mathrm{A}^2 = (\lambda\mathrm{I} - \mathrm{A})(\lambda\mathrm{I} + \mathrm{A})$$

With this in mind, we can show what is required to prove, thus:

$$\begin{aligned}
\phi(\mathcal{D}\mathbf{A}) &= \det\begin{pmatrix} -\lambda\mathbf{I} & \mathbf{A} \\ \mathbf{A} & -\lambda\mathbf{I} \end{pmatrix} \\
&= \det(-\lambda\mathbf{I})\det(-\lambda\mathbf{I} - \mathbf{A}\left(-\frac{1}{\lambda}\mathbf{I}\right)\mathbf{A}) \text{ using Lemma 2.1} \\
&= (-\lambda^n)\det\left(-\frac{1}{\lambda}(\lambda^2\mathbf{I} - \mathbf{A}^2)\right) \\
&= (-\lambda^n)\left(-\frac{1}{\lambda^n}\right)\det(\lambda^2\mathbf{I} - \mathbf{A}^2) \\
&= \det(\lambda^2\mathbf{I} - \mathbf{A}^2) \\
&= \det((\lambda\mathbf{I} - \mathbf{A})(\lambda\mathbf{I} + \mathbf{A})) \\
&= \det(\lambda\mathbf{I} - \mathbf{A})\det(\lambda\mathbf{I} + \mathbf{A}) \\
&= \phi(\mathbf{A})\det(\lambda\mathbf{I} + \mathbf{A})
\end{aligned}$$

Therefore, $\phi(\mathcal{D}\mathbf{A}) \mid \phi(\mathbf{A})$, as required.

From the above theorem, we can easily find what is the spectrum of the duplex of a graph from its spectrum. We showed above that $\phi(\mathcal{D}\mathbf{A}) = \phi(\mathbf{A})\det(\lambda\mathbf{I} + \mathbf{A})$, where $\mathcal{D}\mathbf{A}$ and $\mathbf{A}$ are the adjacency matrices of $\mathcal{D}G$ and G respectively. Since $\phi(\mathbf{A}) = \det(\lambda\mathbf{I} - \mathbf{A}) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$ (where $\lambda_i : 1 \leq i \leq n$ are the elements of the spectrum of G with possible repetitions), it is intuitive to conclude that $\det(\lambda\mathbf{I} + \mathbf{A}) = (\lambda + \lambda_1)(\lambda + \lambda_2)\cdots(\lambda + \lambda_n)$. Indeed, this is the case.

Lemma 2.3 *If $\phi(\mathbf{A}) = \det(\lambda\mathbf{I} - \mathbf{A}) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$, where $\lambda_i : 1 \leq i \leq n$ are the elements of the spectrum of G (with possible repetitions), then $\det(\lambda\mathbf{I} + \mathbf{A}) = (\lambda + \lambda_1)(\lambda + \lambda_2)\cdots(\lambda + \lambda_n)$.*

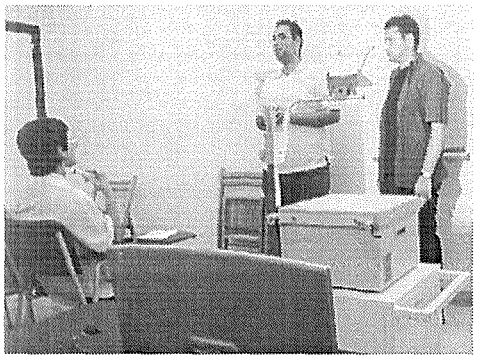
Proof: Consider $\phi(\mathbf{A}) = \det(\lambda\mathbf{I} - \mathbf{A}) = (\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_n) = \alpha_0 + \alpha_1\lambda + \alpha_2\lambda^2 + \cdots + \alpha_n\lambda^n$. In the case of $\det(\lambda\mathbf{I} + \mathbf{A})$, it will be exactly the same, except that the coefficients of the odd powers of λ will be $-\alpha_i$ instead of α_i, i.e. it will be equal to $\alpha_0 - \alpha_1\lambda + \alpha_2\lambda^2 - \cdots + (-1^n)\alpha_n\lambda^n$. Also, $(\lambda - \lambda_i)$ for all i was a factor for $\det(\lambda\mathbf{I} - \mathbf{A})$, which means that $\alpha_0 + \alpha_1\lambda_i + \alpha_2\lambda_i^2 + \cdots + \alpha_n\lambda_i^n = 0$. This implies that for $\det(\lambda\mathbf{I} + \mathbf{A})$, $(\lambda + \lambda_i)$ for all i is a factor, since $\alpha_0 - \alpha_1(-\lambda_i) + \alpha_2(-\lambda_i)^2 + \cdots + (-1)^n\alpha_n(-\lambda_i)^n = 0$, as required.

The following result follows immediately from the above lemma:

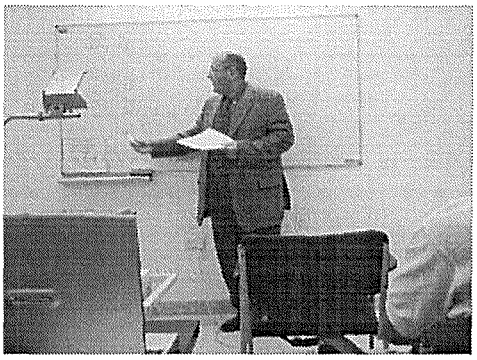
Proposition 2.4 *If the spectrum of G is $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ (with possibly repeated elements), then the spectrum of $\mathcal{D}G$ is $(\lambda_1, \lambda_2, \ldots, \lambda_n, -\lambda_1, -\lambda_2, \ldots, -\lambda_n)$ (with possibly repeated elements).*

Remark: A graph is bipartite if its spectrum is symmetric about 0. From the above corollary, we confirm again that the duplex of any graph is always bipartite.

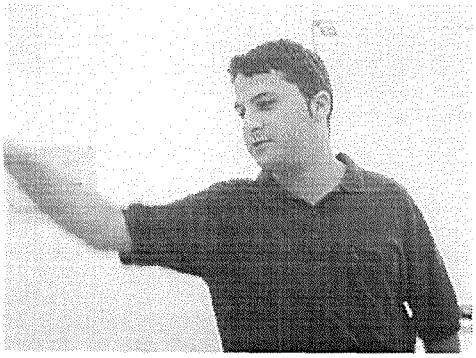
The Speakers



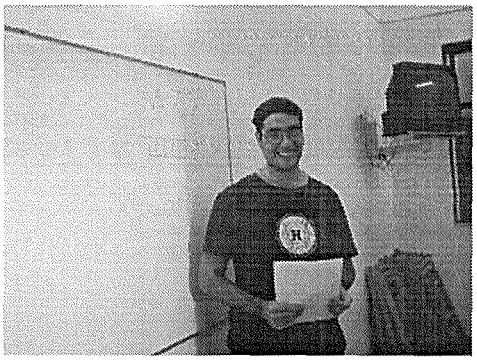
Andrew Duncan and Andrew Cortis



Prof. Anton Buhagiar

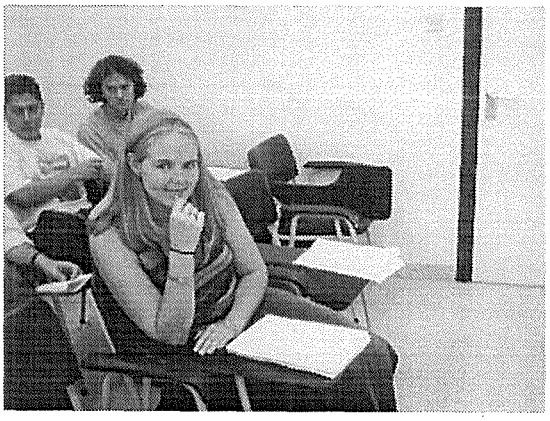


Matthew Sant



Alexander Farrugia

Pamela Cohen and Cheryl Zerafa



Anne Yancey

The Audience