*

ISSN 1350-8539

M500 209

A1 A14 A27 B0	A2 A15 A28 B0	A3 A16 A29 B0	A4 A17 A30 B0	A5 A18 A31 B0	A6 A19 A32 B0
A7 A20 A33 B0	A8 A21 A34 B0	A9 A22 A35 B0	A11 A24 A37 B0		
A0 A2 A14 B10	A13 A15 A27 B13	A26 A28 A1 B15	A0 A3 A10 B8	A13 A16 A23 B18	A26 A29 A36 B12
A0 A4 A7 B6	A13 A17 A20 B4	A26 A30 A33 B9	A0 A5 A27 B16	A13 A18 A1 B17	A26 A31 A14 B5
A0 A6 A9 B7	A13 A19 A22 B11	A26 A32 A35 B1	A0 A8 A11 B15	A13 A21 A24 B10	A26 A34 A37 B13
A0 A12 A21 B12	A13 A25 A34 B8	A26 A38 A8 B18	A0 A13 A37 B2	A13 A26 A11 B14	A26 A0 A24 B3
A0 A15 A30 B18	A13 A28 A4 B12	A26 A2 A17 B8	A0 A18 A28 B5	A13 A31 A2 B16	A26 A5 A15 B17
A0 A19 A20 B17	A13 A32 A33 B5	A26 A6 A7 B16	A0 A22 A34 B9	A13 A35 A8 B6	A26 A9 A21 B4
A0 A29 A38 B13	A13 A3 A12 B15	A26 A16 A25 B10	A0 A31 A35 B11	A13 A5 A9 B1	A26 A18 A22 B7
A0 A33 A36 B14	A13 A7 A10 B3	A26 A20 A23 B2	A1 A3 A15 B12	A14 A16 A28 B8	A27 A29 A2 B18
A1 A4 A21 B9	A14 A17 A34 B6	A27 A30 A8 B4	A1 A5 A32 B3	A14 A18 A6 B2	A27 A31 A19 B14
A1 A6 A11 B8	A14 A19 A24 B18	A27 A32 A37 B12	A1 A7 A22 B18	A14 A20 A35 B12	A27 A33 A9 B8
A1 A8 A20 B11	A14 A21 A33 B1	A27 A34 A7 B7	A1 A9 A23 B16	A14 A22 A36 B17	A27 A35 A10 B5
A1 A10 A31 B13	A14 A23 A5 B15	A27 A36 A18 B10	A1 A12 A34 B14	A14 A25 A8 B3	A27 A38 A21 B2
A1 A16 A35 B7	A14 A29 A9 B11	A27 A3 A22 B1	A1 A19 A37 B4	A14 A32 A11 B9	A27 A6 A24 B6
A1 A29 A33 B10	A14 A3 A7 B13	A27 A16 A20 B15	A1 A30 A38 B6	A14 A4 A12 B4	A27 A17 A25 B9
A2 A3 A20 B14	A15 A16 A33 B3	A28 A29 A7 B2	A2 A4 A11 B13	A15 A17 A24 B15	A28 A30 A37 B10
A2 A5 A37 B7	A15 A18 A11 B11	A28 A31 A24 B1	A2 A6 A38 B5	A15 A19 A12 B16	A28 A32 A25 B17
A2 A7 A9 B17	A15 A20 A22 B5	A28 A33 A35 B16	A2 A8 A19 B12	A15 A21 A32 B8	A28 A34 A6 B18
A2 A10 A23 B9	A15 A23 A36 B6	A28 A36 A10 B4	A2 A12 A36 B1	A15 A25 A10 B7	A28 A38 A23 B11
A2 A24 A32 B11	A15 A37 A6 B1	A28 A11 A19 B7	A2 A30 A35 B15	A15 A4 A9 B10	A28 A17 A22 B13
A2 A33 A34 B4	A15 A7 A8 B9	A28 A20 A21 B6	A3 A4 A32 B18	A16 A17 A6 B12	A29 A30 A19 B8
A3 A6 A35 B4	A16 A19 A9 B9	A29 A32 A22 B6	A3 A8 A36 B7	A16 A21 A10 B11	A29 A34 A23 B1
A3 A9 A24 B2	A16 A22 A37 B14	A29 A35 A11 B3	A3 A11 A31 B10	A16 A24 A5 B13	A29 A37 A18 B15
A3 A17 A18 B16	A16 A30 A31 B17	A29 A4 A5 B5	A3 A23 A30 B3	A16 A36 A4 B2	A29 A10 A17 B14
A3 A25 A33 B11	A16 A38 A7 B1	A29 A12 A20 B7	A3 A34 A38 B17	A16 A8 A12 B5	A29 A21 A25 B16
A4 A6 A10 B17	A17 A19 A23 B5	A30 A32 A36 B16	A4 A8 A22 B16	A17 A21 A35 B17	A30 A34 A9 B5
A4 A19 A25 B1	A17 A32 A38 B7	A30 A6 A12 B11	A4 A23 A35 B14	A17 A36 A9 B3	A30 A10 A22 B2
A4 A31 A34 B15	A17 A5 A8 B10	A30 A18 A21 B13	A4 A37 A38 B3	A17 A11 A12 B2	A30 A24 A25 B14
A5 A6 A21 B14	A18 A19 A34 B3	A31 A32 A8 B2	A5 A10 A20 B18	A18 A23 A33 B12	A31 A36 A7 B8
A5 A11 A22 B4	A18 A24 A35 B9	A31 A37 A9 B6	A5 A12 A25 B6	A18 A25 A38 B4	A31 A38 A12 B9
A5 A19 A33 B2	A18 A32 A7 B14	A31 A6 A20 B3	A5 A34 A36 B11	A18 A8 A10 B1	A31 A21 A23 B7
A6 A25 A36 B15	A19 A38 A10 B10	A32 A12 A23 B13	A7 A11 A25 B5	A20 A24 A38 B16	A33 A37 A12 B17
A7 A12 A35 B10	A20 A25 A9 B13	A33 A38 A22 B15	A7 A23 A24 B4	A20 A36 A37 B9	A33 A10 A11 B6
A8 A23 A37 B8	A21 A36 A11 B18	A34 A10 A24 B12	A9 A11 A38 B12	A22 A24 A12 B8	A35 A37 A25 B18
A26 A27 A4 C0	A11 A20 A30 C0	A1 A2 A25 C0	A15 A34 A35 C0	A0 A16 A32 C0	A28 A3 A5 C0
A17 A31 A33 C0	A18 A9 A12 C0	A10 A14 A37 C0	A24 A29 A8 C0	A19 A21 A7 C0	A6 A22 A23 C0
A13 A36 A38 C0					
A0 A1 A17 C1	A24 A33 A4 C1	A14 A15 A38 C1	A28 A8 A9 C1	A13 A29 A6 C1	A2 A16 A18 C1
A30 A5 A7 C1	A31 A22 A25 C1	A23 A27 A11 C1	A37 A3 A21 C1	A32 A34 A20 C1	A19 A35 A36 C1
A10 A12 A26 C1					
A13 A14 A30 C2	A37 A7 A17 C2	A27 A28 A12 C2	A2 A21 A22 C2	A26 A3 A19 C2	A15 A29 A31 C2
A4 A18 A20 C2	A5 A35 A38 C2	A36 A1 A24 C2	A11 A16 A34 C2	A6 A8 A33 C2	A32 A9 A10 C2
A0 A23 A25 C2					C0 C1 C2 B0
B0 B1 B4 A0	B0 B7 B9 A13	B0 B11 B6 A26	B1 B2 B5 A1	B1 B8 B10 A20	B1 B12 B7 A30
B2 B3 B6 A2	B2 B13 B8 A35	B3 B10 B12 A22	B3 B14 B9 A28	B4 B15 B10 A32	B5 B6 B9 A3
B5 B16 B11 A37	B6 B13 B15 A19	B7 B8 B11 A4	B7 B14 B16 A14	B7 B18 B13 A33	B8 B9 B12 A5
B8 B0 B14 A38	B9 B10 B13 A6	B10 B17 B0 A23	B10 B2 B16 A34	B11 B12 B15 A7	B11 B18 B1 A17
B11 B3 B17 A27	B12 B0 B2 A25	B12 B4 B18 A31	B13 B14 B17 A8	B13 B5 B0 A36	B14 B15 B18 A9
B14 B2 B4 A15	B15 B16 B0 A10	B15 B3 B5 A21	B16 B17 B1 A11	B16 B4 B6 A16	B17 B5 B7 A24
B17 B9 B4 A29	B18 B0 B3 A12	B18 B6 B8 A18			
B2 B9 B11 C0	B3 B4 B7 C0	B8 B15 B17 C0	B12 B13 B16 C0	B14 B6 B1 C0	B18 B10 B5 C0
B4 B5 B8 C1	B6 B17 B12 C1	B9 B16 B18 C1	B10 B11 B14 C1	B13 B1 B3 C1	B15 B7 B2 C1
B4 B11 B13 C2	B5 B12 B14 C2	B6 B7 B10 C2	B9 B1 B15 C2	B16 B8 B3 C2	B17 B18 B2 C2

The M500 Society and Officers

The M500 Society is a mathematical society for students, staff and friends of the Open University. By publishing M500 and 'MOUTHS', and by organizing residential weekends, the Society aims to promote a better understanding of mathematics, its applications and its teaching. Web address: www.m500.org.uk.

The magazine M500 is published by the M500 Society six times a year. It provides a forum for its readers' mathematical interests. Neither the editors nor the Open University necessarily agree with the contents.

MOUTHS is 'Mathematics Open University Telephone Help Scheme', a directory of M500 members who are willing to provide mathematical assistance to other members.

The September Weekend is a residential Friday to Sunday event held each September for revision and exam preparation. Details available from March onwards. Send s.a.e. to Jeremy Humphries, below.

The Winter Weekend is a residential Friday to Sunday event held each January for mathematical recreation. For details, send a stamped, addressed envelope to Diana Maxwell, below.

Advice to authors. We welcome contributions to M500 on virtually anything related to mathematics and at any level from trivia to serious research. Please send material for publication to Tony Forbes, above. We prefer an informal style and we usually edit articles for clarity and mathematical presentation. If you use a computer, please also send the file on a PC diskette or via e-mail.

Ceva's theorem to the rescue

Rob Evans

This article is intended to accompany my previous contribution, which was entitled 'Long live Geometry: 30° revisited', and which appeared in M500 **202**. In that article the existence of (±) Brocard points was taken on trust. The task of *proving* their existence was deliberately left for another time. That article has now been written—you are reading it!

Interestingly, despite the fact that on the Web one can find a considerable amount of fascinating information about Brocard points there seems to be nothing there that directly addresses the question of their existence. So, in search of help, I made a special trip to Southampton University library. But, despite coming across a 1916 edition of J. L. Coolidge's *A Treatise on the Geometry of the Circle and Sphere*, which has a whole section devoted to Brocardian geometry, I still found nothing! However, within this section of Coolidge's book there were some interesting equations which turn out to be relevant to the question at hand. Specifically, those equations turn out to be relevant to an approach based upon the use of Ceva's theorem. Such an approach is taken in this article.

Before restating the relevant definitions from the previous article, I need to say something about the notation we shall use.

To avoid unnecessary repetition and possible ambiguities in our proofs we shall adopt the following convention. The symbol string 'ΔABC' denotes an arbitrary fixed non-degenerate triangle with vertices A, B, C which—on traversal of its sides exactly once in an anti-clockwise sense starting at some point between C and A—appear in the order A, B, C. (A triangle is referred to as 'non-degenerate' if and only if its vertices are not collinear.) With this interpretation of 'ΔABC' we restate the relevant definitions from the previous article as follows.

The point Ω^+ is a positive Brocard point of ΔABC if and only if Ω^+ lies inside ΔABC and $\angle CB\Omega^+ = \angle AC\Omega^+ = \angle BA\Omega^+$; see Figure 1.

The point Ω^- is a negative Brocard point of ΔABC if and only if Ω^- lies inside ΔABC and $\angle BC\Omega^- = \angle CA\Omega^- = \angle AB\Omega^-$; see Figure 2.

Before stating Ceva's theorem and the main theorem that we shall prove, I need to say something more about the notation we shall use.

We define $[B, C \rightarrow)$ to be the closed half-line whose (included) endpoint is B and which passes through C.

Let P be an arbitrary point on the line which passes through B and

C. Then, for $P \neq C$ we have that BP is the real number whose value is determined according to the following rule:

$$BP \;=\; +|BP| \text{ if } P \in [B, C \rightarrow); \qquad BP \;=\; -|BP| \text{ if } P \in [B, C \rightarrow).$$

In the exceptional case where $P = C$ we have the following dual usage: BC is either the real number whose value is determined according to the above rule or the line which passes through B and C. However, in a given context, which of these two things BC is shall be obvious.

As usual, (B, C) is the open line segment whose (excluded) endpoints are B and C.

Finally, everything said about notation in the previous seven sentences remains true under each permutation of the set of symbols $\{A, B, C\}$.

The wording of Ceva's theorem that we shall use is as follows.

Ceva's theorem. Let L, M, N be arbitrary fixed points on the lines BC, CA, AB respectively. Then, the necessary and sufficient condition that the three lines AL, BM, CN be concurrent at one, and only one, point is that $(BL)(CM)(AN) = (CL)(AM)(BN)$; see Figure 3.

The wording of the main theorem that we shall prove is as follows.

The main theorem. Let L^+, M^+, N^+ be fixed points on the lines BC, CA, AB respectively such that they are defined implicitly by the following equality between ordered triples (of positive real numbers):

$$(a^2 BL^+, b^2 CM^+, c^2 AN^+) \;=\; (c^2 CL^+, a^2 AM^+, b^2 BN^+). \qquad \text{(MT)}$$

Then the three lines AL^+, BM^+, CN^+ are concurrent at one, and only one, point. Moreover, this point of concurrence lies *inside* ΔABC and is a positive Brocard point of ΔABC; see Figure 4.

With regards to the wording of the main theorem note the following.

(1) The letters a, b, c denote $|BC|$, $|CA|$, $|AB|$ respectively.

(2) The treatment of negative Brocard points would entail the implicit definition of fixed points L^-, M^-, N^- by the following equality between ordered triples (of positive real numbers):

$$(a^2 BL^-, b^2 CM^-, c^2 AN^-) \;=\; (b^2 CL^-, c^2 AM^-, a^2 BN^-).$$

This treatment of negative Brocard points is completely analogous to our treatment of positive Brocard points. Consequently, we have restricted our attention to the latter.

Everything that was said about notation before the statements of Ceva's theorem and the main theorem applies in the proof of the main theorem also.

Proof of the main theorem. From the respective definitions of L^+, M^+, N^+ we have

$$(a^2BL^+, b^2CM^+, c^2AN^+) \;=\; (c^2CL^+, a^2AM^+, b^2BN^+)$$

(equation (MT)). Consequently, we must have

$$\begin{aligned} a^2BL^+b^2CM^+c^2AN^+ \;&=\; c^2CL^+a^2AM^+b^2BN^+ \\ \Rightarrow\quad (BL+)(CM+)(AN+) &= (CL+)(AM+)(BN+). \end{aligned}$$

Hence, from Ceva's theorem we must have that the three lines AL^+, BM^+, CN^+ are concurrent at one, and only one, point. Moreover, from equation (MT) and the obvious fact that $a^2, b^2, c^2 > 0$ it is clear that

L^+, M^+, N^+ lie on the open line segments (B, C), (C, A), (A, B) respectively. (OLs)

Consequently, the point at which the three lines AL^+, BM^+, CN^+ concur lies inside ΔABC.

Hence, in order to show that this point of concurrence is a positive Brocard point of ΔABC it only remains to demonstrate that it possesses the angular property that characterizes such a point. In other words, where P^+ is the point at which the three lines AL^+, BM^+, CN^+ concur, it only remains to show that $\angle ACP^+ = \angle BAP^+ = \angle CBP^+$. However, we know that the cotangent function is 1–1 on the open interval $(0, \pi)$. Consequently, in order to show that $\angle CBP^+ = \angle ACP^+ = \angle BAP^+$ it is sufficient to demonstrate that $\cot(\angle CBP^+) = \cot(\angle ACP^+) = \cot(\angle BAP^+)$. This we do as follows. (cot)

Let F be the foot of the perpendicular from M^+ on to the line BC. Then, from inspection of Figure 5 it is clear that we have

$$\cot(\angle CBP^+) \;=\; \frac{BF}{|FM^+|} \;=\; \frac{a - CF}{|FM^+|}.$$

In turn, from inspection of the same figure it is clear that we have

$$CF \;=\; |CM^+|\cos(\angle BCA); \qquad |FM^+| \;=\; |CM^+|\sin(\angle BCA).$$

Consequently, from the last three equations we must have

$$\cot(\angle CBP^+) \;=\; \frac{a - |CM^+|\cos(\angle BCA)}{|CM^+|\sin(\angle BCA)}. \qquad \text{(cot1)}$$

However, from statement (OLs) it is clear that we have

$$|CM^+| \;=\; CM^+ \qquad \text{and} \qquad |AM^+| \;=\; AM^+,$$

where, moreover, $|CM^+| + |AM^+| = b$. In turn, from equation (MT) we have that $b^2 CM^+ = a^2 AM^+$.

From these four equations and some algebraic manipulation we obtain

$$|CM^+| \;=\; \frac{a^2 b}{a^2 + b^2}.$$

Consequently, from equation (cot1) and some more algebraic manipulation we obtain

$$\cot(\angle CBP^+) \;=\; \frac{(a^2 + b^2) - ab\cos(\angle BCA)}{ab\sin(\angle BCA)}. \qquad \text{(cot2)}$$

However, from the law of cosines we know that

$$c^2 \;=\; a^2 + b^2 - 2ab\cos(\angle BCA).$$

Also from elementary trigonometry we know that

$$\Delta = \frac{1}{2}\, ab\sin(\angle BCA),$$

where Δ is the area of ΔABC. Consequently, from these two equations, equation (cot2) and still more algebraic manipulation we obtain

$$\cot(\angle CBP^+) \;=\; \frac{a^2 + b^2 + c^2}{4\Delta}.$$

However, from inspection of the proof since the end of paragraph (cot) it is clear that the truth of the last equation is invariant under each of the permutations $(ABC)(abc)$ and $[(ABC)(abc)]^2$, where $(ABC)(abc)$ is the permutation of the set of symbols A, B, C, a, b, c which maps A, B, C, a, b, c to B, C, A, b, c, a respectively. Consequently, from these two sentences we have

$$\begin{aligned} \cot(\angle CBP^+) \;&=\; \frac{a^2 + b^2 + c^2}{4\Delta}, \\ \cot(\angle ACP^+) \;&=\; \frac{b^2 + c^2 + a^2}{4\Delta}, \\ \cot(\angle BAP^+) \;&=\; \frac{c^2 + a^2 + b^2}{4\Delta}. \end{aligned}$$

However, it is obvious that the right-hand sides of these three equations are equal to each other. Consequently,

$$\cot(\angle CBP^+) \;=\; \cot(\angle ACP^+) \;=\; \cot(\angle BAP^+),$$

Q.E.D.

Hence, from the argument put forward in paragraph (cot), we must have that the point at which the three lines AL^+, BM^+, CN^+ concur (i.e. P^+) is a positive Brocard point of ΔABC. □

For the sake of completeness, we shall now prove the following corollary of the main theorem.

The four cotangents corollary. Let $p^+ = \angle CBP^+ = \angle ACP^+ = \angle BAP^+$. Then

$$\cot(p^+) = \cot(\angle BCA) + \cot(\angle CAB) + \cot(\angle ABC).$$

From the proof of the main theorem we know that P^+ is a positive Brocard point of ΔABC. Consequently, we know that p^+ is properly defined.

Proof. From the law of cosines we know $c^2 = a^2 + b^2 - 2ab\cos(\angle BCA)$. Also from elementary trigonometry we know $\Delta = \frac{1}{2}ab\sin(\angle BCA)$, where Δ is the area of ΔABC. Consequently, from the last two equations and some algebraic manipulation we obtain

$$\cot(\angle BCA) \;=\; \frac{a^2+b^2-c^2}{4\Delta}.$$

However, from inspection of this proof it is clear that the last equation remains true under each of the permutations $(ABC)(abc)$ and $[(ABC)(abc)]^2$. Consequently,

$$\begin{aligned}
\cot(\angle BCA) &= \frac{a^2+b^2-c^2}{4\Delta},\\
\cot(\angle CAB) &= \frac{b^2+c^2-a^2}{4\Delta},\\
\cot(\angle ABC) &= \frac{c^2+a^2-b^2}{4\Delta}.
\end{aligned}$$

From these three equations we have

$$\cot(\angle BCA) + \cot(\angle CAB) + \cot(\angle ABC) \;=\; \frac{a^2+b^2+c^2}{4\Delta}.$$

However, in the proof of the main theorem we had

$$\cot(p^+) \;=\; \frac{a^2+b^2+c^2}{4\Delta}.$$

Consequently, from these two equations we have

$$\cot(p^+) \;=\; \cot(\angle BCA)+\cot(\angle CAB)+\cot(\angle ABC). \qquad \square$$

Next, we shall consider the question of the uniqueness of Brocard points. It is (as readers can confirm) a straightforward matter to prove, via the method of *reductio ad absurdum*, the following theorem.

The uniqueness theorem. Let Ω^+ be a positive Brocard point of ΔABC. Then, Ω^+ is *the* positive Brocard point of ΔABC.

In light of the uniqueness theorem, we obtain from the four cotangents corollary the following theorem.

The four cotangents theorem (+). Let Ω^+ be the positive Brocard point of ΔABC. In turn, let $\omega^+ = \angle CB\Omega^+ = \angle AC\Omega^+ = \angle BA\Omega^+$. Then,

$$\cot(\omega^+) = \cot(\angle BCA)+\cot(\angle CAB)+\cot(\angle ABC).$$

A similar treatment of negative Brocard points would yield the following analogous theorem.

The four cotangents theorem (−). Let Ω^- be the negative Brocard point of ΔABC. In turn, let $\omega^- = \angle BC\Omega^- = \angle CA\Omega^- = \angle AB\Omega^-$. Then,

$$\cot(\omega^-) = \cot(\angle BCA)+\cot(\angle CAB)+\cot(\angle ABC).$$

In light of the last two theorems and the fact that the cotangent function is 1–1 on the open interval $(0,\pi)$, we obtain the following theorem.

The Brocard angle theorem. Let Ω^+ and Ω^- be the positive and negative Brocard points of ΔABC respectively. In turn, let $\omega^+ = \angle CBW+ = \angle ACW+ = \angle BAW+$ and $\omega^- = \angle BCW- = \angle CAW- = \angle ABW-$. Then, $\omega^+ = \omega^-$.

Hence we refer to $\omega = \omega^+ = \omega^-$ as the Brocard angle of ΔABC.

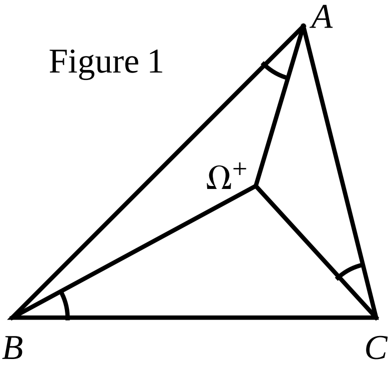


Figure 1

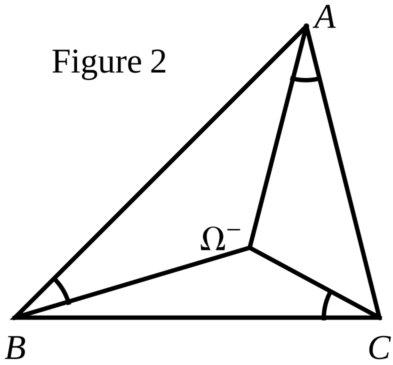


Figure 2

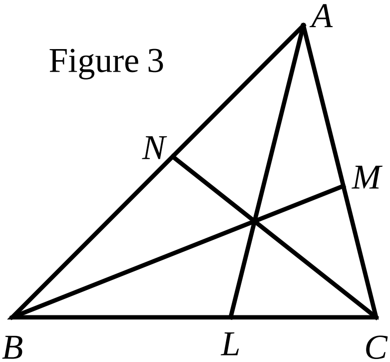


Figure 3

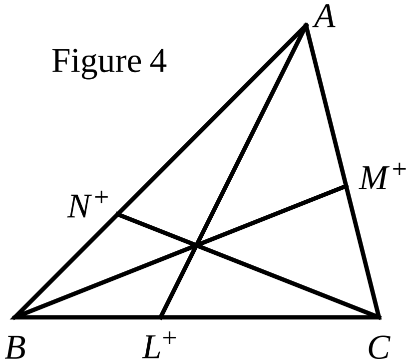


Figure 4

$$(BL)(CM)(AN) = (CL)(AM)(BN)$$

$$(a^2BL^+, b^2CM^+, c^2AN^+) = (c^2CL^+, a^2AM^+, b^2BN^+)$$

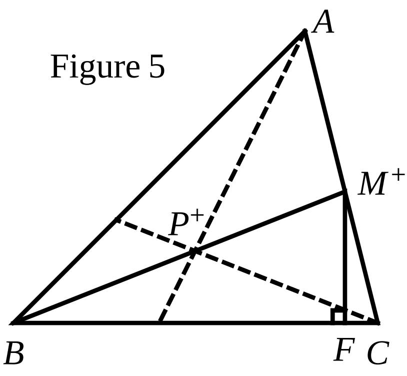


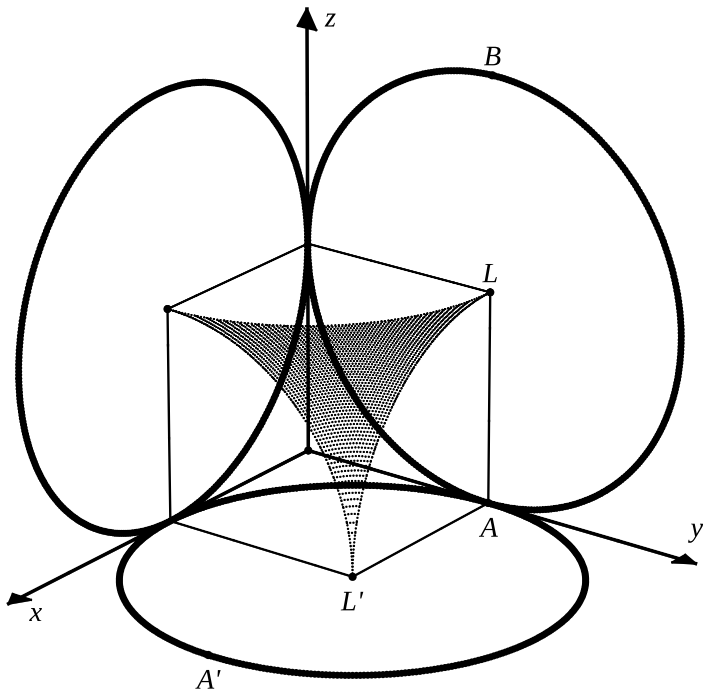
Figure 5

Solution 200.4 – Circle in a box

What is the locus of the centre of a unit-radius circle placed such that the circumference touches the positive (x, y)-plane, the positive (x, z)-plane and the positive (y, z)-plane?

Steve Moon

First, consider one of the boundary positions, where the 2p piece is rotated about the y-axis while touching the (x, z)-plane. The locus of L, the centre of the 2p piece is the positive quadrant of the circle in the plane $y = 1$, $x^2 + z^2 = 1$.



This is also the locus of L if the bottom of the coin, A, slides to A' across the (x, y)-plane while the point B slides down the (y, z)-plane to position A. (By symmetry, each position of the coin under this transformation has a mirror image under the rotation in a plane parallel to the (y, z)-plane; so the position of L under each transformation is coincident.)

By symmetry, other boundary loci are the positive quadrants of the circles $x = 1$, $y^2 + z^2 = 1$ and $z = 1$, $x^2 + y^2 = 1$. So, since all boundary loci are parts of circles, the overall locus of L is that part of a sphere of which they are cross-sections parallel to planes (x, y), (x, z) and (y, z).

The general form of a sphere is $x^2 + y^2 + z^2 = k$. Using the fact that the points satisfy $y = 1$, $x^2 + z^2 = 1$, we deduce that $k = 2$.

Hence the locus of L is that part of the sphere $x^2 + y^2 + z^2 = 2$ in the positive octant, bounded by the points of the circles $y^2 + z^2 = x = 1$, $x^2 + z^2 = y = 1$ and $x^2 + y^2 = z = 1$ in that positive region. This is the shaded surface in the diagram.

Problem 209.1 – 50p in a box

Now that we have the solutions to Problems 200.4 (2p coin in a 3-dimensional box) and 203.5 (50p in a 2-dimensional corner), how about combining the worst aspects of the two?

Drop a 50p piece into the corner of a box and move it about whilst ensuring that it remains touching the three sides of the box which meet there. What is the locus of the centre of the coin?

Problem 209.2 - Five-card trick

I am a magician. I randomize a deck of 52 cards and deal five. I look at them, I choose one and pass the other four to my assistant. She studies the four cards and then correctly identifies the fifth card. How is it done?

One answer: The five cards will contain two of the same suit. Retain the one that minimizes d, where d is the distance from the passed card to the retained card in the ordering …, 10, J, Q, K, A, 2, 3, 4, 5, …. Note that $1 \le d \le 6$. Order the other three cards to indicate d. For example, the assistant sees $(\mathrm{Q}\spadesuit, 8\clubsuit, 3\heartsuit, \mathrm{J}\diamondsuit)$; so she deduces that (i) the missing card is a spade, and (ii) since $(8\clubsuit, 3\heartsuit, \mathrm{J}\diamondsuit)$ is third in possible orderings of these three cards, its rank is $\mathrm{Q} + 3 = 2$.

Now devise similar tricks involving n cards from which you retain k and pass the other $n - k$.

Problem 209.3 – $x^y + y^x$

Show that if x is an odd power of 4, then $x^y + y^x$ is composite for all integers $y \ge 2$. What if x is an even power of 4?

Real space is hyperbolic, Euclidean space is imaginary

Dennis Morris

We take two real numbers and form an ordered pair, (a,b). The set of such ordered pairs of real numbers forms a linear space. As such, all we need to do to form an algebra is to add an appropriate multiplication operation. The multiplication operation associated with linear transformations is matrix multiplication. It is the multiplication associated with spatial motions, and we choose it for the multiplication operation. We now have

$$(a,b)(c,d) \;=\; (ac+bd, ad+bc)$$

which, in more familiar terms is the hyperbolic complex numbers.

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix} : |a| > |b| \;\cong\; a + b\hat{r} : |a| > |b|,$$

where $\hat{r} = \sqrt{+1}$. They have polar form $r(\cosh\chi + \hat{r}\sinh\chi)$. The space associated with the hyperbolic complex numbers is 2-dimensional hyperbolic space.

We repeat the above except that instead of two real numbers we take one real number and one imaginary number, $(a, b\hat{\imath})$. The set of these ordered pairs also forms a linear space, and we also choose matrix multiplication as the multiplication operation. We now have the complex numbers:

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} : a + b\hat{\imath},$$

where $\hat{\imath} = \sqrt{-1}$ and have polar form $r(\cos\chi + \hat{\imath}\sin\chi)$. The space associated with the complex numbers is 2-dimensional Euclidean space.

We now correct a widespread error of mathematical thinking. The error is the (until now unquestioned) assumption that ordered pairs of real numbers, (x,y), are associated with 2-dimensional Euclidean space. We walk to the blackboard and draw two axes, the horizontal one and the vertical one, at right angles to each other, and we proceed to associate every position on the blackboard with a pair of ordered real numbers, (x,y). But this is nonsense. If we are to use ordered pairs of real numbers, we must put them into 2-dimensional hyperbolic space.

The special theory of relativity is that we live in hyperbolic space (plus some bits). Clearly, relativity prefers real numbers to imaginary ones.

The concept of a shadow algebra

with possible implications for our view of space and quantum mechanics

Dennis Morris

The 3-dimensional algebra

$$L^1H_1^2 = \begin{bmatrix} a & b & c \\ c & a & b \\ b & c & a \end{bmatrix} : |a| > |b| + |c| \cong a + b\hat{p} + c\hat{q} : |a| > |b| + |c|$$

has multiplicative relations: $\{\hat{p}\hat{q} = 1,\, \hat{p}^2 = \hat{q},\, \hat{q}^2 = \hat{p},\, \hat{q}^3 = 1\}$. This algebra has a 1-dimensional sub-algebra which is the real numbers, but it does not have a 2-dimensional sub-algebra:

$$\begin{bmatrix} a & b & 0 \\ 0 & a & b \\ b & 0 & a \end{bmatrix} \begin{bmatrix} c & d & 0 \\ 0 & c & d \\ d & 0 & c \end{bmatrix} = \begin{bmatrix} ac & ad + bc & bd \\ bd & ac & ad + bc \\ ad + bc & bd & ac \end{bmatrix}.$$

This algebra is associated with the group C_3, which does not have a C_2 subgroup.

Because this algebra does not have a 2-dimensional sub-algebra, the 3-dimensional space associated with this algebra (through the norm and the polar form) does not have a 2-dimensional subspace. Since this algebra does have a 1-dimensional sub-algebra (the real numbers, when $b = c = 0$; the real number part of the matrix is

$$\begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{bmatrix}),$$

the 3-dimensional space associated with it does have a 1-dimensional subspace. So what remains when we remove (we do this by setting $a = 0$) the 1-dimensional subspace from the 3-dimensional space?

Important bit

The multiplicative relations of the $L^1H_1^2$ algebra are the same as the multiplicative relations of the complex numbers: $\{1, -\frac{1}{2} + \frac{1}{2}\sqrt{3}\,i, -\frac{1}{2} - \frac{1}{2}\sqrt{3}\,i\}$.

We have

$$a + b\hat{p} + c\hat{q} \sim a + b\left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right) + \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right)$$

$$= \ a - \frac{b+c}{2} + \frac{\sqrt{3}}{2}\, i\,(b-c),$$

where we have used the $\sim$ sign to indicate equivalence of multiplicative relations only. These expressions are not equal. With $a = 0$, this is

$$\frac{b+c}{2} + \frac{\sqrt{3}}{2} i(b-c).$$

We call this complex expression the 2-dimensional complex shadow of the $L^1 H_1^2$ algebra.

The norm of the $L^1 H_1^2$ algebra is the cube root of the determinant of the matrix representation of the algebra. The norm is, of course, the distance function of the 3-dimensional space of the $L^1 H_1^2$ algebra. We have

$$d^3 \ = \ a^3 + b^3 + c^3 - 3abc \ = \ (a+b+c)(a^2 + b^2 + c^2 - ab - ac - bc).$$

With $a = 0$,

$$d^3 \ = \ b^3 + c^3 \ = \ (b+c)(b^2 + c^2 - bc).$$

The determinant of the 2-dimensional complex shadow is

$$\left(-\frac{b+c}{2}\right)^2 + \left(\frac{\sqrt{3}}{2}(b-c)\right)^2 \ = \ b^2 + c^2 - bc.$$

Note: If we had not set $a = 0$, the determinant of the 2-dimensional complex shadow would still be the second factor.

The polar form of the 2-dimensional complex shadow is

$$\begin{bmatrix} e^{-\frac{1}{2}(b+c)} & 0 \\ 0 & e^{-\frac{1}{2}(b+c)} \end{bmatrix} \begin{bmatrix} \nu_{[\square]}A(b-c) & -\nu_{[\square]}B(b-c) \\ \nu_{[\square]}B(b-c) & \nu_{[\square]}A(b-c) \end{bmatrix},$$

where

$$\nu_{[\square]}A(b-c) \ = \ \cos\left(\frac{\sqrt{3}}{2}(b-c)\right),$$

$$\nu_{[\square]}B(b-c) \ = \ \sin\left(\frac{\sqrt{3}}{2}(b-c)\right).$$

The polar form of the $L^1 H_1^2$ algebra is

$$\begin{bmatrix} e^a & 0 & 0 \\ 0 & e^a & 0 \\ 0 & 0 & e^a \end{bmatrix} \begin{bmatrix} \nu_{[L^1H_1^2]}A(b,c) & \nu_{[L^1H_1^2]}B(b,c) & \nu_{[L^1H_1^2]}C(b,c) \\ \nu_{[L^1H_1^2]}C(b,c) & \nu_{[L^1H_1^2]}A(b,c) & \nu_{[L^1H_1^2]}B(b,c) \\ \nu_{[L^1H_1^2]}B(b,c) & \nu_{[L^1H_1^2]}C(b,c) & \nu_{[L^1H_1^2]}A(b,c) \end{bmatrix},$$

where

$$\begin{aligned}
\nu_{[L^1H_1^2]}A(b,c) &= \frac{1}{3}\left(a^{b+c} + 2e^{\frac{1}{2}(b+c)}\cos\frac{\sqrt{3}}{2}(b-c)\right),\\
\nu_{[L^1H_1^2]}B(b,c) &= \frac{1}{3}\left(a^{b+c} + e^{\frac{1}{2}(b+c)}\left(\sqrt{3}\sin\frac{\sqrt{3}}{2}(b-c) - \cos\frac{\sqrt{3}}{2}(b-c)\right)\right),\\
\nu_{[L^1H_1^2]}C(b,c) &= \frac{1}{3}\left(a^{b+c} - e^{\frac{1}{2}(b+c)}\left(\sqrt{3}\sin\frac{\sqrt{3}}{2}(b-c) + \cos\frac{\sqrt{3}}{2}(b-c)\right)\right).
\end{aligned}$$

This matrix is the product of two angle matrices whose elements are the 3-dimensional hyper-trig functions. With $a = 0$,

$$\begin{bmatrix} e^a & 0 & 0 \\ 0 & e^a & 0 \\ 0 & 0 & e^a \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We change the polar form of the 2-dimensional complex shadow so that the length matrix is 1. We get

$$\begin{aligned}
&\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} e^{\frac{1}{2}(b+c)}\nu_{[\square]}A(b-c) & -e^{\frac{1}{2}(b+c)}\nu_{[\square]}B(b-c) \\ e^{\frac{1}{2}(b+c)}\nu_{[\square]}B(b-c) & e^{\frac{1}{2}(b+c)}\nu_{[\square]}A(b-c) \end{bmatrix}\\
&= \begin{bmatrix} \ell\nu_{[\square]}A(b-c) & \ell\nu_{[\square]}B(b-c) \\ \ell\nu_{[\square]}B(b-c) & \ell\nu_{[\square]}A(b-c) \end{bmatrix},
\end{aligned}$$

where we have introduced the symbol ℓ to represent the length part of the trigonometric functions in the rotation matrix. We have

$$\begin{aligned}
\nu_{[\square]}A(b-c) &= \frac{3\nu_{[L^1H_1^2]}A(b,c) - e^{b+c}}{2\ell},\\
\nu_{[\square]}B(b-c) &= \frac{\sqrt{3}\nu_{[L^1H_1^2]}A(b,c) + 2\sqrt{3}\nu_{[L^1H_1^2]}B(b,c) - \sqrt{3}e^{b+c}}{2\ell},\\
\nu_{[\square]}C(b-c) &= \frac{-\sqrt{3}\nu_{[L^1H_1^2]}A(b,c) - 2\sqrt{3}\nu_{[L^1H_1^2]}C(b,c) + \sqrt{3}e^{b+c}}{2\ell}.
\end{aligned}$$

So, the 2-dimensional complex shadow is such that:

(1) The 2-dimensional distance function is without the factor.

(2) The 2-dimensional trigonometric functions are distorted versions of the 3-dimensional ones.

Perhaps the reader would like to see this phenomenon as similar to the 2-dimensional shadow of a 3-dimensional object.

Implications for our perception of space. When we combine real numbers into ordered pairs and multiply them using matrix multiplication, we get hyperbolic space. Relativity theory tells us the real space–time is hyperbolic. Yet we perceive ourselves to be surrounded by Euclidean space. Perhaps we are seeing a shadow space.

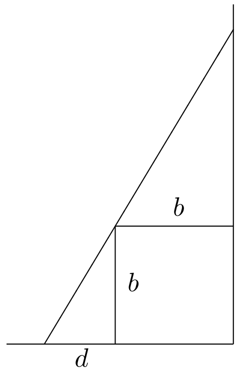
Possible implications for quantum mechanics. The mathematics of quantum mechanics is done in complex number algebra. Perhaps we are seeing just a shadow of the true algebra of quantum mechanics.

The work above has not yet been peer reviewed, and the reader should bear this in mind.

Problem 209.4 – Ladder

Norman Graham

A ladder of length 1 stands against a vertical wall just touching a shed of height and width both equal to b. Find d, the distance of the ladder bottom from the shed.



ADF—There is a similar problem, which asks, 'What is the length of the longest ladder that can be transported around a corner in a corridor?'

I was reminded of this when one day at Kingston station I saw several pieces of railway rail laid along the track. They were long—perhaps 100 metres. So I asked myself, 'How did they get there?' I suppose long pieces of iron can be transported by train if throughout the entire journey the useable width of the railway is sufficiently large and the radius of curvature is not too small. One can imagine flat wagons at regular intervals supporting the lines on some kind of roller-bearing mountings so that they can be shifted laterally to avoid track-side obstacles whenever the train goes around a bend. If both tracks are available (as is the case when there are no scheduled passenger services) it would appear that quite long lengths of rail can be moved in this way. Is this what really happens?

Solution 205.4 – abc

For integers n, a and b, define

$$q(n) = \prod_{\substack{p|n \\ p \text{ prime}}} p \quad \text{and} \quad L(a,b) = \frac{\log(a+b)}{\log q(ab(a+b))}.$$

Find triples of positive integers (a, b, c) for which $c = a + b$, $\gcd(a, b) = 1$, and $L(a, b)$ is as large as possible.

Steve Moon

After the establishment of a (sort of) strategy, a lot of 'trial and error' followed.

Assume a, b are each integer powers of small primes. I focused on 2, 3, 5, 7, 11 and stopped there as bigger p seemed to offer little benefit. Then I tried to find powers of primes p_1, p^2 such that

$$p_1^m + p_2^n \;=\; 5^k \cdot (\text{some multiple of other ideally smallish primes}).$$

The best I found were (2, 243, 245),

$$a = 2, b = 243 = 3^5, c = 245 = 5 \cdot 7^2 : \; L(a,b) = \frac{\log 245}{\log 2 \cdot 3 \cdot 5 \cdot 7} = 1.029.$$

Then $a = 2^{11}, b = 3^7, c = 5 \cdot 7 \cdot 11^2$: $L(a, b) = 1.078$ and $a = 2^6, b = 3^8, c = 5^3 \cdot 53$: $L(a, b) = 1.194$. However, these were bettered by

$$a = 3, b = 5^3, c = 2^7 : \; L(a,b) = 1.427$$

but I was unable to find other powers of primes p_1, p_1 such that $p_1^m + p_2^n = 2^k$ which might improve on this.

Problem 209.5 – Duelling lovers

Norman Graham

Ambrose, Bertram and Christopher all love Deirdre and decide to fight a 3-way duel until only one survives. They draw lots to determine who will shoot first, second and third, and in the same sequence thereafter. They stand at the vertices of an equilateral triangle. The probability of each hitting his target on every shot is A: 0.9, B: 0.8, C: 0.7. What are their chances of survival if while all three are alive A and B shoot at each other and C shoots (i) at A, or (ii) in the air?

Solving sudoku puzzles mathematically

Tony Forbes

Following on from M500 **207**, we develop a mathematical approach to the solving of sudoku puzzles such as the one on this page. (Fill in the blanks to make a Latin square on $\{1, 2, \ldots, 9\}$ with the extra constraint that the nine 3×3 boxes also contain $\{1, 2, \ldots, 9\}$.) The solution is unique—out of all the 9^{81-29} ways of filling in the empty squares with single digits taken from $\{1, 2, \ldots, 9\}$ there is precisely one which works.

	6		1		5		4	
	5			7		1		2
		6		3				4
	8	5	4	6	2	9	7	
4				5		2		
8		7		4			5	
	1		9		3		2	

We represent a sudoku puzzle by a vector $\mathcal{S} = (S_0, S_1, \ldots, S_{80})$, indexed by the set $I = \{0, 1, \ldots, 80\}$, whose elements S_i are sets of integers. Certain special subsets of I are called *regions*. If we imagine $\mathcal{S}$ arranged as a 9×9 array, a region is precisely the set of indices of a row, column, or 3×3 box. Thus index i corresponds to the cell i of the array, numbering the cells 0 to 80 from top-left to bottom-right. For example, the top row is represented by the set of indices $\{0, 1, \ldots, 8\}$, the left-hand column is $\{0, 9, 18, \ldots 72\}$ and the bottom right-hand box is $\{60, 61, 62, 69, 70, 71, 78, 79, 80\}$.

We now make some definitions, starting with an extension of the notation S_i to sets of indices. If $J \subseteq I$, define $S_J = \bigcup\{S_j : j \in J\}$. This is merely a shorthand way of referring to the whole collection of numbers that occur in a set of positions.

A vector $\mathcal{S}$ is *inconsistent* if S_i is empty for some $i \in I$.

A vector $\mathcal{S}$ is *valid* if $S_I \subseteq \{1, 2, \ldots, 9\}$ and for each region R and each $n \in \{1, 2, \ldots, 9\}$ there is at most one $i \in R$ such that $S_i = \{n\}$. This last property is important, and it essentially captures the sudoku-ness of the array.

In its initial state, the vector $\mathcal{S}$ representing a sudoku puzzle is valid, there is a set $H \subseteq I$ such that $|S_h| = 1$ for $h \in H$ and $S_i = \{1, 2, \ldots, 9\}$ for $i \in I \setminus H$. Of course, in the actual printed version of the puzzle the positions $h \in H$ will have S_h printed as a big number whilst all the other S_i will appear as blank squares.

It might help to think of S_i as the set of all those numbers which have not yet been barred from occupying position i in the sudoku array. However, in our representation S_i has no special significance—it is just a set of numbers.

To solve a puzzle we transform $\mathcal{S}$ by making changes to the S_i under two rules:

(i) the transformations must preserve the validity of $\mathcal{S}$;

(ii) they must leave S_h unchanged for $h \in H$.

But otherwise you can do anything you like. This is really true. If you want to be perverse, you can simply empty the S_i for all $i \notin H$—this is a perfectly legitimate transformation of $\mathcal{S}$ even if it doesn't achieve very much. (Remember, you are not allowed to touch the S_i for $i \in H$.) On the other hand, there do exist transformations which genuinely help to solve puzzles. The objective is to achieve a state where $|S_i| = 1$ for all $i \in I$. Ideally, but this is not vital, we want our transformations to have the additional property that the S_i do not grow in size and remain non-empty.

Let us begin by considering a specific transformation.

> **The critical set strategy**. If R is a region and $P \subset R$ such that $|P| = |S_P|$, then replace S_q by $S_q \setminus S_P$ for each $q \in R \setminus P$.

It's a combinatorics term. The P for which $|P| = |S_P|$ are called *critical sets*. Imagine applying the critical set strategy repeatedly until $\mathcal{S}$ is stable. Then it is not too difficult to prove that each region can be partitioned into critical sets $A, B, C, \ldots$ such that $|S_A| = |A|$, $|S_B| = |B|$, $|S_C| = |C|$, $\ldots$, and that a finer partitioning with the same property is impossible.

Whilst the phrase 'critical set strategy' might be unfamiliar, I am sure that all sudoku addicts use an elementary method based on a couple of special cases, which I shall call *basic transformations*. First let us restrict the sizes of the critical sets to 1. Then we have the simple rule:

Basic transformation 1. If $\{n\}$ is in region R, $\{n\}$ cannot go anywhere else in R.

In our language, the critical set is $\{i\} \subset R$ and $S_{\{i\}} = \{n\}$. For instance, if you apply this rule to the above puzzle, you can put $\{8\}$ in row 8, column 5. And for the other well-known case we have:

Basic transformation 2. If $\{n\}$ can't go anywhere else in a region, then it must go here.

Thus we can put $\{5\}$ in row 4, column 7 because 5 is blocked from everywhere else in row 4 by the $\{4\}$ at column 9 and by $\{5\}$s present in regions which intersect row 4. Again, this is the critical set strategy in action, this time with sets of size 8: $P = \{27, 28, 29, 30, 31, 32, 34, 35\}$, $S_P = \{1, 2, 3, 4, 6, 7, 8, 9\}$, and hence we can remove 1, 2, 3, 4, 6, 7, 8 and 9 from S_{33}.

By the way, I describe puzzles as 'easy' if they can be completely solved by the two basic transformations. The main reason is that you do not need to make notes. Stare at the puzzle until you see where a basic transformation applies; write a number in a square; repeat until solved. Curiously, I find that a considerable number of puzzles in books, magazines and newspapers have this property even though their publishers describe them with adjectives like 'fiendish', 'advanced', 'difficult', 'tough', etc.

The critical set strategy is more powerful than the basic transformations, and indeed it is sufficient to dispose of the majority of published puzzles. Our example, however, is an exception. There are 29 starter digits, the basic transformations yield 20 new numbers, and the general critical set strategy produces three more, making a total of 52. You might like to confirm this by actually attempting a solution. See if you end up with the array on the next page, where I have shown the entire contents of the vector $\mathcal{S}$ (using small digits for S_i, $i \notin H$). As you can see, each region is partitioned into critical sets. In row 6, for example, the partitioning is

$$\{\{4\},\ \{5\},\ \{2\},\ \{3,7\} \cup \{3,9\} \cup \{7,8\} \cup \{1,9\} \cup \{1,6,8\} \cup \{6,8\}\},$$

three sets of size 1 and one set of size 6.

Once you can see the 'marked up' puzzle, completing it is not difficult. If you look at the left-hand column, you will notice that there are no 3s except in the top left-hand box. Therefore the 3 in these regions must go somewhere in the overlap. Hence we can eliminate 3 from all other cells in the box. This is an example of another general transformation:

Intersecting regions. Let $A, B \subset I$ be distinct regions and suppose $n \in S_{A\cap B}$ but $n \notin S_{A\setminus B}$. Then we can remove n from all S_j for $j \in B \setminus A$.

2379	4	1	36	29	8	67	369	5
2379	**6**	239	**1**	29	**5**	8	**4**	79
39	**5**	8	36	**7**	4	**1**	369	**2**
279	27	**6**	78	**3**	19	5	18	**4**
1	**8**	**5**	**4**	**6**	**2**	**9**	**7**	3
4	37	39	78	**5**	19	**2**	168	68
8	9	**7**	2	**4**	6	3	**5**	1
5	**1**	4	**9**	8	**3**	67	**2**	67
6	23	23	5	1	7	4	89	89

In our example $S_{11} = \{2, 3, 9\}$ can be replaced by $S_{11} = \{2, 9\}$. This creates a new critical set, $\{11, 13\}$ with $S_{\{11,13\}} = \{2, 9\}$, in row 2, which allows you to remove the 9 from $S_{17} = \{7, 9\}$ at the end of row 2. The rest is easy.

However, there exist puzzles for which critical sets combined with intersecting regions will not work. So we now define one transformation which will always work, no matter what other methods fail.

Backtracking. If $\mathcal{S}$ is inconsistent, do nothing. Otherwise choose any i such that $|S_i| > 1$. If there are none, the puzzle is solved, so report the solution. Otherwise for each $n \in S_i$, perform the following.

> Save $\mathcal{S}$. Replace S_i by $\{n\}$. Apply the critical sets and intersecting regions procedures until S is stable. Perform the backtracking transformation. Restore $\mathcal{S}$.

Backtracking is simply the process of trying things out in a systematic fashion, discarding any choice that leads to an inconsistent $\mathcal{S}$. It works on any sudoko-type array, not just genuine puzzles and if allowed to run to completion, it will eventually report the entire solution set although in some cases human mortality may prevent you from seeing the final result.

Note that the changes made by backtracking are only temporary; the overall effect on $\mathcal{S}$ is to leave it unchanged.

Although I have concentrated on critical sets and intersecting regions, it is worth noting that there are other strategies which do not involve trial. However, I have not investigated them to any great extent.

Summarizing what we have so far, there are four strategies:

(0) the basic transformations,
(1) critical sets,
(2) intersecting regions,
(∞) backtracking.

When I started getting interested in these things I assumed, somewhat naïvely as it turned out, that puzzles with many starter digits would in general be easy to solve. Consider (1) in the above list. How big must a puzzle be for the critical set strategy to guarantee a solution?

Given a puzzle $\mathcal{S}$, let $\phi(\mathcal{S})$ denote the number of cells (including starter digits) that can be determined by the critical set strategy. Apart from 81, how big can $\phi(\mathcal{S})$ be? My initial impression was that $\phi(\mathcal{S})$ should be quite small. Most published puzzles have 20–30 starter digits, and most of them yield to (1). Therefore, allowing for a generous amount of slack, I expected $\phi(\mathcal{S})$ to be somewhat less than about 50. So it came as a bit of a shock to discover puzzles with $\phi(\mathcal{S})$ considerably greater than this. Indeed, I reported in M500 **207** that I had found puzzles with $\phi(\mathcal{S})$ taking every value in the range 22 to 70. The large ones, such as the 70-digit example in M500 **207**, are easy enough to complete, requiring only a modicum of backtracking. However, I find their existence exceedingly surprising. I can't help thinking that there ought to be a way of constructing these things logically rather than by trial, the method I have been using so far.

Similarly, we can define $\psi(\mathcal{S})$ to be the total number of cells that can be determined by a combination of critical sets and intersecting regions ((1) and (2) in the list). Even more surprisingly, there is approximately the same range of possible values. So far I haven't found an example with $\psi(\mathcal{S}) < 23$; the only puzzle in my collection with $\phi(\mathcal{S}) = 22$ can be completed with intersecting regions. However, I do know that $\psi(\mathcal{S})$ takes all values from 23 to 70. As with $\phi(\mathcal{S})$, upper and lower limits must exist for $\psi(\mathcal{S})$, but as yet I do not have any proofs.

Finally, I offer six puzzles and invite you to confirm that they are all closed under strategies (1) and (2). The last one is interesting. It is the only example I know of (apart from trivial variations) where my computer program requires two levels of backtracking.

			6					
				9	5	1	7	
	4			7				
5						7		3
	9		7					8
3	7			1			2	
		6	8	4			5	
		3						6

				8			9	
6					4			
		5	2		1	3		
			8		7			
			4		6			
3	8	1						
		4						1
	7			4				2
	3			5	9		6	

							8	7
	4				5		1	
	7			8			6	
		4	7	3		8		
3		8		5		7		9
7			4					
	1				2		7	5
		3	5			2		

			4					5
			1		2	8		
							6	
9		8						2
	3	7						1
		3	9	5	8			
	7	9	3	2	4	5		
8	5	2	7	6	1		4	

					9		5	
9							2	8
						7	9	
			9		4	2		
2	9	6		1	7	8		
4	5	3		8		9	1	7
1		9		4	5		8	2
3			1	2	8		7	9
	8	2		9				1

1		6	2		4	3		
	4			5				
5		7			8	4		9
2		4	1	3	6			
	5		8	4	9			
		8	5	2	7			4
3			4			1		6
							4	
4		9						7

A note on Pell's equation

Sebastian Hayes

The mathematical level of articles in M500 is becoming so high that puzzling out steps in a given argument takes me weeks (or months) and involves me in revision of long forgotten topics. Bryan Orman's stylish article, 'Machin's formula' (M500 **204**), at one point involves the solution of $C^2 - 2N^2 = -1$ in integers, and this recalled Pell's equation, so-called, $X^2 - NY^2 = 1$. It would seem incidentally that history has been overkind to Pell, who is described by the historian Bell as 'mathematically a non-entity and humanly an egregious fraud' (!).

We look for non-trivial solutions in positive integers and thus reject $X = 1, Y = 0$. We call the number $\alpha = r + s\sqrt{N}$, where r, s are positive integers and N is not a perfect square—an important point—a *solution* to $X^2 - NY^2 = 1$ if and only if $r^2 - Ns^2 = 1$. Nothing is lost by sticking to positive values since the solutions come in quadruplets because of the squaring and once we have the absolute values of r and s, we can derive the others. Thus, if $r + s\sqrt{N}$ is a solution, so is $r - s\sqrt{N}$, $-r + s\sqrt{N}$ and $-r - s\sqrt{N}$. Note that if $\alpha = r + s\sqrt{N}$, $1/a = r - s\sqrt{N}$—the conjugate' if you like—is also a solution because the denominator r^2Ns^2 is equal to 1.

The master theorem for the solution of Pell's equation is

Theorem. If θ is the generator for all solutions to $X^2 - NY^2 = 1$, then all non-trivial solutions in positive integers are given by θ^k, $k = 1, 2, 3, \ldots$.

I shall not attempt to prove this but an example makes it plausible. If α is a solution to $X^2 - NY^2 = 1$ then so is α^2. For

$$\alpha^2 = (a + b\sqrt{N})(a + b\sqrt{N} = (a^2 + Nb^2) + (2ab)\sqrt{N};$$

$$(a^2 + Nb^2)^2 - N(2ab)^2 = (a^2 - Nb^2)^2 = 1^2 = 1.$$

More generally it can be easily shown that if α and β are solutions, then so is $\alpha\beta$—and if we want to include negative exponents we set $\beta = 1/\alpha$.

By induction we conclude that, given one solution α, all the powers of α also provide solutions. This, however, does not show that such solutions are the only ones, nor even that a generator must exist. In a few cases a solution can be found at once by trial—or rather by scanning lists of squares. However, even quite small values of N can require a lot of effort—apparently $X^2 - 61Y^2 = 1$ has no solutions until we get to $X = 1766319049$.

The theory of continued fractions does in fact guarantee a solution, and thus a generator of indefinitely many solutions (see Theorem 13.16 and the preceding lemma on pages 335–6 of Burton, *Elementary Number Theory*). In brief, if N is not a perfect square, $\sqrt{N}$ can be developed as an infinite continued fraction with recurring period $\sqrt{N} = [a_0; a_1, a_2, \ldots, 2a_0]$, where a_0 is an integer and the fractional part has period running from a_1 to the last term of the period, which is always equal to $2a_0$. The period will either contain an odd or an even number of terms—the case of a single term period is best considered apart. If the number of terms in the period is even, the penultimate convergent of every cycle provides a solution to the equation $X^2 - NY^2 = 1$.

Thus $\sqrt{7} = [2; 1, 1, 1, 4]$, where the recurring part has four terms. If we select the convergent associated with the 1 just before the 4, i.e. p_3/q_3, we have 8/3 and, as it happens, $8^2 - 7 \cdot 3^2 = 1$.

Another example: $\sqrt{6} = [2; 2, 4]$. The convergents are

$$2/1,\ 5/2,\ 22/9,\ 49/20,\ 218/89,\ \ldots.$$

The period consists of two terms, an even number. The penultimate convergents of all cycles should thus provide solutions to $X^2 - 6Y^2 = 1$. We have in effect

$$5^2 - 6 \cdot 2^2 \ = \ 25 - 24 = 1; \quad \text{also} \quad 49^2 - 6 \cdot 20^2 \ = \ 1.$$

However, if the number of terms in the period is odd, the penultimate convergents alternate, giving -1 and $+1$. Thus, for example, $\sqrt{13} = [3; 1, 1, 1, 1, 6, \ldots]$ with period five. The convergents are

$$3/1,\ 4/1,\ 7/2,\ 11/3,\ 18/5,\ 119/33,\ \ldots$$

and taking the penultimate 18/5 we obtain $18^2 - 13 \cdot 5^2 = -1$.

If $\sqrt{N}$ has a period of *one*, it will be of the form $[a_0; 2a_0]$ and in this case the convergents will give -1 and $+1$ alternately, beginning with $a_0/1$. Thus in the case of $\sqrt{2}$ we have $[1; 2]$ with convergents 1/1, 3/2, 7/5, … giving $1^2 - 2 \cdot 1^2 = -1$, $3^2 - 2 \cdot 2^2 = +1$, and so on.

If the continued fraction representation of $\sqrt{N}$ has an even period, then there are no solutions in integers to $X^2 - N \cdot Y^2 = -1$.

As an example of generating solutions, take $5 - 2\sqrt{6}$ as generator and the third power. Then we have

$$\begin{aligned} (5 - 2\sqrt{6})^3 &= 5^3 - 3 \cdot 5^2(2\sqrt{6}) + 3 \cdot 5(2\sqrt{6})^2 + (2\sqrt{6})^3 \\ &= 125 + 15 \cdot 24 - \sqrt{6}(6 \cdot 5^2 + 8 \cdot 6) \quad = \quad 485 - 198\sqrt{6}. \end{aligned}$$

Trying this out we find that $485^2 - 6 \cdot 198^2 = 235225 - 235224 = 1$.

Letters to the Editor

The complex numbers are not an algebraic field extension of the real numbers

I write to correct a misunderstanding that is prevalent within modern mathematics.

The Misunderstanding. The complex numbers are the real numbers together with the number $\hat{i} = \sqrt{-1}$.

The Correct Understanding. The complex numbers are the 2×2 matrices of the form $\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$ together with the 2×2 matrices of the form $\begin{bmatrix} 0 & b \\ b & 0 \end{bmatrix}$. (All elements within the matrices are real numbers.)

Post-amble. (a) We have $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}^2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$.

(b) The 2×2 matrices of the form $\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$ are algebraically isomorphic to the real numbers. However, algebraic isomorphism is not identity. Two algebras are algebraically isomorphic if the operations of addition and multiplication produce the same results in both algebras. (We also need the operation of multiplication by a scalar, but that obscures the point at issue here.) Square-rooting is not an algebraic operation, and square-rooting need not produce the same answers in algebraically isomorphic algebras. The positive real numbers have only two square roots; the 2×2 matrices of the form $\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$ have an infinite number of square roots.

(c) Conventional wisdom has it that the complex numbers are an algebraic field extension of the real numbers because they are associated with the polynomial $x^2 + 1$, which will not split into linear factors over the real numbers. The algebra of the Study numbers is the 2×2 matrices of the form $\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$ together with the 2×2 matrices of the form $\begin{bmatrix} 0 & b \\ b & 0 \end{bmatrix}$ such that $|a| > |b|$. This is a bona fide algebra, but it is not an algebraic field extension of the real numbers because the polynomial with which it is associated, $x^2 - 1$, splits into linear factors over the real numbers. Clearly, conventional wisdom does not apply to 2×2 matrices.

Methinks. It seems to me that the root of the misunderstanding that I have above corrected is the habit of writing a complex number in the form $a + \hat{i}b$ as if it were a 1-dimensional object. It would be better written as (a, b) or in matrix form. In such form, we notice the potential errors more easily.

For example, while it is obvious that $e^{a+\hat{\imath}b} = e^a e^{\hat{\imath}b}$, it is not immediately obvious that $e^{(a,b)} = e^a e^b$. It does because 2×2 matrices of the form $\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$ commute with matrices of the form $\begin{bmatrix} 0 & b \\ b & 0 \end{bmatrix}$. At a deeper level, the root of such misunderstandings is the misunderstanding that 2-dimensional space is two 1-dimensional spaces fixed together—an incorrectitude made obvious by the existence of both 2-dimensional hyperbolic space and 2-dimensional Euclidean space.

Dennis Morris

Coins

I don't see why Boris [who is convinced that a straight line drawn on a coin becomes curved when the coin is heated even though Marina disagrees [M500 206 p. 19]] should for a moment suppose that the line on the coin should curve when it is heated, any more than would happen when he looks at the coin under a magnifying glass. Am I missing something here?

Ralph Hancock

Zoe's design

Tony Forbes

The cover of this magazine shows a type B 3-colourable Steiner system $S(2, 4, 61)$ with colour class sizes 39, 19 and 3. It was found by one of my computers on 25 March 2005, by coincidence the 21st birthday of my youngest daughter. Hence the name, 'Zoe's design'.

The 61 points A_0, A_1, ..., A_{38}, B_0, B_1, ..., B_{18}, C_0, C_1 and C_2 appear in 305 blocks of four. Each pair of points occurs in exactly one block. The 'colours' are denoted by letters A, B and C, and 'type B' refers to the pattern XXXY, where each block contains three elements of one colour and one element of different colour. Also the triples of a given colour form the block set of a Steiner triple system. So you can also regard it as a 'stitching together' of an STS(39), an STS(19) and an STS(3).

The cover of M500 **205** indicates a similar structure involving 100 points and two colours, known as the *Design of the Century*, found in June 2005. Details are in the forthcoming paper: A. D. Forbes, M. J. Grannell & T. S. Griggs, 'The design of the century', *Mathematica Slovaca.*

The existence of these combinatorial designs has been an unsolved problem for over 25 years.

Contents — M500 209 – April 2006