Mathematical Spectrum

2003/2004 Volume 36 Number 1

- **Areas of continents**
- **Minkowski space**
- **Catalan numbers**
- **Taxicab geometry**

A magazine for students and teachers of mathematics
in schools, colleges and universities

MATHEMATICAL SPECTRUM

This is a magazine for students and teachers in schools, colleges and universities, as well as the general reader interested in mathematics. It is published by the Applied Probability Trust, a non-profit-making organisation established in 1963 with the support of the London Mathematical Society. The object of the Trust is the encouragement of study and research in the mathematical sciences.

One volume of *Mathematical Spectrum* is published in each British academic year consisting of three issues, which appear in September, January and May.

Articles published in *Mathematical Spectrum* deal with the entire range of mathematical disciplines (pure mathematics, applied mathematics, statistics, operational research, computing science, numerical analysis, biomathematics). Both expository and historical material may be included, as well as elementary research and information on educational opportunities and careers in mathematics. There are also sections devoted to problems, to mathematics in the classroom, and to computing.

From the Editor

In Code: A Mathematical Journey

Sarah Flannery lives on a farm in County Cork, Ireland, with her mum and dad and four brothers. Nothing very remarkable about that. What is a bit special is that she wrote a book entitled *In Code: A Mathematical Journey*, with the help of her father. In the book she charts her growing interest in mathematics, first aroused by her father who teaches mathematics at Cork Institute of Technology. She mentions some of the puzzles that she encountered:

- How can you time the boiling of an egg for 4 minutes if you have two egg timers, a 3 minute timer and a 5 minute timer?
- How can you send something valuable in a strong box to a friend through the post if the postal service is such that every unsecured item is opened and the contents stolen?
- How many squares are there on a chessboard?

And many more.

Sarah went along to an evening class run by her father and became interested in numbers. Just the simple whole numbers $1, 2, 3, \ldots$. In October 1997 a teacher at her school was looking for students to do a project for the Young Scientist Competition to be held the following January in Dublin. Sarah's father suggested that she should do a project on cryptography. She learned how number theory had been used to develop a method of sending messages by a code that cannot be broken, called the RSA cryptosytem after its inventors, Ronald Rivest, Adi Shamir and Leonard Adelman at Massachusetts Institute of Technology in 1977. Its security depends on the impracticability of factorising a large number even if it is known to be the product of two primes. By 'large' we mean with 200 or more digits. The book introduces the mathematics behind the RSA cryptosystem, but in a gentle way suitable for a non-mathematically literate reader. Some readers may find the style in this part a bit patronizing.

The rest, as they say, is history. Sarah won a prize at the Young Scientist Competition, which led to further competitions and the development of her project. She developed a new coding system using matrix theory which she called the CP system after the famous nineteenth-century English mathematician Sir Arthur Cayley, the inventor of matrices, and Dr Michael Purser, the founder of a Dublin cryptography company where Sarah did some work. The details of the CP system are not explained, and the website was unobtainable when I tried it. The CP system is 20 times faster than the RSA system. It brought Sarah more prizes and international recognition, culminating in her becoming Ireland's Young Scientist of the Year in 1999 and winning a first prize at the European Union Contest for Young Scientists in that year in Greece. She was in demand as a speaker from Singapore to Stockholm, where she attended the Nobel Prize ceremonies in December 1999. For her, a highlight was to get a phone call from Ronald Rivest to talk about her CP system.

But, as Sarah was always at pains to emphasize in her presentations, her CP system had not been tested rigorously to check whether it was safe from attack. Could it be broken by an 'enemy'? In the end it turned out not to be secure, and couldn't be modified to make it so. Sarah writes: 'Its public element simply gives just a little too much information to the bad guys.' So was it all for nothing? Sarah again: 'My beloved CP algorithm may not have turned out to be earth-shatteringly amazing, but it was still good science.' As for Sarah's future, she writes: 'I would like to be one of those lucky people who get paid for doing what they love', in her case mathematics.

This book is good read, a mathematical whodunnit. It's worth it just for the story of Daniel Gorenstein and his chauffeur!

In Code: A Mathematical Journey. By Sarah Flannery with David Flannery. Profile Books, London, 2000. Pp. 271. Hardback £14.99 (ISBN 1-86197-222-9).

Positive integers

For positive integers m, p with $p < 2m+1$, prove that

$$\sqrt{m^2+p} = m + \cfrac{p}{2m + \cfrac{p}{2m + \cfrac{p}{2m + \cdots}}}$$

Bablu Chandra Dey
Kolkata, India.

Distribution of Areas of Continents and Islands

A. TAN, W. LYATSKY and SUSAN XU

1. Introduction

The theory of continental drift, first proposed by Wegener in 1912, is universally accepted today. According to this theory, there was once a single landmass called Pangaea which was surrounded by the world ocean called Panthalassa. Around 200 million years ago, Pangaea started to break up into two supercontinents named Laurasia and Gondwanaland, which subsequently fragmented into continents. These continued to drift, but remained largely intact to form the present-day continents.

The formation of the islands of the world is less discussed. It may be safely assumed that they were formed in secondary and tertiary processes by plate tectonics and other geological upheavals. Fracture and drift around the more fragile edges of the continents probably gave rise to the majority of the islands. For example, Madagascar separated from the African mainland to become an island just as Tasmania separated from Australia. This process continues today and it is predicted that one day the Baja California peninsula will detach from North America to form an island.

2. Distribution in areal fragmentation

In this article, we treat the formation of continents and islands as the result of fragmentation of the original landmass Pangaea. We study the size distributions of these land fragments in terms of the area A, or its equivalent diameter which is proportional to $\sqrt{A}$. Distributions are customarily discussed in terms of the cumulative number N of fragments having sizes equal to or greater than a given size. Using this convention, the cumulative number of the largest fragment is 1, that of the second largest fragment is 2, and so on. This is also called the areal rank of the fragments, which is the term that we will use here.

It is a law of nature that the random fragmentation of an object produces fragments of different sizes which have a certain characteristic distribution. In the random fragmentation of a linear object, the resulting distribution is usually exponential (see reference 1). The cumulative number of fragments N as a function of the length L is given by

$$N(L) = N_0 \mathrm{e}^{-N_0 L}\,, \tag{1}$$

where N_0 is a constant. For this distribution, the plot of N in logarithmic scale against L is a straight line with negative gradient. However, in the areal fragmentation of a two-dimensional object, various types of distributions are possible, including one given by (1) with A replacing L (see reference 2). A logical extension of (1) in two-dimensional space was given by Mott and Linfoot (reference 3),

$$N(A) = N_0 \mathrm{e}^{-(2N_0 A)^{1/2}}\,.$$

For this distribution, the plot of N in logarithmic scale against $\sqrt{A}$ is a straight line with negative gradient (see reference 2).

Other distributions in areal fragmentation include analytical solutions in terms of modified Bessel functions and computer-generated distributions of various types. However, the simple Mott distribution is generally the most successful in validating experimental data (see reference 2).

3. The data

The data for this study were compiled from *The New York Times Atlas of the World* (reference 4) and *The New Headline World Atlas* (reference 5) and entered in table 1. The continental and islands data were also taken from these. The continental areas include the adjacent islands; North America includes Greenland, the Canadian islands and the Caribbean islands, and Australasia includes New Guinea, New Zealand and neighbouring Pacific islands.

4. The distribution of continents and islands

Plots using the data from table 1 show that none of the distributions mentioned in section 2 gives any satisfactory trends for the observed data. Figure 1 gives the most interesting analysis of the data, in which N in linear scale is plotted against the effective diameter $\sqrt{A}$. The data points fall on three straight lines representing three distinct groups, firstly, the continents and Greenland, secondly, the large islands from New Guinea to Great Britain, and thirdly, the smaller islands. In figure 1, Greenland, the largest island, is elevated to the status of a continent.

The equation of the straight line for the continents can be written as

$$N(A) = m\sqrt{A} + c\,, \tag{2}$$

where m is the gradient and c the intercept on the y-axis. The intercept on the x-axis is 7800, and on the y-axis is 9, which gives $m = -0.00113$ and $c = 9$. If we regard (2) as defining a distribution, then its probability density ρ is N/S, where S is the area under the continental straight line. Here $S = \frac{1}{2} \times 9 \times 7870 = 35\,415$. Therefore,

$$\rho(A) = (-3.189 \times 10^{-8})\sqrt{A} + 2.541 \times 10^{-4}\,. \tag{3}$$

If we assume the continents to be square, then the average continental side is 3935 km, and the total area of the continents is thus $9 \times (3935)^2$ km^2 which is 139 358 025 km^2.

Table 1. Areas of continents and largest islands.

Continents/islands	Area A (km^2)	$\sqrt{A}$ (km)	Area rank N
EURASIA	54 106 000	7 356	1
AFRICA	30 335 000	5 508	2
N. AMERICA	25 349 000	5 035	3
S. AMERICA	17 611 000	4 197	4
ANTARCTICA	13 340 000	3 652	5
AUSTRALASIA	8 923 000	2 987	6
Greenland	2 175 600	1 475	7
New Guinea	789 950	889	8
Borneo	751 100	867	9
Madagascar	586 376	766	10
Baffin	507 454	712	11
Sumatra	424 760	652	12
Honshu	227 920	477	13
Great Britain	218 896	468	14
Victoria	217 290	466	15
Ellesmere	196 236	443	16
Celebes	189 034	435	17
South Island	151 238	389	18
Java	126 501	356	19
North Island	114 444	338	20
Newfoundland	108 860	330	21
Cuba	104 981	324	22
Luzon	104 688	324	23
Iceland	103 000	321	24
Mindanao	94 631	308	25
Ireland	82 214	287	26
Sakhalin	76 405	276	27
Hispaniola	76 143	276	28
Hokkaido	75 066	274	29
Banks	70 028	263	30
Ceylon	65 610	256	31
Tasmania	63 710	252	32
Svalbard	62 049	249	33
Devon	55 247	235	34
Novaya Zemlya	48 200	220	35
Marajo	46 597	216	36
Tierra del Fuego	46 360	215	37
Alexander	43 250	208	38
Axel Heiberg	43 170	208	39
Melville	42 150	205	40
Southampton	41 215	203	41
New Britain	36 519	191	42
Taiwan	35 835	189	43
Kyushu	35 664	189	44
Hainan	33 999	184	45
Prince of Wales	33 338	183	46
Spitsbergen	31 999	179	47
Vancouver	31 285	177	48
Timor	29 855	173	49
Sicily	25 708	160	50
Somerset	24 786	159	51
Sardinia	24 090	155	52
Shikoku	17 767	133	53
New Caledonia	16 913	130	54
Nordaustlandet	16 599	129	55
Samar	13 080	114	56
Negros	12 707	113	57

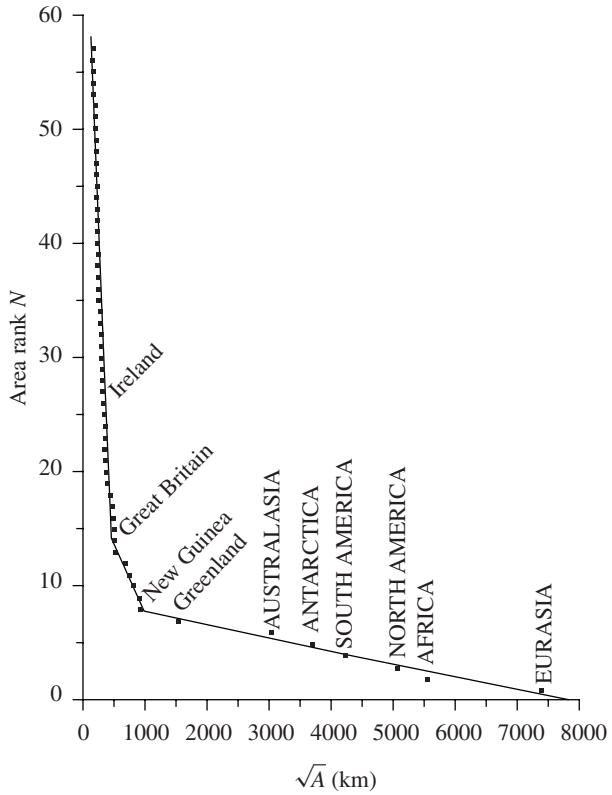


Figure 1. Plot of area rank N against the effective diameter $\sqrt{A}$. Three least-square straight lines define the continental group (including Greenland) and the large and small islands groups.

Since the continents include the islands, this is also the total area of the land surface of the globe. This crude calculation is within 6.43% of the actual figure of 148 941 000 km^2 (see reference 5). A more rigorous calculation involves expressing A as a quadratic function of N from (3) and integrating over N from $-c$ to 0; this is left to the reader.

The linear distribution of N in terms of $\sqrt{A}$ places an upper limit for the area of the largest continent as $(7870\,\text{km})^2$, which is 61 936 900 km^2. A continent larger than this area would be unstable and break up according to this model. In an exponential distribution, the corresponding upper limit is far greater, and is theoretically equal to the total area of the land surface of the world.

In spite of their attractiveness, the three separate linear distributions are lacking in completeness. A single distribution which will account for all landmasses, i.e. both continents and islands, is needed. However, this is not an easy task and no simple functional relationship describes the observed distribution. Among the common distributions found in the literature, the long-tailed Zipf distribution (cf. reference 6, pp. 465–471) appears to offer most promise of achieving this objective. In this distribution, the probability density function is given by

$$\Pr(N) = \frac{C}{N(A)^{1+\alpha}}\,, \qquad \alpha > 0\,,$$

which follows from the deterministic relation

$$\sqrt{A} = \frac{C}{N(A)^{\beta}}\,, \tag{4}$$

where $\beta = 1 + \alpha$. Taking logarithms, we obtain

$$\log C - \beta \log N(A) = \log \sqrt{A}\,, \tag{5}$$

and then multiplying both sides by $N(A)$ gives

$$N(A) \log C - \beta N(A) \log N(A) = N(A) \log \sqrt{A}\,. \tag{6}$$

The two normal equations are obtained by summing (5) and (6) over the data points, giving

$$n \log C - \beta \sum \log N(A) = \sum \log \sqrt{A}\,, \tag{7}$$

and

$$\log C \sum N(A) - \beta \sum N(A) \log N(A) = \sum N(A) \log \sqrt{A}\,, \tag{8}$$

where n is the number of data points. For a least-square fit of the data points, the constants C and β are obtained by elimination from (7) and (8), giving

$$\beta = \frac{\sum N(A) \sum \log \sqrt{A} - n \sum N(A) \log \sqrt{A}}{n \sum N(A) \log N(A) - \sum N(A) \sum \log N(A)} \tag{9}$$

and

$$C = \exp\left(\frac{\beta \sum \log N(A) + \sum \log \sqrt{A}}{n} \right). \tag{10}$$

Figure 2 shows the effective diameter $\sqrt{A}$ against the area rank N. The least-square line is given by (4) with constants $C = 11136.61$ and $\beta = 1.101553$ which are obtained from (9) and (10) by summing over the data from table 1. Evidently, the fit is good for the small islands and only fair for the large islands and Greenland, but poor for the continents except Africa. Other visually better fits were possible with $\beta = 1.2$ say, but such fits had greater least-square errors. The fact that the rectilinear fits in figure 1 gave far better agreement with the data suggests that the continents and islands belong to three separate distributions and were formed from different geological processes.

5. Summary and conclusions

Various types of areal fragmentation of two-dimensional objects including Mott, exponential, theoretical and computer-generated distributions are found in the literature. The distribution of continents and islands formed from the fragmentation of the original landmass Pangaea fits into a curious new category. Distinct groups of continents and large and small islands clearly separate themselves out in a distribution plot of the effective diameter. In this distribution, Greenland is elevated to the status of a continent. The overall distribution of all landmasses can be approximated by a Zipf distribution.

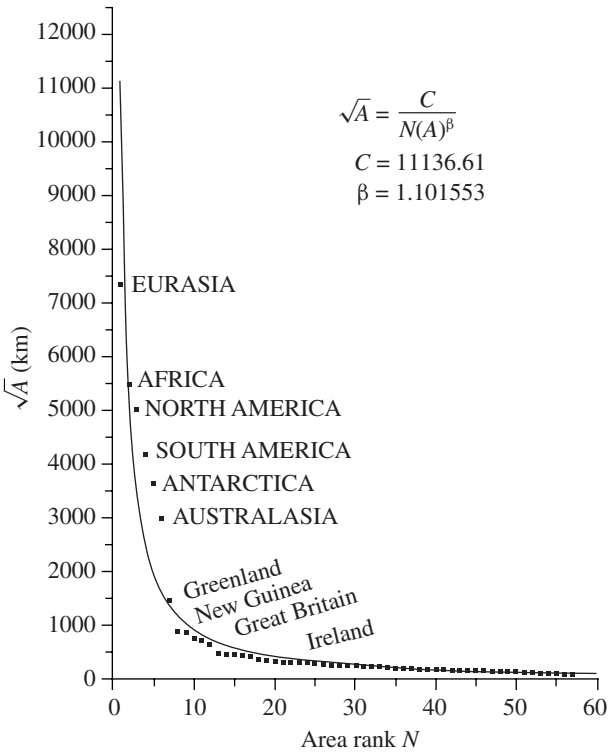


Figure 2. Least-square Zipf distribution fit of data in the plot of $\sqrt{A}$ against N. The agreement is good for the small islands, fair for the large islands, but generally poor for the continents.

Acknowledgements

This study was partially supported by NASA grant NAG5-10202 and Office of Naval Research grant N00014-97-1-0267.

References

1. C. C. Lienau, Random fracture of brittle solids, *J. Franklin Inst.* **221** (1936), pp. 485–494, 674–686, 769–783.
2. D. E. Grady and M. E. Kipp, Geometric statistics and dynamic fragmentation, *J. Appl. Phys.* **58** (1985), pp. 1210–1222.
3. N. F. Mott and E. H. Linfoot, AC 3348, Ministry of Supply, January 1943.
4. *The New York Times Atlas of the World* (John Bartholomew and Sons and Times Books, London, 1986).
5. *New Headline World Atlas* (Hammond, Maplewood, NJ, 1990).
6. N. L. Johnson, S. Kotz and A. W. Kemp, *Univariate Discrete Distributions*, 2nd edn (John Wiley, New York, 1992).

A. Tan *is a professor of physics at Alabama Agricultural and Mechanical University. He has special interests in applied mathematics and has published many articles in* Mathematical Spectrum.

W. Lyatsky *is a Visiting Scholar from the Polar Research Institute in Russia. He specialises in space physics and has a keen interest in geophysical sciences.*

Susan Xu *is finishing her Masters degree in computer science.*

An Introduction to Minkowski Space

CĂLIN GALERIU

1. Introduction

The theory of special relativity was created by Einstein in 1905 (see references 1 and 2), and is based on two postulates. The first postulate states that the laws of physics are the same for any inertial (non-accelerating) reference frame. The second postulate states that the speed at which light propagates is constant, and does not depend on the velocity of the inertial reference frame in which it is measured. Since light signals determine the way in which we experience space localization and time synchronization, the second postulate probes deeply into the nature of space and time.

Poincaré (see reference 3) has shown that the transformations of special relativity (the Lorentz transformations), relating measurements in one inertial frame to measurements in another provide a group of transformations for which the expression $x^2 + y^2 + z^2 - \mathrm{c}^2 t^2$ is invariant (where c is the speed of light in a vacuum). This expression is similar to $x^2 + y^2 + z^2$, the invariant of rotations in three-dimensional Euclidean space.

Investigating this analogy, Minkowski (see reference 4) realized that space and time could be linked into a four-dimensional continuum, with coordinates $(x, y, z, \mathrm{i}ct)$ where the fourth variable is 'pure imaginary' ($\mathrm{i} = \sqrt{-1}$). He postulated such a continuum, and then rederived all the known results of electrodynamics and special relativity in a four-dimensional treatment (reference 5). The Lorentz transformations are no more than rotations in Minkowski space.

The constancy of the speed of light can now easily be understood. Indeed, consider two events, the emission of a light signal at point A (at position x_A, y_A, z_A, at time t_A) and the reception of the same signal at point x_B (at position x_B, y_B, z_B, at time t_B). The speed of light is then given by

$$\mathrm{c} = \frac{\sqrt{(x_\mathrm{B} - x_\mathrm{A})^2 + (y_\mathrm{B} - y_\mathrm{A})^2 + (z_\mathrm{B} - z_\mathrm{A})^2}}{t_\mathrm{B} - t_\mathrm{A}} .$$

This equation can be rewritten as

$$(x_\mathrm{B} - x_\mathrm{A})^2 + (y_\mathrm{B} - y_\mathrm{A})^2 + (z_\mathrm{B} - z_\mathrm{A})^2 + (\mathrm{i}ct_\mathrm{B} - \mathrm{i}ct_\mathrm{A})^2 = 0 ,$$

and interpreted as expressing the invariance of a null 'distance' in a space with three real dimensions and one imaginary dimension.

The space–time continuum provided a framework for all later mathematical work in relativity. It was the foundation on which Einstein developed the theory of general relativity. Unfortunately, the simpler Minkowski space was not able to deal with accelerating reference frames, and was to be replaced by the more versatile Riemannian space.

The purpose of this article is to give an informal and detailed introduction to the geometry of Minkowski space in the light of the parallelism with Euclidean space, and to discuss some simple applications to special relativity. We will talk about motion only in the x-direction, so that we can plot events in an $(x, \mathrm{i}ct)$ plane, called the *Minkowski plane*. We will call the trajectory of a particle in this plane a *world line*.

2. The Euclidean and the Minkowski planes

The Minkowski plane, like the Euclidean plane, is a two-dimensional plane. A basis has two vectors, $\boldsymbol{x}$ and $\boldsymbol{y}$, which satisfy the orthonormality conditions $\boldsymbol{x} \cdot \boldsymbol{x} = 1$, $\boldsymbol{y} \cdot \boldsymbol{y} = 1$, $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{y} \cdot \boldsymbol{x} = 0$. A vector in the Euclidean plane is a linear combination of the basis vectors, with real coefficients. But for a vector in the Minkowski plane, only the first coefficient is real, the last being pure imaginary. (Alternatively, we might keep the coefficient real and absorb the imaginary numbers into the condition $\boldsymbol{y} \cdot \boldsymbol{y} = -1$.)

2.1. Orthogonality

For two vectors in the Euclidean plane, $\boldsymbol{r}_\mathrm{OA} = x_\mathrm{A}\boldsymbol{x} + y_\mathrm{A}\boldsymbol{y}$ and $\boldsymbol{r}_\mathrm{OB} = x_\mathrm{B}\boldsymbol{x} + y_\mathrm{B}\boldsymbol{y}$, the scalar product is given by $\boldsymbol{r}_\mathrm{OA} \cdot \boldsymbol{r}_\mathrm{OB} = x_\mathrm{A}x_\mathrm{B} + y_\mathrm{A}y_\mathrm{B}$. The two vectors are said to be orthogonal if their scalar product is zero, that is, $x_\mathrm{A}x_\mathrm{B} + y_\mathrm{A}y_\mathrm{B} = 0$. In this case

$$\frac{x_\mathrm{A}}{y_\mathrm{A}} = -\frac{y_\mathrm{B}}{x_\mathrm{B}} \quad (\text{for } x_\mathrm{B}, y_\mathrm{A} \neq 0) . \tag{1}$$

For two vectors in the Minkowski plane, $\boldsymbol{r}_\mathrm{OA} = x_\mathrm{A}\boldsymbol{x} + \mathrm{i}y_\mathrm{A}\boldsymbol{y}$ and $\boldsymbol{r}_\mathrm{OB} = x_\mathrm{B}\boldsymbol{x} + \mathrm{i}y_\mathrm{B}\boldsymbol{y}$, the scalar product is given by $\boldsymbol{r}_\mathrm{OA} \cdot \boldsymbol{r}_\mathrm{OB} = x_\mathrm{A}x_\mathrm{B} - y_\mathrm{A}y_\mathrm{B}$. The two vectors are orthogonal if their scalar product is zero, that is, $x_\mathrm{A}x_\mathrm{B} - y_\mathrm{A}y_\mathrm{B} = 0$. In this case

$$\frac{x_\mathrm{A}}{y_\mathrm{A}} = \frac{y_\mathrm{B}}{x_\mathrm{B}} \quad (\text{for } x_\mathrm{B}, y_\mathrm{A} \neq 0) . \tag{2}$$

We can correlate the coefficients x_A, y_A, x_B, y_B with the coordinates of some Euclidean points A, B, and construct the geometrical representation from figure 1.

The relations (1) and (2) reflect the similarity of two right-angled triangles (OAC and OBD), and the equality of corresponding angles ($\angle$AOC and $\angle$BOD). The minus sign in (1) tells us that the two vectors are in different but adjacent quadrants. The plus sign in (2) tells us that the two vectors are in the same or in opposite quadrants. In the Euclidean plane between two orthogonal vectors we have the usual right angle, but in the Minkowski plane two orthogonal vectors — as plotted on this Euclidean sheet of paper — are

characterized as making the same angle with the bisecting line of the first quadrant. Pythagoras's theorem applies for the Minkowski plane as well.

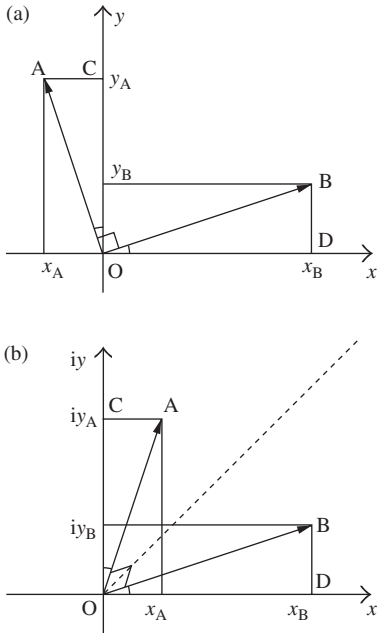


Figure 1. Orthogonality in (a) the Euclidean and (b) the Minkowski planes.

2.2. The trigonometric circle

The circle is the geometric locus of the points equally distant from a given point. For the Euclidean plane, we have

$$x^2 + y^2 = a^2,$$

and for the Minkowski plane we have

$$x^2 - y^2 = a^2.$$

This equation describes a hyperbola, as represented in figure 2.

The unit trigonometric circle has radius $a = 1$, and we define the value of the angle $\angle$AOB in figure 2 to be the distance — measured along the circumference — from the intersection of the circle with the x-axis to the point A. For the Euclidean plane, this is a real number α. The co-ordinates of the point A are $x_A = \cos(\alpha)$ and $y_A = \sin(\alpha)$, where the sine and cosine functions have Taylor expansions and can be expressed as complex exponentials as follows:

$$\sin(\alpha) = \alpha - \frac{\alpha^3}{3!} + \frac{\alpha^5}{5!} - \cdots = \frac{e^{i\alpha} - e^{-i\alpha}}{2i},$$
$$\cos(\alpha) = 1 - \frac{\alpha^2}{2!} + \frac{\alpha^4}{4!} - \cdots = \frac{e^{i\alpha} + e^{-i\alpha}}{2}.$$

For the Minkowski plane, the angle is a pure imaginary number $i\phi$ for some $\phi \in \mathbb{R}$. The co-ordinates of the point A are $x_A = \cos(i\phi)$ and $iy_A = \sin(i\phi)$, and we have

$$\begin{aligned}\sin(i\phi) &= i\phi - \frac{(i\phi)^3}{3!} + \frac{(i\phi)^5}{5!} - \cdots = \frac{e^{-\phi} - e^{\phi}}{2i} \\ &= i\sinh(\phi),\end{aligned}$$
$$\begin{aligned}\cos(i\phi) &= 1 - \frac{(i\phi)^2}{2!} + \frac{(i\phi)^4}{4!} - \cdots = \frac{e^{-\phi} + e^{\phi}}{2} \\ &= \cosh(\phi).\end{aligned}$$

The *hyperbolic sine* and *hyperbolic cosine* introduced above are both real functions. We can similarly define other hyperbolic functions such as the hyperbolic tangent $\tanh(\phi) = \sinh(\phi)/\cosh(\phi)$. All trigonometric formulae for the sine and cosine functions of real arguments apply to pure imaginary arguments as well. In this way, the relation $\cos^2(\alpha) + \sin^2(\alpha) = 1$ gives us the relation $\cosh^2(\phi) - \sinh^2(\phi) = 1$.

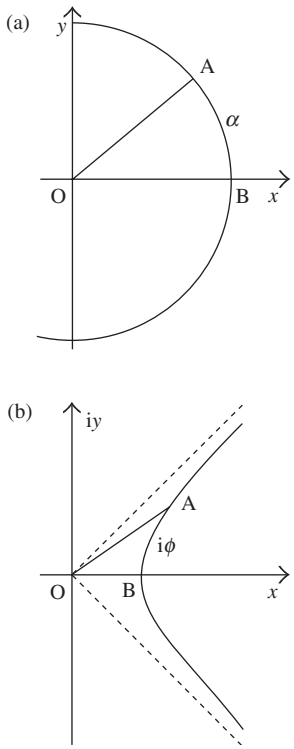


Figure 2. The unit trigonometric circle in (a) the Euclidean and (b) the Minkowski planes.

2.3. Rotation of the reference frame

Rotation of the reference frame means a change in the direction of the x-axis, and at the same time a change in the direction of the y-axis, such that the two axes remain orthogonal and continue to pass through the origin O, as shown in figure 3.

When we rotate a Euclidean reference frame through an angle α, we have new co-ordinates

$$x' = x\cos(\alpha) + y\sin(\alpha),$$
$$y' = -x\sin(\alpha) + y\cos(\alpha).$$

A rotation of a Minkowski frame through an angle $\mathrm{i}\phi$ gives new co-ordinates

$$\begin{aligned} x' &= x\cos(\mathrm{i}\phi) + \mathrm{i}y\sin(\mathrm{i}\phi) \\ &= x\cosh(\phi) - y\sinh(\phi)\,, \qquad (3) \\ \mathrm{i}y' &= -x\sin(\mathrm{i}\phi) + \mathrm{i}y\cos(\mathrm{i}\phi) \\ &= -\mathrm{i}x\sinh(\phi) + \mathrm{i}y\cosh(\phi). \qquad (4) \end{aligned}$$

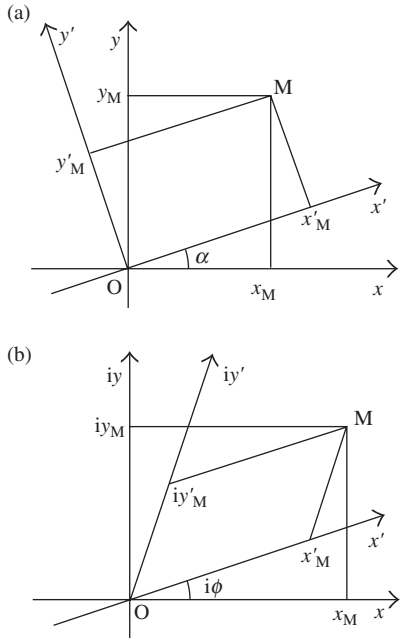


Figure 3. Rotation of the reference frame in (a) the Euclidean and (b) the Minkowski planes.

2.4. Classification of separations

We should think of the points $(x, \mathrm{i}y)$ in the Minkowski plane as corresponding to events $(x, \mathrm{i}ct)$ in space–time. In the Euclidean plane, we define the distance D between two points (x_A, y_A) and (x_B, y_B) by $D^2 = (x_A - x_B)^2 + (y_A - y_B)^2$, so D^2 is either positive (for distinct points) or zero (for coincidental points). But in the Minkowski plane we have the separation S given by $S^2 = (x_A - x_B)^2 - (y_A - y_B)^2$ and now S^2 may be positive (in which case we say that the separation is *space-like* because two events occur at the same time if $y_A = y_B$) or negative (in which case we say that the separation is *time-like* because two events happen at the same place if $x_A = x_B$) or zero (in which case we call the separation *null* and a light signal could pass through both events).

3. Applications to special relativity

We now present a few applications to special relativity, in which we exploit the geometry and trigonometry of the Minkowski plane. We will study only the simplest case of one spatial and one temporal dimension.

3.1. The Lorentz transformation

Consider a transformation from an inertial frame K to an inertial frame K' which moves with velocity $\boldsymbol{V} = V\boldsymbol{x}$ relative to K. Their origins coincide at $t = t' = 0$. The Lorentz transformation describes how measurements of position and time in K are related to measurements in K' and vice versa. In the Minkowski plane $(x, \mathrm{i}ct)$ the origin of the reference frame K' is seen as a point moving along a straight line whose slope is $V/\mathrm{i}c$. Along this line, which is the trajectory of the origin of the frame K', we have $x' = 0$. Therefore, this line is the new axis of time $\mathrm{i}ct'$. The new axis of the coordinate x' is simply a line orthogonal to $\mathrm{i}ct'$. The Lorentz transformation takes the form of a rotation, through angle $\mathrm{i}\phi$, of the axes in the Minkowski plane.

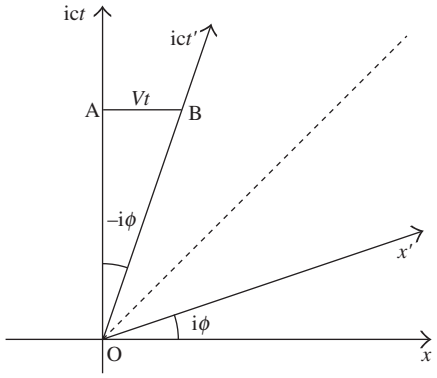


Figure 4. The Lorentz transformation is a rotation in the Minkowski space.

Referring to figure 4, we see that

$$\frac{\mathrm{AB}}{\mathrm{OA}} = \frac{Vt}{\mathrm{i}ct} = \tan(-\mathrm{i}\phi)$$

so the angle $\mathrm{i}\phi$ is given by

$$\tan(\mathrm{i}\phi) = \frac{\mathrm{i}V}{\mathrm{c}}\,,$$

and thus $V = \mathrm{c}\tanh(\phi)$. Now,

$$\begin{aligned} \cosh(\phi) = \cos(\mathrm{i}\phi) &= \frac{1}{\sqrt{1+\tan^2(\mathrm{i}\phi)}} \\ &= \frac{1}{\sqrt{1 - V^2/\mathrm{c}^2}} \qquad (5) \end{aligned}$$

and

$$\begin{aligned} \mathrm{i}\sinh(\phi) = \sin(\mathrm{i}\phi) &= \frac{\tan(\mathrm{i}\phi)}{\sqrt{1+\tan^2(\mathrm{i}\phi)}} \\ &= \frac{\mathrm{i}V/\mathrm{c}}{\sqrt{1 - V^2/\mathrm{c}^2}}\,. \qquad (6) \end{aligned}$$

Introducing (5) and (6) into (3) and (4), after substituting ic$t$ for iy, we obtain the Lorentz transformation:

$$x' = x\frac{1}{\sqrt{1-V^2/c^2}} + ict\frac{iV/c}{\sqrt{1-V^2/c^2}} = \frac{x-Vt}{\sqrt{1-V^2/c^2}}, \tag{7}$$

$$ict' = x\frac{-iV/c}{\sqrt{1-V^2/c^2}} + ict\frac{1}{\sqrt{1-V^2/c^2}} = ic\frac{t-Vx/c^2}{\sqrt{1-V^2/c^2}}. \tag{8}$$

The inverse transformation is obtained by substituting $-V$ for V:

$$x = \frac{x'+Vt'}{\sqrt{1-V^2/c^2}}, \tag{9}$$

$$ict = ic\frac{t'+Vx'/c^2}{\sqrt{1-V^2/c^2}}. \tag{10}$$

This transformation shows that space and time are not independent concepts, and that 'only a kind of union of the two will preserve an independent reality' (see reference 5).

3.2. Time dilation

Consider an observer at rest at the origin in the frame K', but uniformly moving in the frame K. A clock at rest at the origin of K' will have $x' = 0$ so will move along the line OB in figure 4. Such a path is called a *world line*. An observer in the K reference frame will regard the event B as occurring at time t and position $x = Vt$, but will see the clock moving along the x'-axis as showing time t'. We can see from figure 4 that

$$\frac{OA}{OB} = \frac{ict}{ict'} = \cos(-i\phi) = \frac{1}{\sqrt{1-V^2/c^2}}$$

and

$$t = t'\frac{1}{\sqrt{1-V^2/c^2}}.$$

This result can also be derived using the Lorentz transformation (10), with $x' = 0$. The frame in which the clock (the observer) is at rest is called its *rest frame* and the time t' shown is called its *proper time*. Observers in other frames will see this clock as going slower than their own clocks ($t > t'$).

3.3. Length contraction

Consider two particles at rest in the reference frame K'. Seen from the reference frame K, they are moving with velocity $\boldsymbol{V} = V\boldsymbol{x}$. In Minkowski space, the two particles are represented by the two world lines of figure 5. Suppose that the two particles represent the ends of a rod. In order to measure the length of the rod, one must find the co-ordinates of the two ends *at the same time*. So, in figure 5, the length of the rod in K' (its rest frame) will be AB (the events A and B have the same value of t'), but its length in K will be AC (the events A and C have the same value of t). The fact that two events are simultaneous in one frame but not in another is called *relativity of simultaneity*.

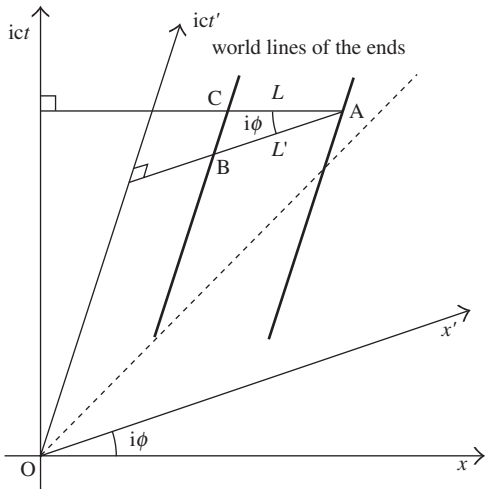


Figure 5. Length contraction and relativity of simultaneity.

We can see that

$$\frac{AB}{AC} = \frac{L'}{L} = \cos(i\phi) = \frac{1}{\sqrt{1-V^2/c^2}}$$

and

$$L = L'\sqrt{1-V^2/c^2}.$$

This result can also be derived using the Lorentz transformation (7), with $t = 0$. Observers in frames moving relative to the rest frame of the rod will see that the length of the rod is less than its length in its rest frame, its *proper length* ($L < L'$).

3.4. Addition of velocities with the same direction

Consider a reference frame K' which moves with velocity $\boldsymbol{V} = V\boldsymbol{x}$ relative to another reference frame K, and a particle moving with velocity $\boldsymbol{v}' = v'\boldsymbol{x}'$ in the reference frame K'. We have to determine the velocity $\boldsymbol{v} = v\boldsymbol{x}$ of the particle in the reference frame K.

We can see from figure 6 that

$$\frac{BA}{OB} = \frac{v'}{ic} = \tan(-i\psi),$$
$$\frac{CB}{OC} = \frac{V}{ic} = \tan(-i\phi),$$
$$\frac{DA}{OD} = \frac{v}{ic} = \tan(-i\theta).$$

These imply, respectively, that

$$\tan(i\psi) = \frac{iv'}{c}, \tag{11}$$

$$\tan(i\phi) = \frac{iV}{c}, \tag{12}$$

$$\tan(i\theta) = \frac{iv}{c}. \tag{13}$$

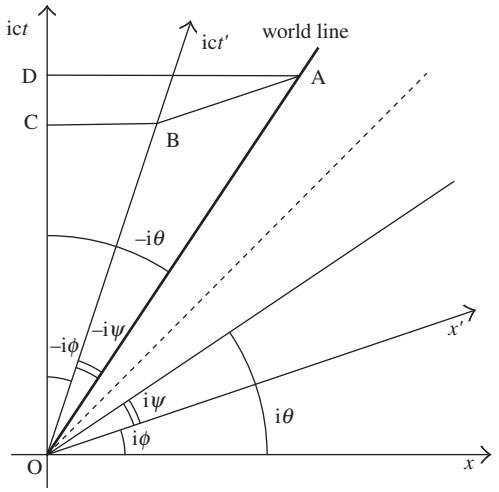


Figure 6. Relativistic addition of velocities with the same direction.

The addition of velocities is based on the addition of angles: if

$$\mathrm{i}\theta = \mathrm{i}\phi + \mathrm{i}\psi\,,$$

then

$$\tan(\mathrm{i}\theta) = \frac{\tan(\mathrm{i}\phi) + \tan(\mathrm{i}\psi)}{1 - \tan(\mathrm{i}\phi)\tan(\mathrm{i}\psi)}\,. \qquad (14)$$

Substituting the tangents from (11)–(13) into (14) we finally obtain the formula for the relativistic addition of velocities:

$$v = \frac{V + v'}{1 + Vv'/\mathrm{c}^2}\,.$$

This result can also be derived using the Lorentz transformation. Dividing (9) by (10) we obtain

$$\frac{x}{t} = \frac{x'/t' + V}{1 + V(x'/t')/\mathrm{c}^2}$$

so

$$v = \frac{v' + V}{1 + Vv'/\mathrm{c}^2}\,.$$

This expression basically states that, by combining two velocities each less than the speed of light, we cannot obtain a velocity greater than the speed of light. This is also apparent from (13), since

$$v = \mathrm{c}\tanh(\theta) = \mathrm{c}\frac{\mathrm{e}^{\theta} - \mathrm{e}^{-\theta}}{\mathrm{e}^{\theta} + \mathrm{e}^{-\theta}} < \mathrm{c}\,.$$

References

1. A. Einstein, On the electrodynamics of moving bodies, reprinted in *The Principle of Relativity*, eds H. A. Lorentz, A. Einstein, H. Minkowski and H. Weyl (Dover, New York, 1952).
2. A. Einstein, *The Meaning of Relativity*, 5th edn (Princeton University Press, 1988).
3. H. Poincaré, Sur la dynamique de l'éléctron, *Rend. Circ. Mat. Palermo* **21** (1906), pp. 129–175.
4. H. Minkowski, Space and time, reprinted in *The Principle of Relativity*, eds H. A. Lorentz, A. Einstein, H. Minkowski and H. Weyl (Dover, New York, 1952).
5. H. Minkowski, Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern, *Nachr. Gesellsch. Göttingen* (1908), pp. 53–111.

Călin Galeriu *is a graduate student at Worcester Polytechnic Institute, Massachusetts. He studies solid state physics, and is also interested in the geometrical foundations of the theory of special relativity.*

Catalan Numbers

FRAZER JARVIS

Introduction

Catalan numbers appear in many places throughout mathematics and are a fairly frequent subject for articles in *Mathematical Spectrum*. Indeed, a very nice introduction to their properties was given in this magazine by Vun and Belcher (reference 1), and a historical account of their discovery in Chinese mathematics (before Catalan!) has also appeared in a recent article by Larcombe (reference 2). The paper by Vun and Belcher gives some combinatorial situations in which Catalan numbers play a role, such as the problem of dissecting a polygon into triangles. In this note, we will briefly review the main properties of Catalan numbers and give another (possibly new) derivation of the generating function.

The main definition for Catalan numbers that we will use is the following.

Definition 1. The *nth Catalan number*, denoted C_n, is the number of ways of multiplying together n symbols.

This isn't a particularly enlightening definition, and we will see the sequence arising in other more interesting situations. But let's just think about this definition and what it means.

Suppose that $n = 3$ for simplicity, and we want to multiply a by b by c. We can essentially do this in two ways, first multiplying a and b and then multiplying by c, or by multiplying b and c and then multiplying by a.

Symbolically, these two operations are given by $(ab)c$ and $a(bc)$. As there are two ways to multiply three objects, we see that $C_3 = 2$.

Note that we are insisting that a, b and c remain in the same order: given three matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$, we will have $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$, but, in general, any permutation of the order in which $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ appear will give rise to a different product.

To multiply four objects, a, b, c and d, there are five different ways:

$$\big(\big((ab)c\big)d\big)\,,\quad \big(\big(a(bc)\big)d\big)\,,\quad \big((ab)(cd)\big)\,,$$
$$\big(a\big((bc)d\big)\big)\quad\text{and}\quad \big(a\big(b(cd)\big)\big)\,.$$

This shows that $C_4 = 5$.

How to compute the nth Catalan number

There are various ways of obtaining a simple formula for C_n, one of which is by using the generating function: we will give a method for computing this in the next section. The derivation we give now is based on one given by Singmaster (reference 3; in turn based on an idea of D. André from 1878), and relies on the idea of a *reverse Polish string*. When we do a sum like $3+4\times 6$, we can easily scan the expression to see which part of the sum we should evaluate first ($4 \times 6 = 24$ in the example), and which part should be evaluated second ($3 + 24 = 27$): we can scan the expression much faster than we can do the calculation, and so the way in which the sum is presented is not important. But computers can carry out the arithmetic (almost) instantly, and if we want the computer to do the sum as quickly as possible — very useful if variants of the calculation must be done many times — we have to think about how to present the data to the computer so that it knows the order to do the calculations, preferably using as little memory as possible. One solution is the reverse Polish method, in which the sum would be presented as a string

$$3\ 4\ 6\ \times\ +$$

in which the computer starts by reading in the values 3, 4 and 6 to memory locations. Then the '$\times$' means 'multiply the last two entries': now the memory locations consist of '3' and '24'. Finally, the '$+$' means 'add the previous two entries', so that 3 and 24 are added and the final answer 27 is output. Brackets are never required in reverse Polish notation.

Remember that C_n is the number of ways of multiplying n symbols. Let's multiply a, b and c. A reverse Polish string will be a string consisting of these symbols (in this order), and involving another symbol X, meaning 'multiply the last two things in the string'.

Consider the string $abXcX$. We first read in a and b, then we multiply to form ab, next read in c, and then multiply these two things to get $(ab)c$. The string $abcXX$ means: read in a, then b, then c; now multiply the last two things so that the string has a and bc, and finally multiply these to get $a(bc)$.

Remark 1. In fact, we can form the reverse Polish string for any product simply by deleting all left parentheses, and substituting X for all right parentheses. Thus the string $(a((bc)d))$ has reverse Polish string $abcXdXX$. Let's check that this works: we read in a, then b and then c, multiply the last two so that the string consists of a and bc, then read in d, multiply the last two to get a and $(bc)d$, and the final multiplication gives $a((bc)d)$ as required.

You might like to write down the reverse Polish strings for all five of the products of four symbols, and go through the interpretation of each product in terms of operations on the objects of the string.

Let's write S for one of the symbols $a, b, \ldots$. Then the two ways of multiplying three objects are represented by the strings $SSSXX$ and $SSXSX$.

A string representing multiplication of n objects must have n copies of the symbol S. Also, as the operation X reduces the number of objects on the string by 1, replacing two objects with their product, and as we have to go from n objects to one product, the string must have $n-1$ copies of the symbol X. There are therefore $2n-1$ symbols in the string, with n of these being S and the other $n-1$ being X.

However, not every possible string represents a valid way of multiplication. We can only apply the symbol X when there are at least two objects in the string. For $n = 2$, the only valid string is SSX; both of the strings XSS and SXS would involve multiplication of fewer than two symbols! It is easy to see that a string of Ss and Xs is legitimate precisely when the number of Ss in the first k symbols of the string is more than the number of Xs for every value of k.

We can represent the string by means of a graph: if the next symbol is S, then we move one unit upwards in the y-direction, and if it is X, then we move one unit downwards.

For example, the string $SSSXSXX$ of remark 1 has the graph in figure 1. The line always begins at $(0, 0)$ and ends at $(2n-1, 1)$.

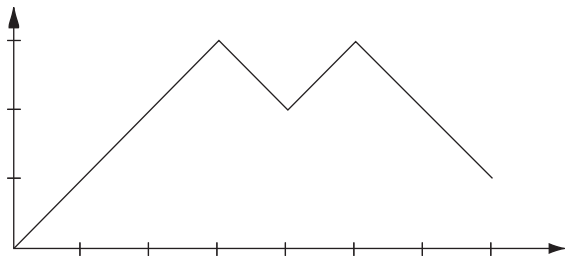


Figure 1. Graph for the string $SSSXSXX$.

The condition that a string be legitimate is that the graph always lies above the x-axis. So the first move *must* be from $(0, 0)$ to $(1, 1)$. Now we have to count the number of walks from $(1, 1)$ to $(2n-1, 1)$ which do not touch (or cross) the axis. There are $\binom{2n-2}{n-1}$ walks from $(1, 1)$ to $(2n-1, 1)$ in total, obtained by choosing $n-1$ steps upwards and making the remaining $n-1$ go downwards; we will count the number that do touch the axis and subtract to get the number we want.

We have the following result:

> The number of walks from $(1, 1)$ to $(2n-1, 1)$ which touch the axis is the same as the number of walks from $(1, 1)$ to $(2n-1, -1)$.

We will show that there is a correspondence between these two sorts of walks. The idea is very simple. As soon as a walk touches the axis, we reflect the rest of the walk in the axis. So if we have a walk such as in figure 2, then we reflect the part of the graph after time $t = 4$ to get the graph in figure 3.

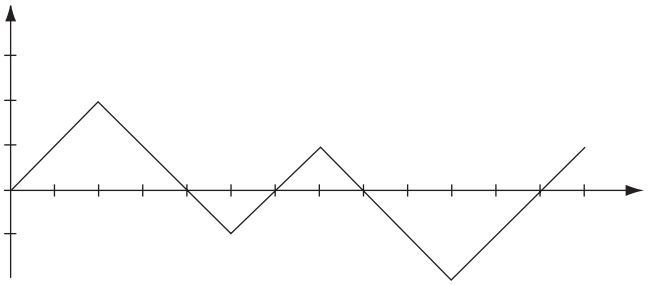


Figure 2.

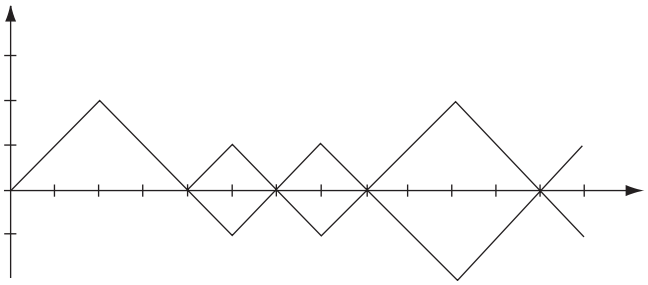


Figure 3.

It is easy to construct a converse: given any walk from $(1, 1)$ to $(2n-1, -1)$, it must hit the x-axis at some point. If we reflect the graph from that point onwards, then we get a walk from $(1, 1)$ to $(2n-1, 1)$ which touches the axis.

Now, any walk from $(1, 1)$ to $(2n-1, -1)$ must go upwards $n-2$ times and downwards n times, and so there are $\binom{2n-2}{n}$ such paths. It follows that the number of walks from $(1, 1)$ to $(2n-1, 1)$ which do *not* touch the x-axis is

$$\begin{aligned}\binom{2n-2}{n-1} - \binom{2n-2}{n} &= \frac{(2n-2)!}{(n-1)!\,(n-1)!} - \frac{(2n-2)!}{n!\,(n-2)!} \\ &= \frac{(2n-2)!}{(n-1)!\,(n-2)!}\left(\frac{1}{n-1} - \frac{1}{n}\right) \\ &= \frac{(2n-2)!}{(n-1)!\,(n-2)!}\frac{1}{n(n-1)} \\ &= \frac{1}{n}\frac{(2n-2)!}{(n-1)!\,(n-1)!} \\ &= \frac{1}{n}\binom{2n-2}{n-1}.\end{aligned}$$

These kinds of graph also appeared in the article of Vun and Belcher: they may be interpreted as representing the problem of counting the number of ways in which a tied election between two candidates can be held in which one candidate was always ahead of the rival, except before the first vote was cast and after the last vote was cast.

The generating function

If we have a sequence of numbers, $a_0, a_1, a_2, \ldots$, then we can form the series

$$a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots .$$

This series is known as the *generating function* for the sequence. Often the generating function allows you to perform calculations with the terms in the sequence, and thus to give easier ways to prove certain identities. There is a lovely book by Wilf (reference 4), just about generating functions and how to manipulate them to get interesting results.

One interesting feature is that it is often possible to give other ways to write generating functions. For example, the sequence $1, 2, 3, \ldots$ has the generating function

$$1 + 2x + 3x^2 + 4x^3 + \cdots ,$$

but we might recognise this as the expansion of $1/(1-x)^2$. This single function therefore encodes the sequence of natural numbers (positive integers). More complicated sequences can also be similarly encoded; the Fibonacci numbers, defined by $F_0 = F_1 = 1$, $F_{n+1} = F_n + F_{n-1}$ for $n \geq 1$, arise as the coefficients of the function $1/(1-x-x^2)$. (To prove this, let $F(x) = 1 + x + 2x^2 + 3x^3 + 5x^4 + \cdots$ be the generating function and write down $xF(x)$ and $x^2F(x)$. Now add them, and note that you get $F(x) - 1$. Now rearrange the equality $x^2F(x) + xF(x) = F(x) - 1$ to get the above expression for $F(x)$.)

Here's a way of finding the generating function for the Catalan numbers, which is not in reference 4.

The generating function is

$$C(x) = C_0 + C_1 x + C_2 x^2 + C_3 x^3 + C_4 x^4 + \cdots ,$$

where C_n, the nth Catalan number, is the number of ways of bracketing products together, as mentioned above. Thus,

$$\begin{aligned}C(x) &= x + x^2 + 2x^3 + 5x^4 + \cdots \\ &= x + x^2 + [x^3 + x^3] + [x^4 + x^4 + x^4 + x^4 + x^4] \\ &\quad + \cdots \\ &= x + (x \cdot x) + \big((x \cdot x)x\big) + \big(x(x \cdot x)\big) \\ &\quad + \Big(\big((x \cdot x)x\big)x\Big) + \Big(\big(x(x \cdot x)\big)x\Big) \\ &\quad + \big((x \cdot x)(x \cdot x)\big) + \Big(x\big((x \cdot x)x\big)\Big) \\ &\quad + \Big(x\big(x(x \cdot x)\big)\Big) + \cdots ,\end{aligned}$$

where each term corresponds to multiplying x by itself using all the possible bracketings. Observe that, apart from the first term, if we strip off the outer pair of brackets, every term is naturally the product of two smaller terms in the series for $C(x)$:

$$\begin{aligned}C(x) = x &+ x \cdot x + (x \cdot x)x + x(x \cdot x) \\ &+ \big((x \cdot x)x\big)x + \big(x(x \cdot x)\big)x + (x \cdot x)(x \cdot x) \\ &+ x\big((x \cdot x)x\big) + x\big(x(x \cdot x)\big) + \cdots\end{aligned}$$

and a term like $((x \cdot x)x)x$ is the product of the two terms $((x \cdot x)x)$ and x. In fact, it is easy to see that, apart from the first term, the other terms are exactly $C(x)^2$ (all terms apart from the first are a product of two smaller terms, and any product of two smaller terms will arise as a way of multiplying x by itself a suitable number of times, so will appear in $C(x)$). Thus

$$C(x) = x + C(x)^2 ,$$

so that $C(x)^2 - C(x) + x = 0$. We can solve this with the quadratic formula to give

$$C(x) = \frac{1 \pm \sqrt{1-4x}}{2} .$$

In fact, we must choose the minus sign here, otherwise the coefficients of the powers of x in the generating function of $C(x)$ are all negative, whereas we want $C(x)$ to be the generating function of the Catalan numbers, all of which are positive. Indeed, if we expand the square root $(1-4x)^{1/2}$ as a series $1 + a_1x + a_2x^2 + \cdots$ using the binomial formula, we have

$$\sqrt{1-4x} = 1 - 2x - 2x^2 - \cdots ,$$

and so we need to choose the minus sign to get positive coefficients in $C(x)$. Then

$$C(x) = \frac{1 - \sqrt{1-4x}}{2} .$$

The binomial formula tells us that the coefficient of x^n (if $n > 0$) in this series is

$$\begin{aligned} C_n &= -\tfrac{1}{2}\left\{ \frac{\frac{1}{2}(-\frac{1}{2})(-\frac{3}{2})\cdots(-(2n-3)/2)}{n!}(-4)^n \right\} \\ &= \tfrac{1}{2}\left\{ \frac{\frac{1}{2}(\frac{1}{2})(\frac{3}{2})\cdots((2n-3)/2)(n-1)!}{n!\,(n-1)!}\,(2^2)^n \right\} \\ &= \tfrac{1}{2}\left\{ \frac{\frac{1}{2}[\frac{1}{2}\cdot 1\cdot\frac{3}{2}\cdot 2\cdots(n-2)((2n-3)/2)(n-1)]}{n!\,(n-1)!}2^{2n} \right\} \\ &= \tfrac{1}{2}\left\{ \frac{\frac{1}{2}(1\cdot 2\cdot 3\cdot 4\cdots(2n-4)(2n-3)(2n-2)}{n!\,(n-1)!}2^2 \right\} \\ &= \frac{(2n-2)!}{n!\,(n-1)!} \\ &= \frac{1}{n}\binom{2n-2}{n-1} . \end{aligned}$$

References

1. I. Vun and P. Belcher, Catalan numbers, *Math. Spectrum* **30** (1997/98), pp. 3–5.
2. P. Larcombe, The 18th century Chinese discovery of the Catalan numbers, *Math. Spectrum* **32** (1999/2000), pp. 5–6.
3. D. Singmaster, An elementary evaluation of the Catalan numbers, *Amer. Math. Monthly* **85** (1978), pp. 366–368.
4. H. S. Wilf, *Generatingfunctionology* (Academic Press, Boston, 1994). (See also http://www.cis.upenn.edu/∼wilf/.)

Frazer Jarvis *is a member of the Department of Pure Mathematics at the University of Sheffield and his research interests are centred on algebraic number theory. Outside mathematics, he is a keen pianist and sings with the Sheffield Philharmonic Chorus.*

Distance From a Point to a Line in the Taxicab Geometry

AUGUSTO J. M. WANDERLEY, JOSÉ PAULO CARNEIRO and EDUARDO WAGNER

1. Introduction

In the usual plane Euclidean geometry, the least distance between two points is given by the measure of the segment connecting both points. On the other hand, if we are in a planned city where all the streets are mutually perpendicular or parallel (see figure 1), the least distance between points A and B is not given by the usual AB but by the sum AC + CB or by other equivalent sums, such as AD + DE + EF + FB.

In general, to work with urban geography, a more convenient model is the so-called 'taxicab geometry', because the distances followed by a taxicab are measured by means of paths like the ones above: the taxicab has to obey the planning of the streets and it cannot fly in a straight line like a bird.

This geometry consists of the usual plane, where we have fixed an orthogonal system of coordinates, and the usual points and lines. Therefore, the 'taxicab distance' from $A = (x_A, y_A)$ to $B = (x_B, y_B)$ is defined by: $d_T(A, B) = |x_A - x_B| + |y_A - y_B|$ (see figure 2).

If a church is located at $O = (0, 0)$ (see figure 3) and a restaurant at $B = (-1, 5)$, then the church is nearer to the restaurant than to a drugstore at $A = (3, 4)$,

since $d_T(A, O) = |3 - 0| + |4 - 0| = 7 > d_T(B, O) = |-1 - 0| + |5 - 0| = 6$. In the Euclidean geometry, we would have $d_E(A, O) = 5 < d_E(B, O) = \sqrt{26}$ and the church would be nearer to the drugstore than to the restaurant. The taxicab geometry represents the urban reality better than the Euclidean geometry.

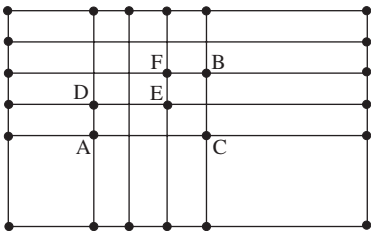


Figure 1.

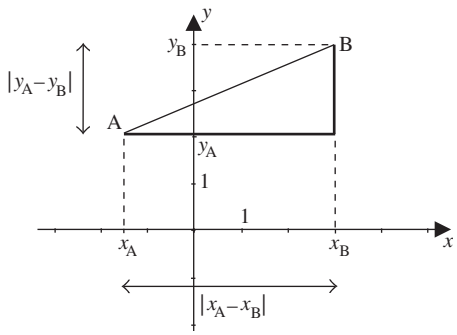


Figure 2.

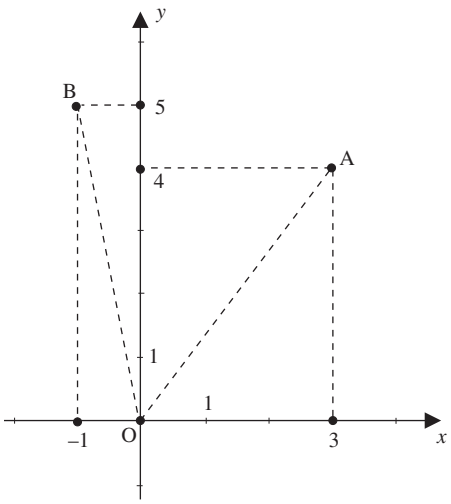


Figure 3.

Many of the properties of the taxicab geometry are similar to those of the Euclidean one. For instance, it is not hard to prove that the taxicab distance is always non-negative and it is zero if and only if the two points are equal, that it is symmetric with respect to the points A and B and that it satisfies the triangular inequality. Moreover, if A and B lie on the same horizontal line or on the same vertical line, then $d_T(A, B) = d_E(A, B)$.

Nevertheless the taxicab distance is, in many aspects, very different from the Euclidean distance and it provides many surprises. For example, if $A = (0, 0)$ and $B = (1, 1)$, then $d_T(A, B) = 2$ and a rotation of 45° of the segment AB around A will keep A fixed, changes B into $B' = (0, \sqrt{2})$ and therefore $d_T(A, B') = \sqrt{2} \neq d_T(A, B)$. Note that this just reflects the importance of the 'street map' in the taxicab distance.

Another surprise is obtained when we try to find the picture that corresponds to a taxicab circle of centre C and radius r. To find out what it looks like, notice that the equation for such a circle with $C = (x_C, y_C)$ is

$$|x - x_C| + |y - y_C| = r\,.$$

Since the taxi distance is translation invariant, we will consider the case of a taxi circle with centre at the origin. Figure 4 suggests that such a circle is the (Euclidean) square shown, since any point on that square has its taxicab distance to the origin equal to r.

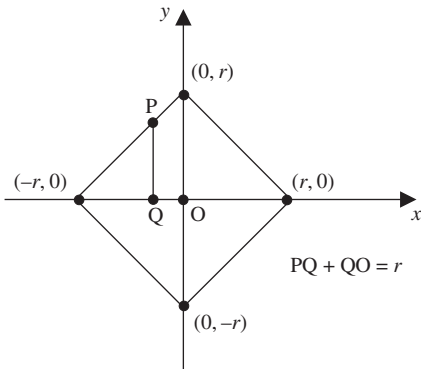


Figure 4.

This problem may be solved algebraically, since the equation $|x| + |y| = r$ is equivalent to the system

$$\begin{cases} x + y = r & \text{if } x \geq 0 \text{ and } y \geq 0\,, \\ -x + y = r & \text{if } x \leq 0 \text{ and } y \geq 0\,, \\ -x - y = r & \text{if } x \leq 0 \text{ and } y \leq 0\,, \\ x - y = r & \text{if } x \geq 0 \text{ and } y \leq 0\,. \end{cases}$$

2. Distance from a point to a line

We will consider the following problem: given a point P and a line l in the plane, determine the distance from P to l in the taxicab geometry, that is, the minimum value of the distance from P to a point of l. If we are working with the Euclidean geometry, we would obtain the answer by means of the point Q where the line perpendicular to l and passing by P intersects l. With such a point,

$$d_E(P, l) = d_E(P, Q)\,.$$

A geometric way to obtain such a point Q would be to imagine circles with centre P with increasing radii. One of these circles must touch the line l (see figure 5). When this happens we would have the point Q of intersection of such a circle with the line. We could visualize this geometric idea with some dynamic geometry software such as Cabri or Cinderella.

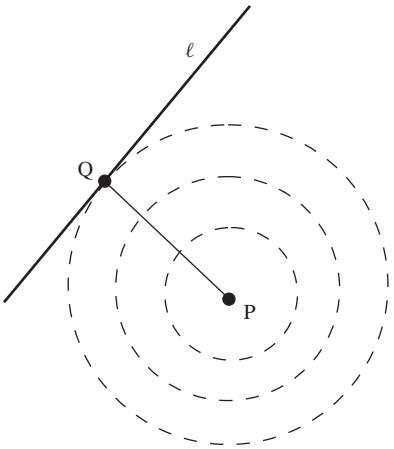
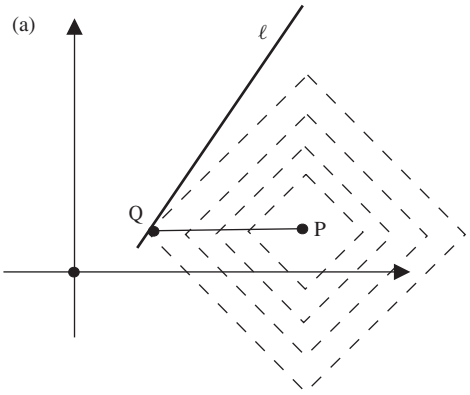


Figure 5.



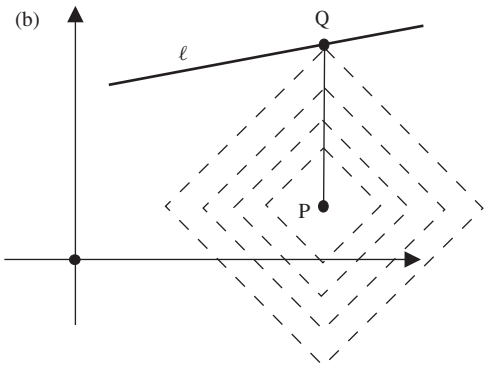


Figure 6.

In a similar way, in the taxicab geometry, let us imagine a sequence of taxicab circles with centre P and radii growing until one of them 'touches' the line l (see figure 6). We can see that the slope of l will have a crucial role in the discussion. For example, figure 6(a) suggests that if l has a slope greater than 1, then the distance from P to l should be the distance from P to Q which is the point of intersection of l with the 'horizontal' line (parallel to the x-axis) through P, and figure 6(b) suggests that if l has a slope less than 1, then the point Q of l where the distance from P to l would attain its minimum would be the intersection of l with the 'vertical' line (parallel to the y-axis) through P.

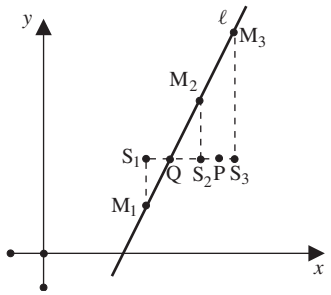


Figure 7.

These considerations will now be confirmed. In figure 7, l has a slope greater than 1. If we take some points M_1, M_2, M_3 on l, we will show that the minimum is indeed attained at Q:

$$\begin{aligned} d_T(P, M_1) &= PS_1 + S_1M_1 \\ &> PS_1 > PQ \\ &= d_T(P, Q)\,, \\ d_T(P, M_2) &= PS_2 + S_2M_2 \\ &> PS_2 + S_2Q = PQ \\ &= d_T(P, Q)\,, \\ d_T(P, M_3) &= PS_3 + S_3M_3 \\ &> S_3M_3 > S_3Q > PQ \\ &= d_T(P, Q)\,. \end{aligned}$$

The reader may consider other situations with slopes greater or smaller than 1. If the slope is 1 or -1, then figure 8 shows that for every point M of the segment Q_1Q_2 we have

$$\begin{aligned} d_T(P, M) &= PS + SM \\ &= Q_1P = PQ_2 \\ &= d_T(P, Q_1) = d_T(P, Q_2)\,. \end{aligned}$$

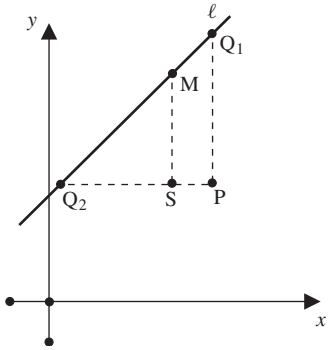


Figure 8.

If l is parallel to one of the axes, then the taxicab distance from P to l will be equal to the Euclidean distance.

Let us now verify such conjectures algebraically.

3. The algebraic approach

Let $P = (x_P, y_P)$ and

$$\begin{cases} x = x_0 + ut\,, \\ y = y_0 + vt\,, \end{cases}$$

be parametric equations for the line l, where $(x_0, y_0) \in l$ and (u, v) is a non-null vector parallel to l, with $t \in \mathbb{R}$. The distance from P to l will be the minimum value of

$$f(t) = |x - x_P| + |y - y_P| \\ = |x_0 + ut - x_P| + |y_0 + vt - y_P| .$$

3.1. Lines parallel to the axes

If $u = 0$, that is, if l is the vertical line $x = x_0$, then $f(t) = |x_0 - x_P| + |y_0 + vt - y_P|$. In this case, $f(t)$ attains its minimum when $y_0 + vt - y_P = 0$, that is, when $t = (y_P - y_0)/v$, and the minimum value of $f(t)$ is $|x_0 - x_P|$. Therefore, in such a case the taxicab distance $d_T(P, l)$ from P to l coincides with the Euclidean distance from P to l (see figure 9(a)).

A similar result occurs when l is the horizontal line $y = y_0$ (see figure 9(b)).

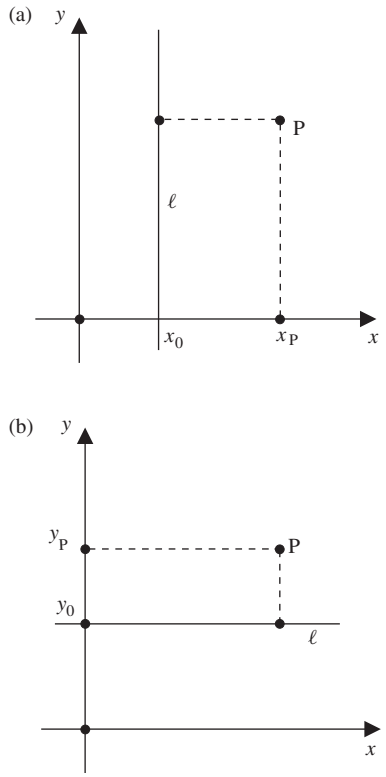


Figure 9.

3.2. Lines not parallel to any one of the axes

From now on we will consider a line l not parallel to any one of the axes, so that $u, v \neq 0$, and we rewrite $f(t)$ as

$$f(t) = |u|\left|t - \frac{x_P - x_0}{u}\right| + |v|\left|t - \frac{y_P - y_0}{v}\right| .$$

To simplify, we write $|u| = w$, $|v| = z$,

$$\frac{x_P - x_0}{u} = c \quad \text{and} \quad \frac{y_P - y_0}{v} = d .$$

We now have

$$f(t) = w|t - c| + z|t - d| .$$

Let us assume, in the first place, that $c < d$. Then

$$f(t) = \begin{cases} -(w + z)t + (wc + zd) & \text{if } t < c , \\ (w - z)t + (zd - wc) & \text{if } c \leq t < d , \\ (w + z)t - (wc + zd) & \text{if } d \leq t . \end{cases}$$

Figure 10 sketches the graphs of f in the cases $w < z$, $w = z$ and $w > z$, showing that the minimum value of f is always $d - c$ multiplied by the smaller of the numbers w and z.

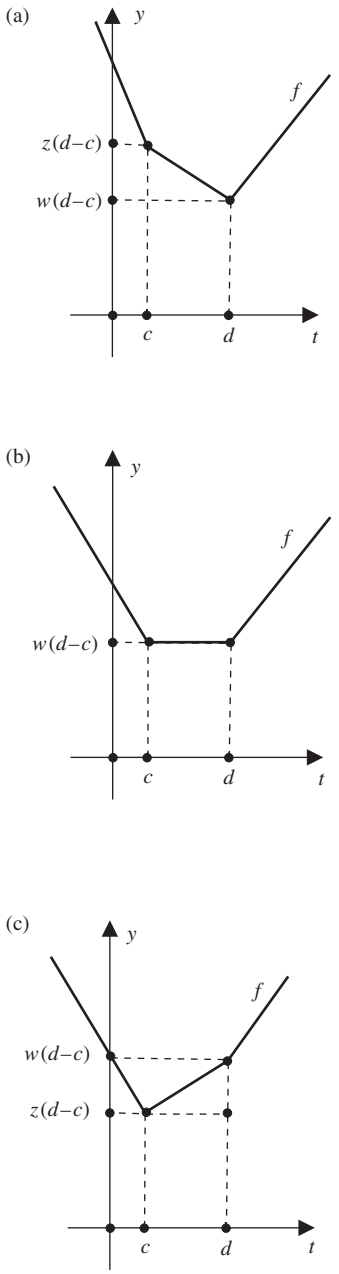


Figure 10. Graphs of f when (a) $w < z$, (b) $w = z$ and (c) $w > z$.

It is interesting to remark that this minimum value is attained if $w < z$ when $t = d$, and if $w > z$ when $t = c$. When $w = z$, the minimum value is attained at an infinity of points, namely at all the points of the interval $[c, d]$.

The case $d < c$ is completely analogous, with the minimum value of f being then equal to $c-d$ multiplied by the smaller of the numbers w and z. This minimum value is attained if $w < z$ when $t = d$, and if $w > z$ when $t = c$. When $w = z$, the minimum value is attained in all the points of the interval $[d, c]$.

Finally, when $c = d$, the minimum value of f is zero and it is attained when $t = c = d$.

In all the cases we see that the minimum value is always $|c-d|$ multiplied by $\min\{w, z\}$. This value is attained when $w < z$ at $t = d$, and when $w > z$ at $t = c$, and when $w = z$ at every point of the interval $[m, M]$, where $m = \min\{c, d\}$ and $M = \max\{c, d\}$.

These results may be translated to our original problem by:

$$d_{\mathrm{T}}(\mathrm{P}, l) = \min\{|u|, |v|\}\left|\frac{x_{\mathrm{P}} - x_0}{u} - \frac{y_{\mathrm{P}} - y_0}{v}\right|.$$

This number is the distance from P to (at least) one point of l. When $|u| \neq |v|$ (which means that the slope of l is different from 1 or –1), this point will be $Q = (x_0 + ut, y_0 + vt)$, where

$$t = \begin{cases} \dfrac{x_{\mathrm{P}} - x_0}{u} & \text{if } |u| > |v| \\ & \text{(slope with absolute value less than 1)}, \\ \dfrac{y_{\mathrm{P}} - y_0}{v} & \text{if } |u| < |v| \\ & \text{(slope with absolute value greater than 1)}. \end{cases}$$

By substitution, we see that such a point may be rewritten as

$$\mathrm{Q}_1 = \left(x_{\mathrm{P}}, y_0 + \frac{v}{u}(x_{\mathrm{P}} - x_0)\right)$$

or

$$\mathrm{Q}_2 = \left(x_0 + \frac{u}{v}(y_{\mathrm{P}} - y_0), y_{\mathrm{P}}\right)$$

respectively. Since the first coordinate of Q_1 is the same as that of P, and the second coordinate of Q_2 is the same as that of P, the points Q_1 and Q_2 are, respectively, the intersections of l with the lines parallel to the axes y and x through P, as was expected from the figures.

4. Cartesian equation

If the line l has a Cartesian equation $ax + by + c = 0$, we may take $(u, v) = (b, -a)$. If $a, b \neq 0$, the point $(-c/a, 0)$ is on l and then

$$\begin{aligned} d_{\mathrm{T}}(\mathrm{P}, l) &= \min\{|a|, |b|\}\left|\frac{x_{\mathrm{P}} + c/a}{b} - \frac{y_{\mathrm{P}} - 0}{-a}\right| \\ &= \min\{|a|, |b|\}\frac{|ax_{\mathrm{P}} + by_{\mathrm{P}} + c|}{|a||b|} \\ &= \frac{|ax_{\mathrm{P}} + by_{\mathrm{P}} + c|}{\max\{|a|, |b|\}}. \end{aligned}$$

This formula can also be used for the case of lines parallel to one of the coordinate axes discussed earlier. Therefore, in any case, if l has equation $ax+by+c = 0$ and $\mathrm{P} = (x_{\mathrm{P}}, y_{\mathrm{P}})$, then

$$d_{\mathrm{T}}(\mathrm{P}, l) = \frac{|ax_{\mathrm{P}} + by_{\mathrm{P}} + c|}{\max\{|a|, |b|\}}.$$

This result could also be obtained from the geometrical interpretation given in section 2. For example, if the slope of l has absolute value greater than 1 (and therefore $|a| > |b|$), the point Q is the point of l with $x = x_{\mathrm{P}}$. From the equation of l, we get $y = -(ax_{\mathrm{P}} + c)/b$ and calculating $d_{\mathrm{T}}(\mathrm{P}, \mathrm{Q})$, we obtain

$$d_{\mathrm{T}}(\mathrm{P}, l) = \frac{|ax_{\mathrm{P}} + by_{\mathrm{P}} + c|}{|a|}.$$

5. Remark

The set $\mathbb{R}^2$, with its usual lines, is an example of an *incidence geometry*. Moreover, if for each line l we define $f_l : l \to \mathbb{R}$ by $f_l(a, y) = y$ if $l = \{(x, y) \in \mathbb{R}^2; x = a\}$ and $f_l(x, y) = (1 + |m|)x$ if $l = \{(x, y) \in \mathbb{R}^2; y = mx + b\}$, then we see that each line has a system of coordinates, that is, for each line l in $\mathbb{R}^2$, there is a bijection $f_l : l \to \mathbb{R}$ such that $|f_l(\mathrm{P}) - f_l(\mathrm{Q})| = d_{\mathrm{T}}(\mathrm{P}, \mathrm{Q})$ for all $\mathrm{P}, \mathrm{Q} \in l$. Therefore $\mathbb{R}^2$, with its usual lines and the distance d_{T}, is a *metric geometry*. Since the d_{T}-segments are the usual Euclidean segments, such a metric geometry satisfies the classical plane separation axiom. The usual Euclidean measure for angles is also an angle measure for such a taxicab plane. However, in such a geometry in $\mathbb{R}^2$, the classical theorem of congruence of triangles (side–angle–side) does not hold. To show this, let us consider the triangles ΔAOB and ΔMOP,

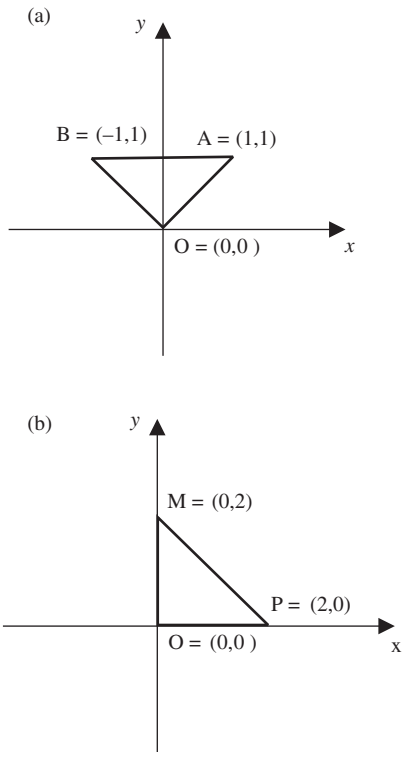


Figure 11.

with $A = (1, 1)$, $O = (0, 0)$, $B = (-1, 1)$, $M = (0, 2)$, $O = (0, 0)$ and $P = (2, 0)$. We have

$$d_T(O, A) = 2 = d_T(O, P) = d_T(O, B) = d_T(O, M),$$

and

$$\angle BOA = \angle MOP = 90^\circ$$

(see figure 11), but such triangles are not d_T-congruent, since there is no bijection $\phi : \{A, O, B\} \to \{P, O, M\}$ with

$$\begin{aligned} d_T(A, B) &= 2 = d_T(\phi(A), \phi(B)), \\ d_T(A, O) &= 2 = d_T(\phi(A), \phi(O)) \\ &= d_T(O, B) = d_T(\phi(O), \phi(B)) \end{aligned}$$

and

$$\begin{aligned} \angle AOB &= \angle\phi(A)\phi(O)\phi(B), \\ \angle OAB &= \angle\phi(O)\phi(A)\phi(B), \\ \angle OBA &= \angle\phi(O)\phi(B)\phi(A). \end{aligned}$$

To persuade yourself of this fact, you may try, for instance, $\phi(A) = P$, $\phi(O) = O$ and $\phi(B) = M$, and see that it does not work.

Therefore, the taxicab plane is not an example of a neutral (or absolute) geometry. As a matter of fact, we have the following theorem.

Theorem. *In a neutral geometry, given a line l and a point $P \notin l$, the distance $d(P, l)$ is equal to $d(P, Q)$, where Q is the foot of the perpendicular line from P to l.*

For a proof, see chapter 18 of reference 1.

Reference

1. G. Martin, *The Foundations of Geometry and the Non-Euclidean Plane*, (Intext Educational Publishers, New York, 1972).

The first two authors obtained their PhD degrees at the Federal University of Rio de Janeiro in the areas of complex functions and theory of approximation respectively, and the third author obtained his MSc degree in the Institute of Pure and Applied Mathematics, Rio de Janeiro. They have also worked on government programmes for the training of teachers in Brazil and in other countries. They have published articles in their research areas and articles in connection with the teaching of mathematics, and work in the State University of Rio de Janeiro (the first two authors) and in the Fundação Getulio Vargas, Rio de Janeiro (the third author).

Mathematics in the Classroom

The recurring problem of nines

If I ask any of my students, from year 7 right up to year 13, to express $0.\dot{3}$ as a fraction, then I am confident that they will come back with the correct answer, i.e. $\frac{1}{3}$. If I follow up with a request for some kind of justification or proof, then I will receive a variety of answers, mainly depending on which year the student is in, and these will usually be based on techniques that I have shown them. I shall describe three different ways of showing that $0.\dot{3} = \frac{1}{3}$, and share my experience of how these are perceived by students when moving on to the slightly trickier problem of showing that $0.\dot{9} = 1$.

Method 1. The simplest way of convincing my year 7 students that $0.\dot{3} = \frac{1}{3}$ is by division:

$$\begin{array}{r} 0.\,3\;\;3\;\;\dot{3} \\ 3\overline{)\,1.{}^{1}0\;{}^{1}0\;{}^{1}0} \end{array}$$

and they should be familiar with this from primary school.

Method 2. The next method comes along in year 8 or 9 when we start to look in more detail at expressing recurring decimals as fractions. Setting $x = 0.333\ldots$ and then multiplying by 10 and subtracting,

$$\begin{array}{rl} 10x = & 3.333\ldots \\ - \quad x = & 0.333\ldots \\ \hline 9x = & 3 \end{array}$$

so $x = 0.\dot{3} = \frac{3}{9} = \frac{1}{3}$.

Method 3. The last method is reserved for year 12 students as a simple application of geometric series. Writing $0.\dot{3}$ as a series,

$$\begin{aligned} 0.\dot{3} &= \frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \cdots \\ &= \frac{3}{10} + \frac{3}{10^2} + \frac{3}{10^3} + \cdots \\ &= 3\Big(\frac{1}{10} + \frac{1}{10^2} + \frac{1}{10^3} + \cdots\Big) \\ &= 3 \times \frac{\frac{1}{10}}{1 - \frac{1}{10}} \\ &= 3 \times \frac{1}{9} \\ &= \frac{1}{3}, \end{aligned}$$

where we have used the sum of the geometric series $a + ar + ar^2 + \cdots = a/(1-r)$ (provided $-1 < r < 1$). Of course, the proof of this sum follows the idea in method 2 of multiplying the sum by the common ratio, r, and subtracting. Not surprisingly, my year 12 students much prefer to use either of the first two methods!

Turning to the problem of recurring nines, most of my students are surprised when I tell them that $0.\dot{9} = 1$. For year 12, I can use method 3 as follows:

$$\begin{aligned} 0.\dot{9} &= \frac{9}{10} + \frac{9}{100} + \frac{9}{1000} + \cdots \\ &= \frac{9}{10} + \frac{9}{10^2} + \frac{9}{10^3} + \cdots \\ &= 9\Big(\frac{1}{10} + \frac{1}{10^2} + \frac{1}{10^3} + \cdots\Big) \\ &= 9 \times \frac{\frac{1}{10}}{1 - \frac{1}{10}} \\ &= 9 \times \frac{1}{9} \\ &= 1\,. \end{aligned}$$

But some of them obviously want to see how this works for the simpler alternative in method 2, and of course younger students are not familiar with geometric series. Setting $x = 0.999\ldots$ and then multiplying by 10 and subtracting,

$$\begin{array}{rl} & 10x = 9.999\ldots \\ - & \quad x = 0.999\ldots \\ \hline & 9x = 9 \end{array}$$

so $x = 0.\dot{9} = \frac{9}{9} = 1$. Even then this result can be met with some sceptism by year 12 students and those lower down the school. What seems to get almost universal approval, however, is the following very simple approach.

Starting with $0.\dot{3} = \frac{1}{3}$ (which is easily obtained by division as shown above using method 1), and then multiplying by 3 shows, very simply, that

$$\begin{aligned} 0.333\cdots &= \frac{1}{3} \\ \implies 3 \times 0.333\cdots &= 3 \times \frac{1}{3} \\ \implies 0.999\cdots &= 1\,, \end{aligned}$$

that is, $0.\dot{9} = 1$. This is readily accessible to all my students, particularly those in year 7 who are comfortable with division but have not yet met method 2.

I should be interested to hear of other teachers' classroom experience with this problem.

Kendrick School, Reading **Elizabeth M. Glaister**

Computer Column

Tough problems

What makes a particular problem hard? In particular, what makes it hard to solve by computer? These may seem like straightforward questions, and it is easy to come up with plausible-sounding intuitive answers to them. However, as with a number of other questions in mathematics, pinning down a complete answer is proving to be, well, hard. In fact, the Clay Mathematics Institute is offering a million-dollar prize to the first person to come up with one!

The problem — as posed by the Clay Institute — concerns two classes of problem, known as P and NP. A problem is said to be in the P (or polynomial-time) class if it is possible to write a program which is guaranteed to find the solution within a length of time which is a polynomial in the size of the problem. For example, imagine a map with N cities and a network of roads running between them, and that I want to write a program to find the shortest route from one city to another. If the time my program takes to run is, at worst, proportional to N^3, say, then I know that the problem is in the P class. (Such programs do exist, forming the basis for practical journey planners.)

The other class, NP (or non-deterministic polynomial-time) is rather different: for a problem to qualify for this class, all that we require is that, given a possible answer to the problem, we can write a program to check whether or not the answer is correct in polynomial time. Now, any problem in P must also be in NP (we could, for example, check an answer by recalculating it ourselves and then comparing the two), but it looks like the NP class ought to be bigger than P, given that the definition looks a lot less restrictive. Not surprisingly, this is what most people believe, but — and this is where the Clay Institute comes in — no one has yet been able to prove it. (The reason for the name is that, if we wrote a program to solve one of these problems that included the ability to make random guesses at certain points, it could succeed in polynomial time by being very lucky.)

There are a wide variety of problems which are commonly believed to be in NP but not P; this is because the best programs anyone has been able to write for them run not in polynomial time but in exponential time. Rather than being proportional to N^3, say, such a program's running time might be proportional to 3^N. The difference this makes is quite startling; for example, imagine two programs, with

running times proportional to N^3 and 3^N respectively. If the polynomial-time program can solve a small example of the problem (say, $N = 10$) in one second, then it would take eight seconds over $N = 20$, 27 seconds over $N = 30$ and a little over a minute for $N = 40$. By contrast, if the exponential-time program can also solve $N = 10$ in one second, it would take more than 16 hours for $N = 20$, 110 years for $N = 30$ and more than six and a half million years for $N = 40$!

Many of these problems often do not seem very different from their P-class brethren; for example if, rather than finding the shortest route between two cities, we want to find a tour round *all* the cities, then we suddenly find ourselves apparently outside the P class. When we ask for the *shortest* such tour, this is known as the travelling salesman problem; when we simply ask whether or not it is *possible* to go round all the cities without revisiting any of them, it is known as the Hamiltonian path problem.

A curious feature of some NP problems (including these two) is that all other NP problems can be restated in terms of them. In other words, if I have a different NP problem to solve, I can restate it in terms of a travelling salesman problem, solve that, and then use the answer to answer my original problem. Of course, converting back and forth takes a certain amount of effort, but it is known that the programs involved can be written to run in polynomial time. Such problems are said to be NP-complete, and are interesting in that they must have within them all the features that make an NP problem hard. If a program could be found which solved just one of these in polynomial time, that would be enough to show that *all* problems in NP also lie in P, i.e. that P and NP are in fact the same class. Conversely, if we want to try and prove that some problems in NP lie outside P, these should be good places to start.

With all this talk of exponential-time programs and running times in the millions of years, it may come as a bit of a surprise to learn that most examples of NP problems are actually fairly easy in practice. If we give up the notion of guaranteeing an exact solution and settle for hopefully being able to find a reasonable one, then it turns out that we can usually come up with a pretty good answer. A real travelling salesman, after all, is likely to be happy with a route that's a few miles longer than the best one if he can find it in a matter of minutes rather than years! In a number of cases, an exact solution is even possible: the current record for the largest exactly-solved example is 15 112 cities, for which the worst-case running time really doesn't bear thinking about!

To see why these problems can be so difficult, think about what happens if you start from a bad solution and try to improve it. If you're trying to find the best route between two places, one way to go about it would be to pick a section of the route and try to find a shorter path between the start of the section and the end of it. If you managed to find one, then you would have a better solution to the problem as a whole. If you try to do the same thing for a travelling salesman problem there's a problem: changing one part of the route affects what you're able to do elsewhere. If you take one section of the route through a different city, then you'll also have to adapt the part that previously went through it. This is the general difficulty with NP-complete problems: changing one part tends to affect the whole problem. For polynomial-time problems, by contrast, it is usually possible to split the problem into pieces which can be solved individually. This also explains why some NP-complete problems turn out to be easy after all: if it so happens that there is a way of dividing the problem into pieces, then life becomes very much easier.

Another reason NP-complete problems can be easy is if they turn out to have a lot of different solutions: finding any given solution is still hard, but finding *one* of them is easier the more there are. In fact, it turns out that there are only usually a few examples of any given problem that are genuinely difficult; in most cases, the problem is either too constrained (making it easy to show that there are no solutions), or too unconstrained (meaning that there are many solutions).

So, where does all this get us? On one hand, it tells us that we shouldn't despair if we run into an NP-complete problem; we just have to settle for approximate solutions. On the other hand, it holds out the possibility of encrypting information by hiding it in one of the few truly hard examples; this is the basis of RSA encryption, which is used to make internet shopping secure, for example. This uses the fact that factoring a large number is known to be difficult[1] if all the factors are also large; unfortunately, this is one of the few problems where we know how to find hard examples reliably. (This kind of encryption also requires there to be a 'trapdoor', known only to the legitimate recipient of the message, which allows the difficult problem to be by-passed. That, however, is another story.)

From helping a travelling salesman to filling a knapsack as full as possible, NP-complete problems turn up everywhere; if you run into a problem that seems hard, maybe there's a reason for it!

Websites

1. Clay Mathematics Institute: http://www.claymath.org/
2. Solving Travelling Salesman Problems: http://www.math.princeton.edu/tsp/
3. A compendium of NP optimization problems: http://www.nada.kth.se/~viggo/problemlist/

Peter Mattsson

[1] Factoring is not NP-complete, but it is still hard!

1. Are there three distinct digits such that the three numbers formed from them are all prime? What if the numbers are not all distinct?

2. How many solutions does the equation

$$\sqrt{x} + \sqrt{y} = 2003$$

have in integers x, y?

ABBAS ROOHOLAMINY
Iran

Letters to the Editor

Dear Editor,

Extracting a square root

This is an alternative to Newton's method for extracting a square root.

To find $r = \sqrt{x}$ we use the following recurrence relation (or iteration):

$$r_{n+1} = \frac{x - g^2}{r_n + g} + g\,,$$

where g is the initial guess and $r_0 = g$. For example, to find $\sqrt{20}$, a first guess is $g = 4 = r_0$. Then

$$\begin{aligned} r_1 &= \frac{20 - 16}{r_0 + 4} + 4 = 4.5\,,\\ r_2 &= 4.47058\ldots\,,\\ r_3 &= 4.47222\ldots\,,\\ r_4 &= 4.47213\ldots\,,\\ r_5 &= 4.47213\ldots\,, \end{aligned}$$

which is correct to 4 decimal places.

Yours sincerely,
BOB BERTUELLO
(12 Pinewood Road,
Midsomer Norton,
Bath BA3 2RG,
UK.)

Dear Editor,

From Pascal to Einstein

In the article 'From Pascal to Groups' by myself and David Sharpe, in Vol. 29, No. 3, of *Mathematical Spectrum*, an example of a group construction on the point of a conic was given. Specifically, given a conic C, a straight line l and a fixed point F of C, we can define the 'product' $\mathrm{P}\cdot\mathrm{Q}$ of points P, Q of C as follows: draw the straight line PQ to meet l at R, then draw RF to meet C again at $\mathrm{P}\cdot\mathrm{Q}$; this is illustrated in figure 1.

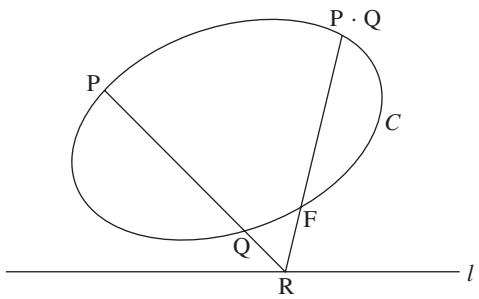


Figure 1.

This binary operation gives a group construction on the points of C, as shown in the article. The proof of the associative law uses Pascal's theorem for conics.

As an example, let C be the rectangular hyperbola with equation $xy = c^2$, let l be the line $y = x$ and let F be the point at infinity in the direction of the y-axis. With $\mathrm{P} = (a, c^2/a)$, $\mathrm{Q} = (b, c^2/b)$, a little calculation shows that the point $\mathrm{P}\cdot\mathrm{Q}$ has x-coordinate

$$\frac{a+b}{1+ab/c^2}.$$

We interpret the line through R and F to be the line through F parallel to the y-axis. This gives the group defined on $\mathbb{R}$ with binary operation given by

$$a * b = \frac{a+b}{1+ab/c^2}.$$

This is the rule for combining velocities in Einstein's theory of relativity, where c is the speed of light. If, however, l is the line $y = 0$, with C and F as before, we obtain the binary operation on $\mathbb{R}$ given by

$$a * b = a + b,$$

which is Newton's rule for combining velocities. Intriguing!

Yours sincerely,
GUIDO LASTERS
(Ganzendries 245,
3300 Tienen/Oplinter,
Belgium.)

Dear Editor,

Stirling numbers of the second kind

The otherwise-nice article 'Suppose Snow White Agreed to Take Part as Well' in *Mathematical Spectrum*, Volume 35, Number 2 has a small error in its concluding paragraph. The number of surjections from a set with n elements to a set with r elements is $r!\,{}_n\mathrm{T}_r$ not just ${}_n\mathrm{T}_r$ as stated in the article. The discrepancy even occurs for $n = r = 2$ since ${}_2\mathrm{T}_2 = 1$, while there are two surjections from a two-element set to a two-element set. The correct formula for the number of surjections, along with other results in the Snow White article, appears in the book *Discrete and Combinatorial Mathematics* by A. Hillman, G. Alexanderson and R. Grassl (Macmillan, London, 1987). This and other books on combinatorics call the numbers ${}_n\mathrm{T}_r$ 'Stirling numbers of the second kind'.

Yours sincerely,
PETER ROSS
(Department of Mathematics
and Computer Science,
Santa Clara University,
500 El Camino Real,
Santa Clara, CA 95053-0290,
USA.)

Problems and Solutions

Students are invited to submit solutions to some or all of the problems below. The most attractive solutions will be published in subsequent issues and are eligible for annual prizes. When writing to the Editorial Office, please state your full name and also the postal address of your school, college or university.

Problems

36.1 Let $A_1A_2 \ldots A_n$ be a regular n-sided polygon. Show that $(PA_1)^2+(PA_2)^2+\cdots+(PA_n)^2$ is the same for all points P on the circumscribed circle of the polygon.

(Submitted by J. Craig, Nottingham High School)

36.2 A 3×3 *magic square* is defined as a square array of natural numbers in which the sums of the elements in each row, each column and on each of the two diagonals are the same, s say. Express the element in the centre in terms of s and show how to obtain different 3×3 magic squares with a given central element. Define a 3×3 *multiplication magic square* by replacing addition by multiplication. What are the corresponding results for these?

(Submitted by Emanuel Emanouilidis, Kean University, New Jersey)

36.3 The positive real numbers $a_1, \ldots, a_n$ are such that $a_1+a_2+\cdots+a_n = n$. Show that $a_1^{3/2}+a_2^{3/2}+\cdots+a_n^{3/2} \geq n$.

(Submitted by Milton Chowdhury, Blackpool)

36.4 Let $s_n = n! - n^n/e^{n-1}$ for $n = 1, 2, 3, \ldots$. Prove that the sequence (s_n) is strictly increasing and unbounded.

(Submitted by Hassan Shah Ali, Tehran)

Solutions to Problems in Volume 35 Number 2

35.5 The real numbers $\alpha_1, \ldots, \alpha_n$ are such that

$$m\lfloor\alpha_1 + \cdots + \alpha_n\rfloor = \lfloor m\alpha_1\rfloor + \cdots + \lfloor m\alpha_n\rfloor$$

holds for infinitely many values of the integer m, where $\lfloor x\rfloor$ denotes the integer part of the real number x. Prove that $\alpha_1, \ldots, \alpha_n$ are rational.

Solution by Hassan Shah Ali, who proposed the problem

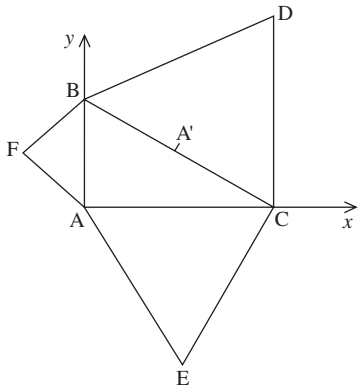
For each real number x, $x = \lfloor x\rfloor + \{x\}$, where $\{x\}$ denotes the fractional part of x, so that $0 \leq \{x\} < 1$. Hence

$$m\{\alpha_1 + \cdots + \alpha_n\} = \{m\alpha_1\} + \cdots + \{m\alpha_n\}$$

holds for infinitely many values of m. Suppose that $\{\alpha_1 + \cdots + \alpha_n\} > 0$. Then the integers m for which this holds must be positive. But $\{m\alpha_1\} + \cdots + \{m\alpha_n\} < n$ and it is impossible to have infinitely many positive integers m such that $m\{\alpha_1 + \cdots + \alpha_n\} < n$. Hence $\{\alpha_1 + \cdots + \alpha_n\} = 0$, so that $\{m\alpha_1\} = \cdots = \{m\alpha_n\} = 0$ for infinitely many values of m. Thus, $m\alpha_1, \ldots, m\alpha_n$ are integers and so $\alpha_1, \ldots, \alpha_n$ are rational.

35.6 The triangle ABC has angles $A = 90°$, $B = 60°$, $C = 30°$. Outward equilateral triangles BCD, CAE, ABF are attached to the triangle. Show that AD, BE and CF are concurrent and that their point P of intersection lies on the circle with diameter $A'B$, where A' is the midpoint of BC. If inward instead of outward equilateral triangles are attached to the triangle, which point corresponds to P? (These two points are called the *first* and *second Fermat points* of the triangle.)

Solution by Bor-Yann Chen, University of California, Irvine



Choose axes and scale so that $A = (0, 0)$, $B = (0, 1)$ and $C = (\sqrt{3}, 0)$. Then $D = (\sqrt{3}, 2)$, $E = (\sqrt{3}/2, -3/2)$, $F = (-\sqrt{3}/2, 1/2)$, so AD has equation $y = (2/\sqrt{3})x$, BE has equation $y = -(5/\sqrt{3})x + 1$ and CF has equation $y = -(1/3\sqrt{3})x + 1/3$. These lines intersect at the point $P(\sqrt{3}/7, 2/7)$. The circle with diameter BA' has equation $(x - \sqrt{3}/4)^2 + (y - 3/4)^2 = 1/4$, and P lies on this circle.

If inward equilateral triangles are attached to the sides of the triangle ABC, the point corresponding to P is B.

35.7 O_1, O_2 and A_0 are three distinct points in the plane. The sequence of points $A_1, A_2, A_3, \ldots$ is constructed by the following rule: given A_{2n}, the point A_{2n+1} lies on O_1A_{2n} produced such that A_{2n} is the midpoint of O_1A_{2n+1} and A_{2n+2} lies on O_2A_{2n+1} produced such that A_{2n+1} is the midpoint of O_2A_{2n+2}. Show that $A_0, A_2, A_4, \ldots$ are collinear, that $A_1, A_3, A_5, \ldots$ are collinear, and that these straight lines are parallel.

Solution by Guido Lasters, who proposed the problem

Let $O_1 = (0, 0)$, $O_2 = (1, 0)$ and $A_0 = (a, b)$. Then $A_1 = (2a, 2b)$, $A_2 = (2a, 2b)+(2a-1, 2b) = (4a-1, 4b)$

and $A_3 = (8a - 2, 8b)$. The straight line A_0A_2, ℓ_1 say, has equation

$$\frac{y-b}{4b-b} = \frac{x-a}{4a-1-a},$$

that is,

$$\frac{y-b}{3b} = \frac{x-a}{3a-1}.$$

The straight line A_1A_3, ℓ_2 say, has equation

$$\frac{y-2b}{8b-2b} = \frac{x-2a}{8a-2-2a},$$

that is,

$$\frac{y-2b}{3b} = \frac{x-2a}{3a-1}.$$

If A_{2n} has coordinates (r, s), then A_{2n+2} has coordinates $(4r - 1, 4s)$. If A_{2n} lies on ℓ_1, then

$$\frac{s-b}{3b} = \frac{r-a}{3a-1},$$

from which

$$\frac{4s-4b}{3b} = \frac{4r-4a}{3a-1}$$

and

$$\frac{4s-b}{3b} - \frac{3b}{3b} = \frac{(4r-1)-a}{3a-1} - \frac{3a-1}{3a-1},$$

so that

$$\frac{4s-b}{3b} = \frac{(4r-1)-a}{3a-1}$$

and A_{2n+2} also lies on ℓ_1. If A_{2n+1} has coordinates (u, v), then A_{2n+3} has coordinates $(4u - 2, 4v)$. If A_{2n+1} lies on ℓ_2, then

$$\frac{v-2b}{3b} = \frac{u-2a}{3a-1},$$

from which

$$\frac{4v-8b}{3b} = \frac{4u-8a}{3a-1},$$

$$\frac{4v-2b}{3b} - \frac{6b}{3b} = \frac{(4u-2)-2a}{3a-1} - \frac{6a-2}{3a-1},$$

$$\frac{4v-2b}{3b} = \frac{(4u-2)-2a}{3a-1}$$

and A_{2n+3} also lies on ℓ_2. Both ℓ_1 and ℓ_2 have slope $3b/(3a - 1)$, and so are parallel. (If $b = 0$, then ℓ_1, ℓ_2 are both the line $y = 0$; if $a = \frac{1}{3}$, then they are $x = \frac{1}{3}$ and $x = \frac{2}{3}$ respectively. If $a = \frac{1}{3}$ and $b = 0$, then $A_0 = A_2 = A_4 = \cdots = (\frac{1}{3}, 0)$ and $A_1 = A_3 = A_5 = \cdots = (\frac{2}{3}, 0)$.)

35.8 Show that the set of natural numbers is the union of two disjoint subsets neither of which contains an infinite arithmetic progression.

Solution by Farshid Arjomandi, who proposed the problem

Consider the two disjoint subsets

$$A = \{1\} \cup \{4, 5, 6\} \cup \{11, 12, 13, 14, 15\} \cup \cdots,$$
$$B = \{2, 3\} \cup \{7, 8, 9, 10\} \cup \{16, 17, 18, 19, 20, 21\} \cup \cdots.$$

Consider any infinite arithmetic progression with common difference d say. Among any d consecutive natural numbers, one of them must belong to this arithmetic progression. But both A and B contain d consecutive natural numbers, so must contain a member of this arithmetic progression, so it does not lie wholly in A nor does it lie wholly in B.

J. Craig (Nottingham High School) considered the disjoint subsets

$$A = \{1\} \cup \{4, 6\} \cup \{7, 9, 11\} \cup \{14, 16, 18, 20\} \cup \cdots,$$
$$B = \{2\} \cup \{3, 5\} \cup \{8, 10, 12\} \cup \{13, 15, 17, 19\} \cup \cdots.$$

Milton Chowdhury (Blackpool) took $A = \{a_k\}_{k\in\mathbb{N}}$ where $a_k = k! + k$, and $B = \mathbb{N} \setminus A$. Consider the arithmetic progression $\{a + nd\}_{n\in\mathbb{N}}$. Since

$$\big((d+1)a\big)! + (d+1)a = a + nd$$

for some n, it has a term in A. Suppose the arithmetic progression lies wholly in A, say

$$a_{k_1}, a_{k_2}, \ldots.$$

Then, for all i,

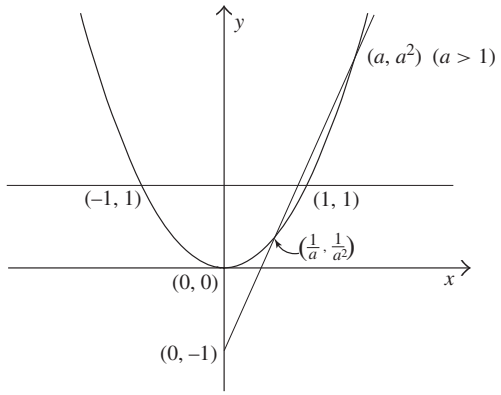
$$\begin{aligned} d = a_{k_{i+1}} - a_{k_i} &\geq a_{k_i+1} - a_{k_i} \\ &= (k_i + 1)! + (k_i + 1) - (k_i! + k_i) \\ &= k_i!\, k_i + 1. \end{aligned}$$

But $k_i \to \infty$ as $i \to \infty$, so this is impossible. Milton Chowdhury points out that, if

$$C = \sum_{n=1}^{\infty} \frac{1}{2^{n!+n}},$$

then C and $1 - C$ are both irrational because they have non-repeating binary expansions. Is C transcendental?

A property of the parabola $y = x^2$ at a glance



GUIDO LASTERS
Tienen, Belgium

Reviews

The William Lowell Putnam Mathematical Competition 1985–2000: Problems, Solutions and Commentary. By KIRAN S. KEDALYA, BJORN POONEN AND RAVI VAKIL. MAA, Washington, DC, 2002. Pp. 354. Hardback $44.95 (ISBN 0-88385-807-X).

This book contains problems from the Putnam competition for the years 1985–2000. This is an annual competition, held in the United States and Canada for university students who have not yet gained their degrees. In addition to the problems, hints for each problem and (sometimes multiple) solutions are given. For interested readers, there is a lot of background material about the competition, such as the names of the winners. Furthermore, the solutions are placed in the context of more advanced concepts, so that, for example, the solution of one problem was linked to the generalised Riemann hypothesis. Thus, it illustrates that whilst such problems do not necessarily require great mathematical knowledge, they can still be linked with some of the hardest problems in mathematics.

The problems are clearly stated and the solutions are presented in a straightforward fashion, making the book eminently readable. Although the mathematics required is not of an extremely high standard, more than is taught in any A-level syllabus is needed, so that whilst some of the easier problems might be interesting and relevant to interested A-level students, they would find those requiring more mathematical knowledge to be somewhat dispiriting. However, for university students or teachers wishing to extend their problem-solving skills, I would definitely recommend the book, as it provides a set of over 150 beautiful problems with various interesting solutions, which can only improve your problem-solving ability.

Student, Berkhamsted Collegiate School PAUL JEFFERYS

Mathematical Treks. By IVARS PETERSON. MAA, Washington, DC, 2002. Pp. 170. Paperback $24.95 (ISBN 0-88385-537-2).

Ivars Peterson is one of the foremost current popularisers of mathematics and his name may be familiar to readers through books such as *Islands of Truth*, *The Mathematical Tourist* and *Jungles of Randomness*. *Mathematical Treks* consists of 33 short miscellaneous articles that are updated versions of *Science News* online columns that appeared in 1996–1997. The weekly 'MathTrek' column is still running and may be accessed from the MAA's website, http://www.maa.org/. Peterson's style is lucid, lively and very readable; each article is well researched with follow-up references and he has a magpie-like eye for a promising theme or interesting entrée. These glimpses of 'mathematics in action' range from recreational mathematics (Conway's game of sprouts, matchstick maths, Möbius strips and recycling logos, river-crossing problems) to recent research (DNA computers, cake division algorithms, circle packing problems, prime number theorems and records) via historical (Erdős, EDSAC's golden jubilee) and humorous pieces (I won't disclose which is the April fool spoof). Readers may enjoy playing with the following problems:

(p. 52.) In a group of N gossipers (N even), each is able to pass on one piece of gossip per day. Show that a piece of news can percolate to everyone in the group within D days, where D is the smallest integer greater than $\log_2 N$.

(p. 135: Thabit ibn Qurra's 9th century theorem.) If $p = 3 \cdot 2^n - 1$, $q = 3 \cdot 2^{n-1} - 1$ and $r = 9 \cdot 2^{2n-1} - 1$ are all prime, then $2^n pq$ and $2^n r$ are amicable numbers (i.e. each is the sum of the proper factors of the other).

I hope that this review conveys the flavour of this enjoyable book which certainly merits a wide readership and a place in sixth-form libraries.

Tonbridge School, Kent NICK LORD

Solve This. By JAMES TANTON. MAA, Washington, DC, 2002. Pp. 240. Paperback $29.95 (ISBN 0-88385-717-0).

What an excellent book! The author is an American college teacher and no doubt many of the sources he quotes are more readily obtainable in the USA. However, this will not detract from its value. His enthusiasm, top rate mathematical explanations and variety of subject matter overwhelm.

Enough material is given for 30 Maths Club sessions of length 90 minutes, mainly suitable for older secondary school children or university students. The material has clearly been trialled as witnessed by the pictures, which enhance the value of the book further. Much of the material is practically based and so ostensibly not too heavy mathematically, and yet the maths is there, and is deep.

The book is in three parts. Firstly the problems for the sessions are given; then, in the second section, hints and some of the answers are provided. Proofs are given in the final section along with the rest of the answers. In all three sections further challenges are given so when the whole book has been digested (which is unlikely to happen!) there is still more to get your teeth into. The author admits that he and his students have not found an answer to some problems. The explanations are clear and the author does not hesitate to give hard proofs for those who want them.

Most sections have about three or four different problems, which are linked in surprising ways. Readers will probably recognise some of the problems, but no doubt will find much that is new and interesting. Fibonacci sequences, Steiner points, map colouring, tiling and probability problems are just some of the topics included. There are ideas that could be used for party tricks (for example, how could you take a T-shirt off someone and put it back inside out whilst all the time his arms are folded?). Here is a sample of problem headings: which way did the bicycle go? a torus with a serious twist (there's a lot about toruses); yo-yo quirk; the money or the goat? path walking; slicing a bagel (an

American toroidal bun by the way); mutilated laundry; skew tetrominoes; congruent halves; on perfect shuffling.

The book will stimulate mathematical thinking and I am sure will prove to be great fun too.

Stamford, Lincolnshire ALASTAIR SUMMERS

Towing Icebergs, Falling Dominoes, and Other Adventures in Applied Mathematics. By ROBERT B. BANKS. Princeton University Press, 2002. Pp. 339. Paperback £11.95 (ISBN 0-691-10285-6).

Slicing Pizzas, Racing Turtles, and Further Adventures in Applied Mathematics. By ROBERT B. BANKS. Princeton University Press, 2002. Pp. 304. Paperback £11.95 (ISBN 0-691-10284-8).

The first of this pair of books takes its readers on a magical mystery tour of some exciting areas of applied mathematics. It avoids techniques beyond A-level standard and is quite readable, due to Banks' relaxed style. A good range of topics are covered, which are often unusual and intriguing.

When reading a book like this, I always experience a burst of pleasure when the author shows me a mathematical simulation that strikes a chord — when it describes a situation which I want to understand more deeply for one reason or another. Most of the book covers simulations. Personally (though perhaps embarrassingly) I found the discussion of density waves in traffic flows fascinating; Banks shows how to calculate the speed of the 'shock wave' created by cars slamming on their brakes to avoid an accident on a motorway, or to catch a glimpse of something curious on the roadside. I think most people will find something that catches their imagination here, whether it be economics, dynamics, probability, combinatorics or modelling. But a list like that makes the whole affair sound too dull; two chapters, for instance, are devoted to a feasibility study of a proposal to tow icebergs from Antarctica to California. This is made to sound plausible (amazingly) right down to planning which US navy ships would need to be procured. We also have a delightful mathematical 'explanation' for the shape of the Eiffel Tower, and estimates of the strength of the most powerful volcanoes in history.

There are some problems with the book, however. It can be a little repetitive; the same example is sometimes used twice in different contexts. Although the book does make itself relevant by dealing with specific cases, it is sometimes a little too specific; for example when giving long numerical calculations. One small problem for the UK audience is the use of units; be prepared to deal in feet, pounds and slugs (yes, honestly, 'slugs'). There are some errors, mostly typographical, but some mathematical. I feel I should point out the mistake in formula (10.17), to the effect that $(\mathrm{d}/\mathrm{d}n)2^n = n \cdot 2^{n-1}$.

One of the strengths of this book, especially for sixth-form readers looking for something a bit challenging, is that it does not shy away from more difficult maths. Banks grapples with partial differential equations, and some interesting fluid dynamics (while addressing the lift force on a baseball and the drag force on an iceberg); the general solution of the wave equation and the Kutta–Joukowski hypothesis are mentioned. So there is certainly something to stretch sixth-formers, and probably something to open the eyes of most undergraduates too. There are plenty of references to books and research papers for further study. All in all, *Towing Icebergs, Falling Dominoes, and Other Adventures in Applied Mathematics* is a book well-crafted to its audience.

The sequel, *Slicing Pizzas, Racing Turtles, and Further Adventures in Applied Mathematics*, is aimed at anyone who has an interest in seeing maths at work, in an often quirky way, and in a wild range of disciplines.

There are twenty-six chapters, all quite short, and each covering a particular area. Most of them walk the reader through a problem. For instance, if you want to stay dry, is it best to run or to walk when it rains? Is the answer different in a shower or in a thunderstorm? This showcases the way in which mathematicians investigate things and gain insights into real-world problems. Banks has useful things to say about the process of simplifying a problem. He also makes points about analogies; connecting population curves and cartography, for instance. There are many insights here, and the examples treated always suggest lots of further investigations.

Having said all this, you might want to check how many of the topics covered are genuinely new to you. As a third-year undergraduate, I had come across almost all the topics elsewhere in the recreational maths literature at some point. Some of them are quite wild flights of fancy, such as 'What happens if we dump all the land into the ocean?', which may seem a touch bizarre and (dare I say it) academic, although this is a matter of personal taste. There is an American bias, so that calculations concerning 'Hershey Chocolate Kisses' may not seem quite as relevant as they should. I did have a few specific criticisms: the notation in the 'slicing pizzas' chapter is confusing, and the author sometimes stretches A-level standard maths too far in deriving results that he should probably just state (again, in regard to 'slicing pizzas', and in the map projections chapter).

The approach is fresh, and the language informal. Many references are given, some of them to academic articles. I would advise sixth-formers dipping into this book not to be discouraged if they find some parts difficult to follow; filling in the details will require some back-of-envelope scribbling. There are also some 'high-tech' uses of geometry and differential equations, which may be challenging.

The more 'mature' mathematician should find interest in this book, although purists may be dismayed by the explicit numerical computations. On the other hand, there are some genuine pure maths gems; formulae for π, and lovely golden section results. And surely no-one can fail to be touched by the fusion of cardioids and lemniscates into valentine hearts!

Student, Queens' College, Cambridge WILL DONOVAN

Other books received

Mathematical Analysis. Edited by ALLADI SITARAM AND VISHWAMBHAR PATI. Universities Press, Hyderabad, 2001. Pp. 142. Paperback £15.95 (ISBN 81-7371-291-3).

Mathematical Spectrum

2003/2004 Volume 36 Number 1

ISSN 0025-5653

Published by the Applied Probability Trust
Printed by Pear Tree Press Ltd, Stevenage, Herts, UK