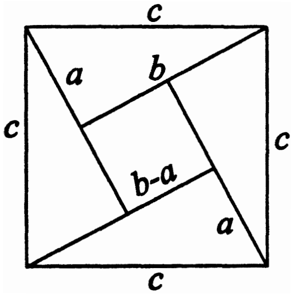
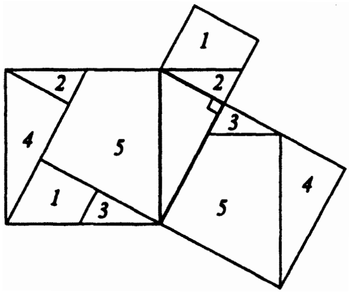
Vol. 73, No. 4, October 2000

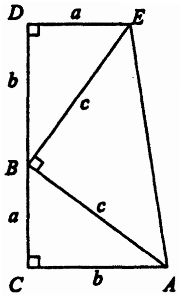
MATHEMATICS MAGAZINE



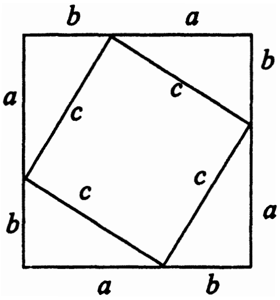
Bhaskara (India) 12th century



Ibn-Cora (Arabia) 9th century



Garfield (US) 1881



Chou-pei Suan-ching (China)
3rd century BCE

- Pythagoras After 2500 Years
- The Gyro-Structure of the Unit Disk
- PIPCIRs

EDITORIAL POLICY

Mathematics Magazine aims to provide lively and appealing mathematical exposition. The *Magazine* is not a research journal, so the terse style appropriate for such a journal (lemma-theorem-proof-corollary) is not appropriate for the *Magazine*. Articles should include examples, applications, historical background, and illustrations, where appropriate. They should be attractive and accessible to undergraduates and would, ideally, be helpful in supplementing undergraduate courses or in stimulating student investigations. Manuscripts on history are especially welcome, as are those showing relationships among various branches of mathematics and between mathematics and other disciplines.

A more detailed statement of author guidelines appears in this *Magazine*, Vol. 71, pp. 76–78, and is available from the Editor. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, or published by another journal or publisher.

Send new manuscripts to Frank Farris, Editor-Elect, Department of Mathematics and Computer Science, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053-0290. Manuscripts should be laser-printed, with wide line-spacing, and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit three copies and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Copies of figures should be supplied on separate sheets, both with and without lettering added.

AUTHORS

Peter Andrews received a B.Sc. from the University of Toronto in 1971 and a doctorate from the University of Maryland in 1975. Since then he has taught at Dartmouth, Williams, McMaster, Wilfrid Laurier, and Middlebury before coming to Eastern Illinois University in 1992. He coordinates the Computer Science program there while dabbling in computer graphics, algorithms, geometry, and topology and marvelling at how flat it is in central Illinois.

Rachel Belinsky received her Ph.D. from Leningrad (now St. Petersburg) University, Russia. She has taught at St. Petersburg Naval Technical University and, for the last seven years, at Morris Brown College in Atlanta. Her primary mathematical interests were in functional analysis and ergodic theory, but she drifted into applications, mostly in hydrodynamics and thermodynamics. The questions addressed in this article brought her back to pure mathematics.

Michael Kinyon received his B.S., M.S., and Ph.D. degrees from the University of Utah in 1986, 1988, and 1991, respectively. Since 1992 he has been on the faculty at Indiana University South Bend. He began his research career studying algebraic methods for differential equations. His collaboration with Ungar, which led to the present paper, sparked a more general interest in quasigroups and loops and their applications to differential geometry. His teaching interests include differential equations, the history of mathematics, and students, not necessarily in that order.

Abraham A. Ungar received his Ms.C. from the Hebrew University in Jerusalem (1968) and his Ph.D. from Tel Aviv University. He has been Professor of Mathematics at North Dakota State University, in Fargo, since 1984. He wishes to share the observation that, although more than 150 years have passed since August Ferdinand Möbius first studied the transformations that now bear his name, the rich structure that he thereby exposed is still far from being exhausted.

Darko Veljan received his B.S. and M.S. from the University of Zagreb, and his Ph.D. from Cornell University in 1979. Since 1971 he has taught at the University of Zagreb, where he is now Professor of Mathematics. His interests include algebraic topology, combinatorics, and, especially, elementary geometry and mathematics education. He is co-author of several textbooks in the Croatian language. According to one legend, the Pythagoreans grilled about 100 oxen to celebrate the discovery of the theorem. Since then, it seems, oxen have feared mathematics. The present article argues that nobody—not even oxen—should fear the beauty of mathematics.

Vol. 73, No. 4, October 2000

MATHEMATICS MAGAZINE

The *MATHEMATICS MAGAZINE* (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.
The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $16 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $68 per year.
Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Advertising correspondence should be addressed to Ms. Elaine Pedreira, Advertising Manager, the Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

Periodicals postage paid at Washington, D.C. and additional mailing offices.

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036-1385.

Printed in the United States of America

ARTICLES

The 2500-Year-Old Pythagorean Theorem

DARKO VELJAN
University of Zagreb
Bijenička c. 30
10000 Zagreb
Croatia

Introduction and a bit of history

When you see a paper with "Pythagorean theorem" in its title, you might say "I know this stuff" and skip it. But I think it's still worthwhile thinking about the "good old Pythagorean theorem," to which this paper is devoted.

Pythagoras was born about 570 B.C. on the island of Samos and died about 490 B.C. Many other well-known philosophers lived and worked around the same time, but in other civilizations. Let us mention only Gautama Buddha in eastern Asia, Confucius (or Kung Fu-tse) in China, Zoroaster (or Zarathustra) in Persia, and the prophet Isaac (or Iitzak) in Judea. Was this simultaneous flowering of philosophy a mere accident?

Pythagoras was well educated, learning to play the lyre and to recite Homer. Most important among his teachers was Thales of Miletus (ca. 624–548 B.C.), who introduced Pythagoras to mathematical ideas and to astronomy, and sent him to Egypt to learn more of these subjects. Pythagoras returned to Samos, but soon left the island to escape the tyranny of Polycrates, settling in Croton in southern Italy, where he founded a philosophical and religious school that had many followers. Pythagoras led the Society, with an inner circle of followers known as *mathematikoi*. The *mathematikoi* lived permanently with the Society, had no personal possessions, and were vegetarians. They were taught by Pythagoras himself and obeyed strict rules. Among the beliefs of Pythagoras were these: (1) reality is fundamentally mathematical; (2) philosophy can lead to spiritual purification; (3) the soul can rise to union with the divine; (4) certain symbols have mystical significance; and (5) all members of the order should observe strict loyalty and secrecy.

Pythagoras and the *mathematikoi* studied mathematics, but they were less interested in formulating and solving problems than in the principles of mathematics, the concepts of number, triangle, and space form, and the abstract idea of proof. Pythagoras believed that all of nature and its order could be reduced to numerical relations. He studied properties of even, odd, triangular, and perfect numbers, and he assigned to each number its own "personality." Numbers might be masculine or feminine, perfect or incomplete, beautiful or ugly, etc. Ten was regarded as perfect: it contained the first four integers ($1 + 2 + 3 + 4 = 10$) and when written in dot notation formed a perfect triangle.

The Society at Croton was not unaffected by political events, despite Pythagoras's desire to avoid politics. After the Society was attacked, Pythagoras escaped to Metapontium. It is believed that he died there, though some authors claim that he committed suicide after the attack on his Society. The Society expanded rapidly after 500 B.C., became political in nature and then split into factions. About 460 B.C. the

Society was violently suppressed. Its meeting houses were sacked and burned, and in the fights between democrats and aristocrats many Pythagoreans were slain. The survivors took refuge at Thebes and elsewhere. Soon thereafter the Pythagorean Society disappeared and emerged again only 5–6 centuries later, in a form heavily influenced by Platonism.

Pythagoras and his followers were the first in history to give geometric considerations a scientific flavor, and they first recognized the need for systematic proof. Only two centuries later did Euclid fully comprehend this approach with his *Elements*, which contained many ideas of Pythagoras, and set new standards for mathematical rigor and logical structure.

The Pythagoreans also studied music, noticing that vibrating strings produce musical tones when ratios of string lengths are whole numbers. In astronomy they realized that the same planet, Venus, appeared both as a morning star and as an evening star.

Pythagoras's most important heritage is his famous theorem. In some form, however, this theorem was known much earlier, as we know from drawings, texts, legends, and stories from Babylon, Egypt, and China, dating back to 1800–1500 B.C. One well-known story (perhaps mythical) holds that Egyptian peasants used a rope with evenly spaced knots to form a 3–4–5 right triangle, which they used to re-measure their agricultural plots flooded each year by the river Nile. Thus, the Pythagorean theorem was an early example of an important fact rediscovered independently and often. But Pythagoras first formulated it in general.

According to one legend Pythagoras discovered "his" theorem while waiting in a palace hall to be received by Polycrates. Being bored, Pythagoras studied the stone square tiling of the floor and imagined the right triangles (half-squares) "hidden" in the tiling together with the squares erected over its sides. Having "seen" that the area of a square over the hypotenuse is equal to the sum of areas of squares over the legs, Pythagoras came to think that the same might also be true when the legs have unequal lengths. In any case, Pythagoras is rightly described as the first pure mathematician in history. (Additional history can be found in **[1]** and **[16]** and in their references.)

The Pythagorean theorem

This is probably the only nontrivial theorem in mathematics that most people know by heart. A good many might even know how to prove it more or less correctly. This "Methuselah" among theorems is one of the most quoted theorems in the history of mathematics, particularly in elementary geometry. But this "folklore" theorem re-

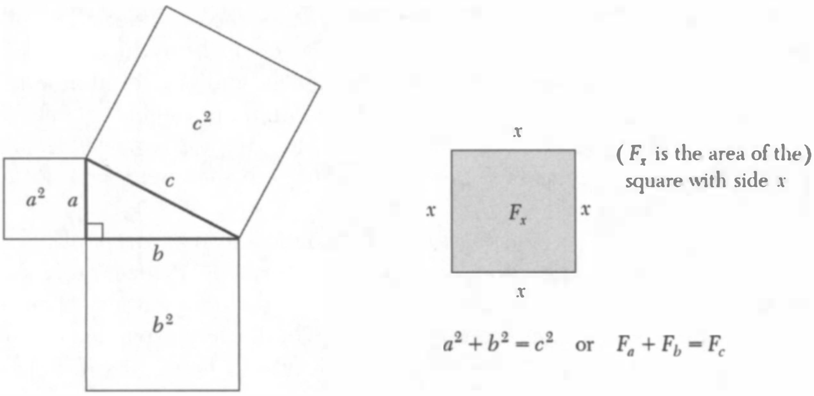


FIGURE 1
The Pythagorean theorem.

mains eternally youthful, as many people continue to find new interpretations, generalizations, analogues, proofs, and applications.

The Pythagorean theorem asserts: *The sum of the squares of the legs of a right triangle is equal to the square of its hypotenuse*. (See FIGURE 1.)

Here is a nice informal interpretation of the Pythagorean theorem: A pizza shop makes three sizes of pizzas; their diameters are the sides of a right triangle. Then the big pizza is equal to the sum of the two smaller pizzas. (See FIGURE 2a.)

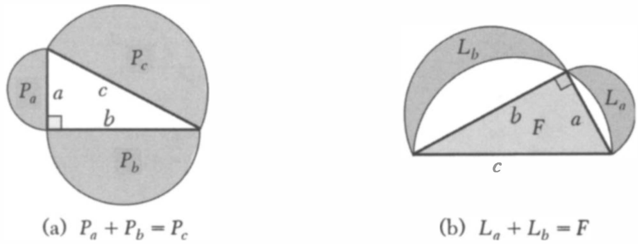


FIGURE 2
Interpreting the theorem.

Another interpretation of the Pythagorean theorem involves *Hippocrates' lunes*. Construct semicircles whose diameters are the sides of a right triangle, as in FIGURE 2b. Then the sum $L_a + L_b$ of the areas of two "lunes" is equal to the area F of the triangle. This follows immediately from the previous interpretation by subtracting from the "half pizzas" the white circular segments. Hippocrates of Chios (ca. 450 B.C.) tried in this way to "square a circle." Since he was able to construct a right triangle whose area is the sum of two lunes formed from circular arcs, he hoped to construct (using only ruler and compass) a square whose area is equal to that of a given circle. Thus originated the famous "squaring the circle" problem. This was shown to be impossible in 1822 by F. Lindemann, who showed that π is not a root of any algebraic equation with rational coefficients. (Some authors claim that the circle-squaring problem was known 200 years before Hippocrates.)

Among the oldest and most important consequences of the Pythagorean theorem is the incommensurability of the diagonal and the side of a square. This fact was the first evidence of the existence of irrational numbers, which in turn led to the basic mathematical concept of real numbers. The Pythagoreans, shocked by the discovery of irrational numbers, tried to keep the concept secret. When Hippasos broke the secret, he was expelled from the Society, and when he died in a ship accident, Pythagoreans recognized it as God's punishment. Soon after Hippasos's treachery, the philosopher Plato (427–347 B.C.), realizing the importance of the discovery, thundered: "One who is not aware that the side and the diagonal of a square are incommensurable does not deserve to be called a man." Hippasos was the first to construct a regular pentagon and to raise the question of constructibility. Only in 1796 was the problem solved by C. F. Gauss (1777–1855): A regular n-gon can be constructed with compass and ruler if and only if n is a power of 2 or is of the form $n = 2^r p_1 p_2 \cdots p_k$, where r is a nonnegative integer and the p_i's are different *Fermat primes*, i.e., primes of the form $2^{2^a} + 1$. (Pierre Wantzel proved the "only if" part some 40 years later, although Gauss asserted it.) It is not known whether infinitely many Fermat primes exist (3, 5, 17, and 257 are the first four).

It is well known (see, e.g., **[6]**) that all integer solutions of the Pythagorean equation $a^2 + b^2 = c^2$ are given by

$$a = (m^2 - n^2)t, \quad b = 2mnt, \quad c = (m^2 + n^2)t, \quad m, n, t \in \mathbb{Z}.$$

With $m = 2$, $n = 1$, and $t = 1$ we get the Pythagorean triple $(a, b, c) = (3, 4, 5)$, the sides of the "Egyptian triangle."

The analogous diophantine equation $a^n + b^n = c^n$, $n \geq 3$, $abc \neq 0$, has no integral solution, according to Fermat's Last Theorem, which was stated by Fermat about 1650 and proved only in 1995 by A. Wiles ([**18**]). Closely related is the *ABC-conjecture* which, roughly speaking, says that a diophantine equation $A + B = C$ has no solutions if A, B, and C are all divisible by many factors (for details and examples, see [**10**]). Another open number-theoretic problem stems from the Pythagorean theorem. Fermat proved that there is no right triangle with rational sides and area 1, but the general problem where 1 is replaced by an arbitrary integer n remains unsolved. The simplest right triangle with all sides rational and area 157 is shown in Figure 3 (see [**18**]):

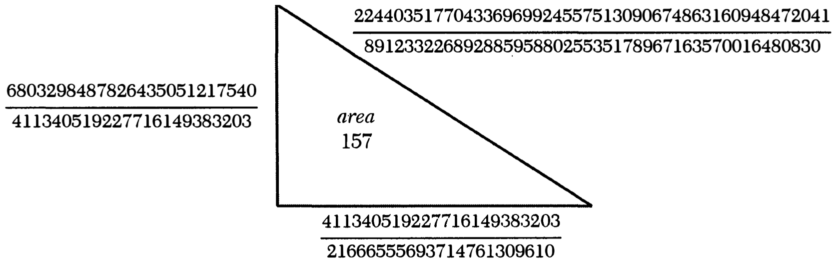


FIGURE 3
A rational right triangle.

In modern times, classical geometry (including the Pythagorean theorem) is being revived in various applications and in new areas such as discrete, combinatorial, and computational geometry. Much credit for this belongs to the Hungarian mathematician Pál Erdős (1913–1996), after Euler the most prolific mathematician of all time, with over 1500 published papers. Erdős raised and solved questions which even Pythagoras and Euclid would understand and appreciate. We quote only one theorem (whose proof relies on the Pythagorean theorem) from Erdős:

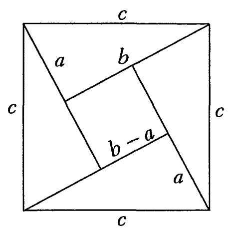
Let n points be given in a plane, not all on a line. Join every pair of points by a line. At least n distinct lines are obtained in this way.

(This is often referred as the Sylvester–Gallai–Erdős' theorem; see [**3**] and [**12**].)

Some proofs of the Pythagorean theorem

About 400 different proofs of the Pythagorean theorem are known today. (See [**2**], [**8**], [**11**].) We shall present here only a few proofs, mostly "without words," but with some short historical comments.

1. Bhaskara (India) XIIth century (Figure 4):



$$(b-a)^2 + \frac{4ab}{2} = c^2 \Rightarrow$$
$$a^2 + b^2 = c^2$$

FIGURE 4
Bhaskara's proof.

2. Cutting squares. Ibn-Cora (Arabia), IXth century (FIGURE 5):

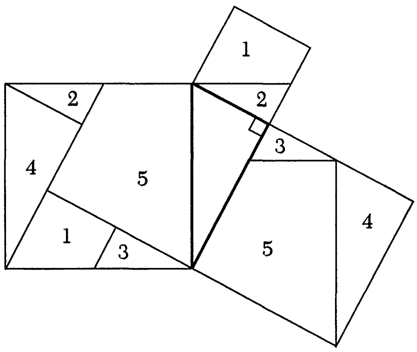
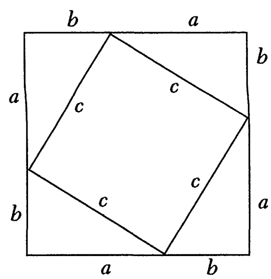


FIGURE 5
Ibn-Cora's proof.

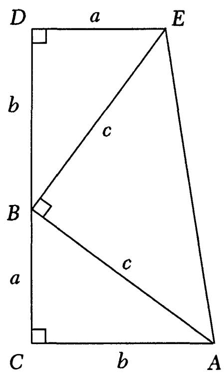
3. Dissecting one square. Chou-pei Suan-ching (China), ca. 250 B.C. (FIGURE 6):



$$(a+b)^2 = c^2 + 4\cdot\frac{ab}{2} \Rightarrow a^2+b^2=c^2$$

FIGURE 6
Chou-pei Suan-ching's proof.

4. Garfield's trapezoid (FIGURE 7):



The area of the trapezoid $CAED$ is found in two ways;

$$\frac{a+b}{2}\cdot(a+b) = 2\cdot\frac{ab}{2} + \frac{1}{2}c^2 \Rightarrow a^2+b^2=c^2$$

FIGURE 7
Garfield's proof.

This proof was given in 1881 by the U.S. President James A. Garfield. This proof was apparently also known in Arabia and India in the 7th century. Note that FIGURE 7 is half of FIGURE 6.

5. Equidecomposability, apparently due to Leonardo da Vinci, 15th century (FIGURE 8):

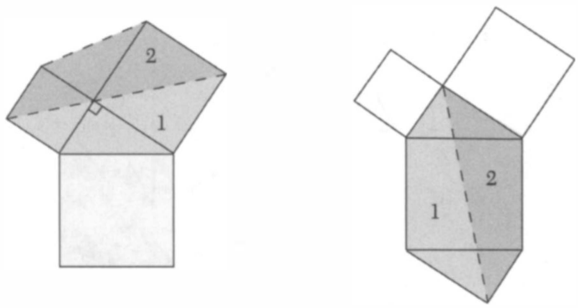


FIGURE 8
Leonardo's proof.

The Pythagorean theorem is a special case of the *law of cosines*:

If $\triangle ABC$ *has sides* a, b, *and* c, *and* $\gamma = \angle C$, *then* $c^2 = a^2 + b^2 - 2ab\cos\gamma$.

If $\gamma = \pi/2$, we get $c^2 = a^2 + b^2$.

The law of cosines is usually derived using the Pythagorean theorem. (Drop a perpendicular from A to BC, apply the Pythagorean theorem to both of the resulting right triangles, and use the definition of the cosine.) So the law of cosines is, in fact, equivalent to the Pythagorean theorem, since each implies the other. The *law of sines* is also equivalent to the law of cosines, so one could say, a bit vaguely, that plane geometry (together with some evident axioms) is nothing but the Pythagorean theorem.

We remark that the cosine law can be phrased without trigonometric functions:

In any triangle, the sum of squares of two sides is equal to the square of the third side increased by twice the product of the first side with orthogonal projection of the second to the first side.

This can be proved in a purely geometric, i.e., Euclidean style.

The converse of the Pythagorean theorem holds as well:

If the sides a, b, c *of* $\triangle ABC$ *satisfy the relation* $a^2 + b^2 = c^2$, *then it is a right triangle with the right angle at vertex* C.

Indeed, let $\triangle A_1B_1C_1$ be a right triangle with legs of lengths $B_1C_1 = a$, $A_1C_1 = b$. Then apply the Pythagorean theorem to $\triangle A_1B_1C_1$. The length of the hypotenuse is then equal to $A_1B_1 = \sqrt{a^2 + b^2}$, which is, by assumption, equal to c. Hence, $A_1B_1 = c$. Using the side-side-side triangle congruence, it follows that $\triangle ABC \cong \triangle A_1B_1C_1$, and so $\triangle ABC$ is a right triangle with the right angle at C.

Generalizations and analogues

We shall now discuss some generalizations and analogues of the Pythagorean theorem in spaces of various dimensions. Consider first a box with side lengths a, b, c. Then the diagonal d of this box is given by $d^2 = a^2 + b^2 + c^2$. (See FIGURE 9.) To prove this, use the ordinary Pythagorean theorem twice.

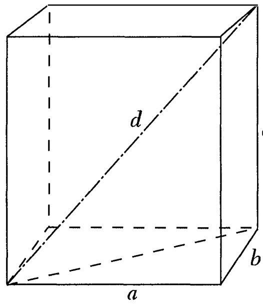


FIGURE 9
The diagonal of a box.

By induction on n, it follows easily that the diagonal d of an n-dimensional box with side lengths $a_1, a_2, \dots, a_n$ is given by $d^2 = \Sigma_{i=1}^n a_i^2$. In an appropriate sense this formula extends to infinite-dimensional spaces in functional analysis; see, e.g., **[15]**.

A direct space analogue of the Pythagorean theorem is as follows:

THEOREM 1. *Let OABC be a right tetrahedron, i.e., such that the edges OA, OB, and OC are mutually orthogonal at the vertex O. Then the square of the face opposite the vertex O is equal to the sum of the squares of the other three faces* (FIGURE 10). *Formally*,

$$\text{area}^2(\,\triangle ABC) = \text{area}^2(\,\triangle OAB) + \text{area}^2(\,\triangle OBC) + \text{area}^2(\,\triangle OAC).$$

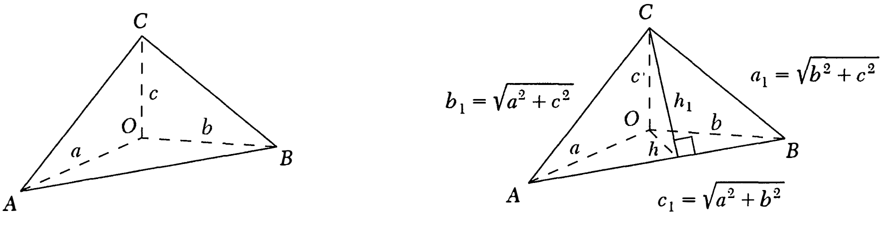


FIGURE 10
A space analogue.

Proof. The areas of the right triangles $\triangle OAB$, $\triangle OBC$, and $\triangle OAC$ are given by area($\triangle OAB) = ab/2$, area($\triangle OBC) = bc/2$, area($\triangle OAC) = ac/2$. The height h of $\triangle OAB$ from the vertex O is given by $h = ab/\sqrt{a^2+b^2}$, while the height h_1 of $\triangle ABC$ from the vertex C is given by $h_1^2 = c^2 + h^2 = c^2 + (a^2b^2)/(a^2+b^2)$. Hence

$$\begin{aligned}\text{area}^2(\,\triangle ABC) &= \frac{1}{4}c_1^2h_1^2 = \frac{1}{4}(a^2+b^2)\left(c^2 + \frac{a^2b^2}{a^2+b^2}\right) = \frac{1}{4}(a^2b^2 + b^2c^2 + c^2a^2)\\ &= \text{area}^2(\,\triangle OAB) + \text{area}^2(\,\triangle OBC) + \text{area}^2(\,\triangle OCA). \qquad \blacksquare\end{aligned}$$

If in the above theorem we call the triangles $\triangle OAB$, $\triangle OBC$, $\triangle OCA$ the "legs" and the triangle $\triangle ABC$ the "hypotenuse" of the right tetrahedron $OABC$, then the theorem says that the sum of the squares of the legs of a right tetrahedron is equal to the square of its hypotenuse.

Another proof of Theorem 1 can be given using *Heron's formula* (ca. 200 B.C., another consequence of the Pythagorean theorem):

If $\triangle ABC$ *has side lengths* a, b, c, *then*

$$\text{area}(\triangle ABC) = \frac{1}{4}\sqrt{4a^2b^2 - (c^2 - a^2 - b^2)^2}\,.$$

There is also an n-dimensional version of Theorem 1. It can be derived from Heron's formula for the volume V of an n-dimensional simplex $A_0 A_1 \dots A_n$ (with $a_{ij} = A_i A_j$):

$$2^n (n!)^2 V^2 = \left| \begin{vmatrix} 0 & a_{01}^2 & a_{02}^2 & \cdots & a_{0n}^2 & 1 \\ a_{10}^2 & 0 & a_{12}^2 & \cdots & a_{1n}^2 & 1 \\ \vdots & & & & & \vdots \\ a_{n0}^2 & \cdots & \cdots & a_{n,n-1}^2 & 0 & 1 \\ 1 & \cdots & \cdots & \cdots & 1 & 0 \end{vmatrix} \right|.$$

For inductive proofs of Heron's formula and Theorem 1, see **[9]**.

A common generalization of the last two 3-dimensional analogues of the Pythagorean theorem is as follows (see **[13]**):

THEOREM 2. *Let* $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k$ *be linearly independent vectors in the n-dimensional space* $\mathbb{R}^n$, $1 \le k \le n$. *Let V be the k-dimensional volume of the parallelepiped P spanned by vectors* $v_1, \dots, v_k$. *Then the square of V is equal to the sum of the squares of volumes of projections of P to all k-dimensional coordinate planes of* $\mathbb{R}^n$.

As an illustration, take $n = 3$ and $k = 2$, as shown in FIGURE 11.

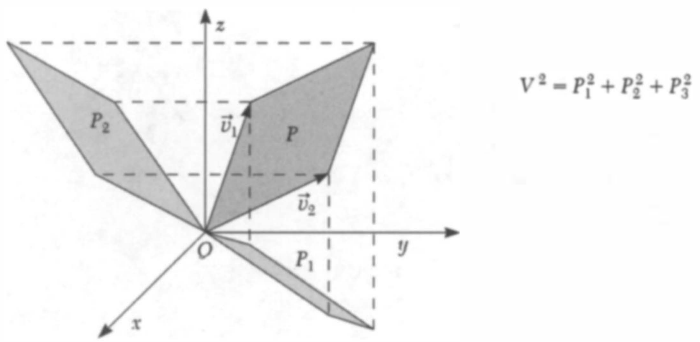


FIGURE 11
A higher-dimensional analogue.

Another space analogue of the Pythagorean theorem is of interest. The ordinary Pythagorean theorem, written as $a^2 - c^2 + b^2 = 0$, can be interpreted as follows: When traveling around the right triangle $\triangle ABC$ in the sense $A \to C \to B$, the alternating sum of the squares of the opposite sides is equal to zero.

Now let's go to space. Let $\triangle ABC$ be a right triangle with a right angle at vertex C. Erect a perpendicular line at A to the plane ABC and take a point D on this line. The resulting tetrahedron $ABCD$ is sometimes called an *orthoscheme* (see FIGURE 12).

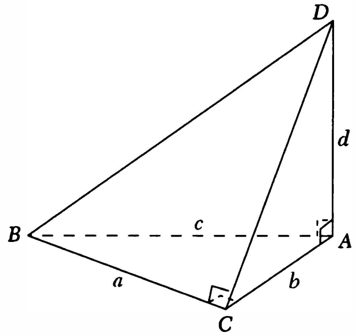


FIGURE 12
An orthoscheme.

Let F_A, F_B, F_C, F_D be the areas of the sides of this tetrahedron opposite to the vertices A, B, C, D, respectively. Note that $\triangle BCD$ is also a right triangle, so $F_B^2 = b^2d^2/4$, $F_C^2 = c^2d^2/4$, $F_D^2 = a^2b^2/4$, and $F_A^2 = a^2(b^2+d^2)/4$, and one checks easily that

$$F_A^2 - F_C^2 + F_B^2 - F_D^2 = 0.$$

This also can be interpreted as saying that if we travel around this tetrahedron in the sense $A \to C \to B \to D$, then the alternating sum of squares of the opposite sides is equal to zero. Here is a challenge for the reader: try to formulate and prove an n-dimensional version of this interpretation of the Pythagorean theorem.

We return to the plane for our next result:

THEOREM 3 [PAPPUS OF ALEXANDRIA, 4TH CENTURY]. *Let $\triangle ABC$ be an arbitrary triangle. Over the sides BC and AC construct (to the outside) parallelograms with areas P_1 and P_2, respectively, and let T be the intersection point of lines parallel to AC and BC* (FIGURE 13). *Over the third side AB construct the parallelogram whose other side is parallel to and of equal length as CT, and let its area be P_3. Then $P_1 + P_2 = P_3$.*

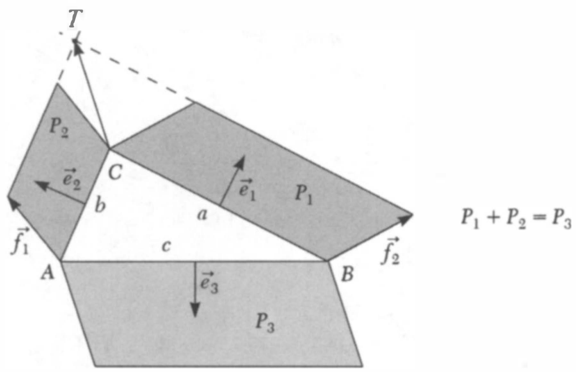


FIGURE 13
Pappus's theorem.

This can again be proved in a purely geometric style, but we offer a different proof. First we need the following simple fact. Let a be the length of the base of a parallelogram, $\vec{e}$ the unit vector perpendicular to that base, and $\vec{f}$ the edge-vector of the other side of this parallelogram. Then the area A of this parallelogram is given by $A = a\vec{e}\cdot\vec{f}$ (where $\vec{e}\cdot\vec{f}$ denotes the dot product). This is obvious, because $\vec{e}\cdot\vec{f}$ is just the height of the parallelogram. Next, we shall make use of the following well-known fact (due first to H. Minkowski, ca. 1900). Let $\vec{e}_1$, $\vec{e}_2$, and $\vec{e}_3$ be the outward-pointing unit normal vectors to the sides a, b, and c, respectively, of an arbitrary triangle $\triangle ABC$. Then $a\vec{e}_1 + b\vec{e}_2 + c\vec{e}_3 = \vec{0}$. To see this, consider the sides a, b, c as vectors, so that $\vec{a} + \vec{b} + \vec{c} = \vec{0}$. Translate the vectors $\vec{a}, \vec{b}, \vec{c}$ to a common point O, and rotate them about O by $90°$. Then they are transformed respectively to $a\vec{e}_1, b\vec{e}_2, c\vec{e}_3$, and since $\vec{a} + \vec{b} + \vec{c} = \vec{0}$, it follows that $a\vec{e}_1 + b\vec{e}_2 + c\vec{e}_3 = \vec{0}$. (In fact, Minkowski's theorem is equivalent to the Pythagorean theorem, since scalar squaring of the relation $-a\vec{e}_1 = b\vec{e}_2 + c\vec{e}_3$ yields the law of cosines, which is equivalent to the Pythagorean theorem, and conversely.)

Proof of Theorem 3. By the above discussion we have $P_1 = a\vec{e}_1 \cdot \vec{f}_1$, $P_2 = b\vec{e}_2 \cdot \vec{f}_2$, and $P_3 = -c\vec{e}_3 \cdot \overrightarrow{CT}$. On the other hand, clearly, $\overrightarrow{CT} \cdot \vec{e}_1 = \vec{e}_1 \cdot \vec{f}_1$ (= height of P_1), and similarly $\overrightarrow{CT} \cdot \vec{e}_2 = \vec{e}_2 \cdot \vec{f}_2$. Hence,

$$P_3 = -c\vec{e}_3 \cdot \overrightarrow{CT} = \left(a\vec{e}_1 + b\vec{e}_2\right) \cdot \overrightarrow{CT} = a\vec{e}_1 \cdot \overrightarrow{CT} + b\vec{e}_2 \cdot \overrightarrow{CT} = a\vec{e}_1 \cdot \vec{f}_1 + b\vec{e}_2 \cdot \vec{f}_2 = P_1 + P_2.$$

∎

Note that if $\triangle ABC$ is a right triangle and P_1 are P_2 squares, then P_3 is also a square, and we get the Pythagorean theorem as a special case.

In the same manner we can prove the space version of Pappus's theorem. First note that the volume V of a prism with base area F, edge-vector $\vec{f}$, and $\vec{e}$ the unit normal vector to the base is given by $V = F\vec{e} \cdot \vec{f}$ (Figure 14a).

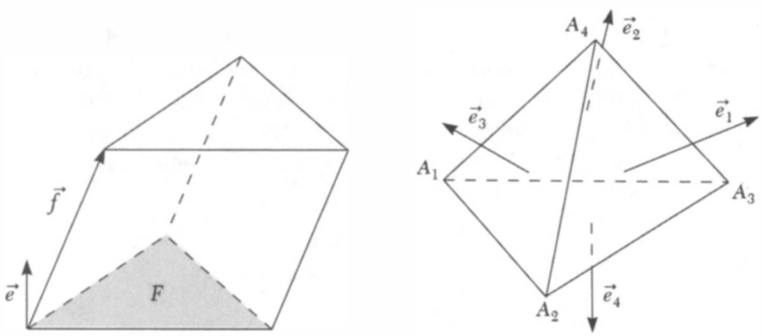


FIGURE 14
Pappus's theorem in space.

Let $A_1 A_2 A_3 A_4$ be any tetrahedron, F_i the area of the side opposite to A_i, and $\vec{e}_i$ the unit (outward) normal vector to that side, $i = 1, 2, 3, 4$ (Figure 14b). Then, again by Minkowski,

$$F_1\vec{e}_1 + F_2\vec{e}_2 + F_3\vec{e}_3 + F_4\vec{e}_4 = \vec{0}.$$

(In fact, a similar formula holds for any convex polytope, and this is not hard to prove using volumes; see, e.g., **[5]**.)

Over any three sides of our tetrahedron, with areas F_i, $i = 1, 2, 3$, erect to the outside triangular prisms with edge-vectors $\vec{f}_i$, $i = 1, 2, 3$, respectively. Let T be the intersection point of the planes containing the other (parallel) bases. Over the fourth side with area F_4 erect the prism with edge-vector $\overrightarrow{TA_4}$. Denote by V_i, $i = 1, 2, 3, 4$ the volumes of the resulting prisms. Then $V_i = F_i\vec{e}_i \cdot \vec{f}_i$, $i = 1, 2, 3$ and $V_4 = F_4\vec{e}_4 \cdot \overrightarrow{TA_4}$

$= -F_4\vec{e}_4 \cdot \vec{A_4T}$. It is clear that $\vec{A_4T} \cdot \vec{e}_i = \vec{f}_i \cdot \vec{e}_i$ (= height of prism), $i = 1, 2, 3$. Hence we have

$$\begin{aligned} V_4 &= -F_4\vec{e}_4 \cdot \vec{A_4T} = \left(F_1\vec{e}_1 + F_2\vec{e}_2 + F_3\vec{e}_3\right) \cdot \vec{A_4T} \\ &= F_1\vec{e}_1 \cdot \vec{A_4T} + F_2\vec{e}_2 \cdot \vec{A_4T} + F_3\vec{e}_3 \cdot \vec{A_4T} \\ &= F_1\vec{e}_1 \cdot \vec{f}_1 + F_2\vec{e}_2 \cdot \vec{f}_2 + F_3\vec{e}_3 \cdot \vec{f}_3 \\ &= V_1 + V_2 + V_3. \end{aligned}$$

This proves the 3-dimensional version of Pappus's theorem. With obvious changes, essentially the same proof applies to an n-dimensional simplex, for any n. Note that in the above proof instead of prisms we can take pyramids, since if a prism and a pyramid have equal area of their bases and equal altitudes, then the volume of such a pyramid is $1/3$ of the volume of the prism.

Pappus's theorem can be described as an "affine" version of the Pythagorean theorem. The point is that an affine map preserves the ratio of volumes. We leave details to the reader (cf. [4]).

Now we turn to a different kind of generalization of the Pythagorean theorem. First recall proof 3 above (by Chou-pei Suan-ching) of the Pythagorean theorem. If F_x denotes the area of a square of side x, and F is the area of right triangle $\triangle ABC$, then Figure 6 shows that

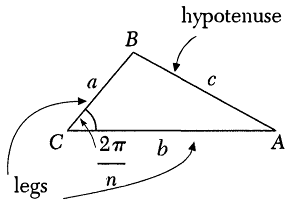
$$F_{a+b} = F_c + 4F. \qquad (*)$$

From Heron's formula it is easy to see that this relation is, in fact, equivalent to the Pythagorean theorem. Therefore, $F_a + F_b = F_c$ if and only if $F_{a+b} = F_c + 4F$, and the Pythagorean theorem can be rephrased as follows:

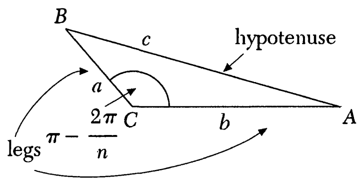
The square over the sum of the legs of a right triangle is equal to the sum of the square over the hypotenuse plus four areas of the triangle.

Now let $\triangle ABC$ be a triangle with (instead of a right angle) an angle $\gamma = 2\pi/n$ or $\gamma = \pi - 2\pi/n$, where $n \geq 3$ is any integer. As in the Pythagorean theorem we erected squares (over a side or sum of sides), now we erect regular n-gons. For $n = 4$, our construction will reduce to the Pythagorean theorem in the form $(*)$.

To this end, let $\triangle ABC$ be a triangle with an angle $\gamma = 2\pi/n$, $n \geq 3$. We call such a triangle a *fraction n-triangle*. In the case of the complementary angle $\gamma = \pi - 2\pi/n$, we call such a triangle a *complementary fraction n-triangle*. The sides a and b adjacent to γ we call *legs*, and the side c opposite to γ *the hypotenuse* of such a triangle. To any such triangle is naturally associated the *complementary triangle* and *complementary hypotenuse* c' (Figure 15 and Figure 16).



(a) fraction n-triangle



(b) complementary fraction n-triangle

FIGURE 15
Complementary triangles.

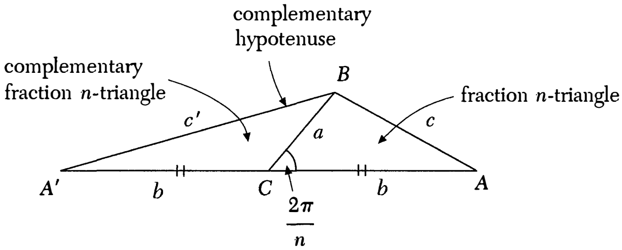


FIGURE 16
Joining complementary triangles.

The length of the complementary hypotenuse c' is given by

$$c'^2 = 2(a^2 + b^2) - c^2.$$

(This holds in any triangle and can be proved easily by applying the law of cosines to $\triangle ABC$ and $\triangle A'BC$, but it can also be done without trigonometry.)

Now construct over the sum of legs $(a + b)$ of a fraction n-triangle a regular n-gon. The internal angle of this n-gon is equal to $\pi - 2\pi/n$. On every edge take the point such that lengths a and b alternate and join all consecutive points by an edge as in Figure 17a for $n = 3$ and 17b for $n = 6$.

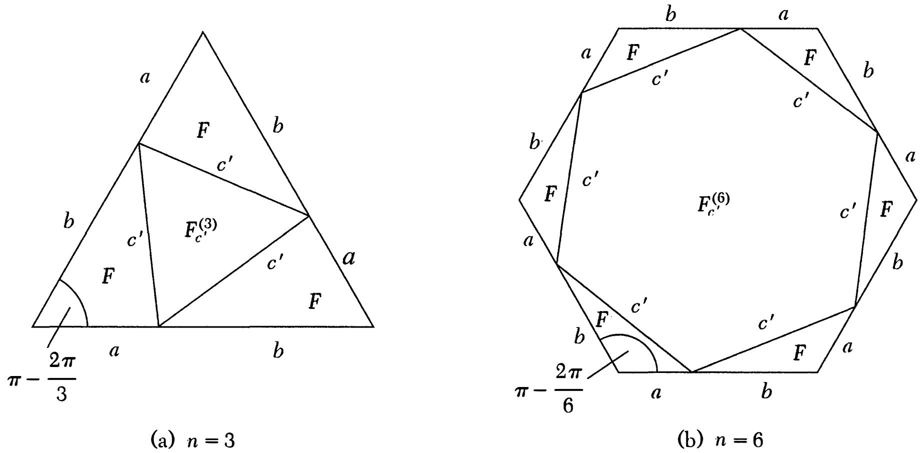


FIGURE 17
Constructing n-gons.

By congruency (side-angle-side) we see that at the vertices of the n-gon we have n complementary triangles, and so the length of every consecutive joining edge is equal to the complementary hypotenuse c' of the original triangle. Denote by $F_x^{(n)}$ the area of a regular n-gon with side length x, and by F the area of our fraction n-triangle. Then, since the area of the complementary triangle is equal to the area of the original triangle, we see from Figure 17 that the following relation holds:

$$F_{a+b}^{(n)} = F_{c'}^{(n)} + nF. \qquad (**)$$

If, instead, we start with a complementary fraction n-triangle (i.e., $\gamma = \pi - 2\pi/n$), then by the same construction we see that the length of every edge is equal to c, the hypotenuse of the triangle, so

$$F_{a+b}^{(n)} = F_{c}^{(n)} + nF. \qquad (***)$$

For $n = 4$, both relations $(**)$ and $(***)$ reduce to the Pythagorean theorem in the form $(*)$, because then, and only then, $c = c'$. Formulas $(**)$ and $(***)$ are summarized in the following theorem (see **[17]**):

THEOREM 4. *Let $\triangle ABC$ be a fraction n-triangle (resp. complementary fraction n-triangle). Then the area of a regular n-gon over the sum of its legs is equal to the sum of n areas of the triangle and the area of a regular n-gon over its complementary hypotenuse (resp. hypotenuse).*

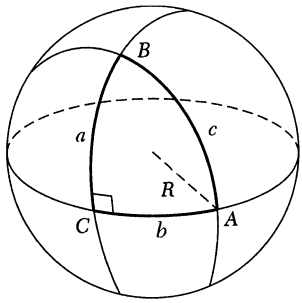
Thus we obtain infinitely many analogues of the Pythagorean theorem, one for every n. A special triangular analogue was considered in [**7**].

It is interesting to note that if we fix a and b and let $n \to \infty$, then, since $2F = ab\sin(2\pi/n)$, from $(**)$ or $(***)$ we get

$$\lim_{n\to\infty}\left(F_{a+b}^{(n)} - F_c^{(n)}\right) = \lim_{n\to\infty} nF = ab\pi.$$

This is the area of an ellipse with half-axes a and b. It would be interesting to find a genuine geometric meaning of this fact.

As usual, a number of interesting questions arise: Are there interesting space analogues of the above construction? What would analogues be in spherical or hyperbolic geometries? Recall that the Pythagorean theorem in spherical geometry on a sphere of radius R has the form $\cos\frac{c}{R} = \cos\frac{a}{R}\cdot\cos\frac{b}{R}$ (see FIGURE 18).



$$\cos\frac{c}{R} = \cos\frac{a}{R}\cdot\cos\frac{b}{R}$$

FIGURE 18
A spherical triangle.

From the power series expansion $\cos x = 1 - x^2/2! + x^4/4! - \cdots$, we get

$$1 - \frac{1}{2!}\left(\frac{c}{R}\right)^2 + \frac{1}{4!}\left(\frac{c}{R}\right)^4 - \cdots$$
$$= \left(1 - \frac{1}{2!}\left(\frac{a}{R}\right)^2 + \frac{1}{4!}\left(\frac{a}{R}\right)^4 - \cdots\right)\left(1 - \frac{1}{2!}\left(\frac{b}{R}\right)^2 + \frac{1}{4!}\left(\frac{b}{R}\right)^4 - \cdots\right).$$

After some manipulations we obtain

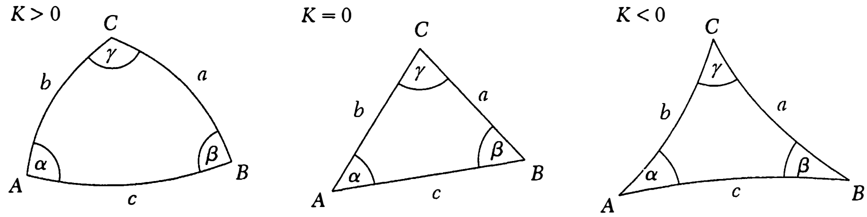
$$c^2 - \frac{1}{12}\frac{c^4}{R^2} + \cdots = a^2 + b^2 - \frac{a^2b^2}{2R^2} - \frac{a^4}{12R^2} - \frac{b^4}{12R^2} + \cdots.$$

If the sides of the triangle are fixed, and the center of the sphere moves farther and farther away, i.e., if $R \to \infty$, the above equation gives, in the limit, the ordinary Pythagorean theorem $c^2 = a^2 + b^2$. In this sense the sphere geometry is closer to Euclidean geometry as its radius becomes greater and greater. Thus, spherical geometry can be viewed as locally Euclidean.

In hyperbolic geometry, the Pythagorean theorem has the form $\cosh c = \cosh a \cdot \cosh b$ (see, e.g., [**14**]). In a similar fashion as above, from the power series expansion $\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots$ and from $\cosh c = \cosh a \cdot \cosh b$, we get $c^2 \approx a^2 + b^2$ for small a and b. This shows that hyperbolic geometry also behaves locally like Euclidean geometry.

How should generalizations and analogues of the Pythagorean theorem look in higher-dimensional spherical or hyperbolic geometry? Just to get an idea, see FIGURE

19, where we exhibit the difference between Euclidean and non-Euclidean geometries for sufficiently small triangles, depending on the (sectional) curvature K of the (Riemannian) manifold in which a triangle sits.



$\alpha + \beta + \gamma > \pi$	$\alpha + \beta + \gamma = \pi$	$\alpha + \beta + \gamma < \pi$
$c^2 < a^2 + b^2 - 2ab \cos \gamma$	$c^2 = a^2 + b^2 - 2ab \cos \gamma$	$c^2 > a^2 + b^2 - 2ab \cos \gamma$
(Spherical)	(Euclidean)	(Hyperbolic)

FIGURE 19
Triangles in various geometries.

So the real question is how to modify the law of cosines (or the Pythagorean theorem) in order to obtain a uniform picture of all three geometries, and hence to obtain, in a sense, a classification of manifolds.

These and similar questions were considered long ago, beginning with Euclid and continued by L. Euler (1707–1783), B. Riemann (1826–1866), A. Einstein (1879–1955) and others, up to our own contemporaries J. Milnor, W. Thurston, M. Gromov, and E. Witten, to mention just a few. And the whole story started with Pythagoras some 2500 years ago.

REFERENCES

1. W. S. Anglin, J. Lambek, *The Heritage of Thales*, Springer-Verlag, New York, NY, 1995.
2. V. N. Beryozin, The good old Pythagorean theorem, *Quantum*, Jan./Feb. 1994, 25–28, 60–61.
3. B. Bollobás, To prove and conjecture: Paul Erdős and his mathematics, *Amer. Math. Monthly* 105 (1998), 209–237.
4. H. S. M. Coxeter, *Introduction to Geometry*, John Wiley & Sons, New York, NY, 1989.
5. B. Grünbaum, *Convex Polytopes*, John Wiley & Sons, New York, NY, 1967.
6. K. Ireland, M. Rosen, *A Classical Introduction to Modern Number Theory*, Springer-Verlag, New York, NY, 1982.
7. J. Kovačević, D. Veljan, A triangular analogue of the Pythagorean theorem, *The Math. Gazette*, 80 (1996), 550–554.
8. W. Lietzmann, *Der Pythagorische Lehrsatz mit einem Ausblick auf das Fermatsche Problem*, Teubner-Verlag, Stuttgart, Germany, 1966.
9. S.-Y. T. Lin, Y.-F. Lin, The n-dimensional Pythagorean theorem, *Linear and Multilinear Algebra*, 26 (1990), 9–13.
10. B. Mazur, Questions about powers of numbers, *AMS Notices*, 47 (2000), 195–202.
11. R. B. Nelson, *Proofs Without Words*, Mathematical Association of America, Washington, DC, 1993.
12. J. Pach, P. K. Agarwal, *Combinatorial Geometry*, John Wiley & Sons, New York, NY, 1995.
13. G. J. Porter, k-volume in $\mathbb{R}^n$ and the generalized Pythagorean theorem, *Amer. Math. Monthly*, 103 (1996), 252–256.
14. J. G. Ratcliffe, *Foundations of Hyperbolic Manifolds*, Springer-Verlag, New York, NY, 1994.
15. W. Rudin, *Functional Analysis*, McGraw Hill, New York, NY, 1991.
16. J. Stillwell, *Mathematics and Its History*, Springer-Verlag, New York, NY, 1989.
17. D. Veljan, An analogue of the Pythagorean theorem with regular n-gons instead of squares, *Elemente der Math.*, 51 (1996), 156–160.
18. A. Wiles, Twenty years of number theory, in *Mathematics: Frontiers and Perspectives*, (Eds. V. Arnold, *et al*), IMU–AMS, 2000, 329–343.

The Gyro-Structure of the Complex Unit Disk

MICHAEL K. KINYON
Indiana University
South Bend, IN 46634

ABRAHAM A. UNGAR
North Dakota State University
Fargo, ND 58105

Introduction

The unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ in the complex plane is rich in mathematical structure. In this paper we will view this set from a new perspective by illuminating the link between the geometry of $\mathbb{D}$ and a natural, but not well-known, algebraic structure which $\mathbb{D}$ possesses. As motivation, we first restrict ourselves to the interval $(-1,1)$ on the real line $\mathbb{R}$. In units where the speed of light is 1, the relativistic addition law for parallel velocities is

$$x \oplus y = \frac{x+y}{1+xy} \tag{1}$$

for $x, y \in (-1,1)$ **[2]**, **[3]**. We also define a multiplication by real scalars by

$$r \odot x = \frac{(1+x)^r - (1-x)^r}{(1+x)^r + (1-x)^r} \tag{2}$$

for $x \in (-1,1)$, $r \in \mathbb{R}$. It is easily checked that the structure $((-1,1), \oplus, \odot)$ is an "exotic" example of a real vector space (**[1]**, **[2]**, **[7]**).

Generalizations of (1) and (2) to $\mathbb{D}$ are given by, respectively,

$$x \oplus y = \frac{x+y}{1+\bar{x}y} \tag{3}$$

$$r \odot x = \frac{(1+|x|)^r - (1-|x|)^r}{(1+|x|)^r + (1-|x|)^r} \frac{x}{|x|} \tag{4}$$

for $x, y \in \mathbb{D}$, $r \in \mathbb{R}$, $\bar{x}$ being the complex conjugate of x. (Here $x \neq 0$ in (4); we also define $r \odot 0 = 0$.) The structure $(\mathbb{D}, \oplus, \odot)$ is *not* a vector space. For one thing, in addition to being obviously noncommutative, the operation $\oplus$ in (3) also turns out to be nonassociative (**[1]**, **[16]**). Instead of being a vector space, $(\mathbb{D}, \oplus, \odot)$ is an example of a more general structure called a *gyrovector space*. As we will explain later, the "gyro" terminology stems from the historical development of these structures in connection with the phenomenon of Thomas precession in relativistic physics.

As is well-known, $\mathbb{D}$ serves as the *Poincaré model* for hyperbolic geometry. Just as vector space structure provides a useful algebraic and analytic foundation for Euclidean geometry, it turns out that so too does the gyrovector space structure (or *gyro-structure* for short) of $(\mathbb{D}, \oplus, \odot)$ turn out to provide a similar foundation for hyperbolic geometry. It is these matters we will survey in this paper. Although our discussion will remain focused on $\mathbb{D}$, the ideas involved are more general. We view this paper as an invitation to the reader to explore further.

Gyrovector spaces

To motivate our definition, we make a few preliminary remarks about the algebraic structure of $(\mathbb{D}, \oplus, \odot)$. The element $0 \in \mathbb{D}$ is clearly an identity element for $\oplus$, and for each $x \in \mathbb{D}$, $-x \in \mathbb{D}$ is an inverse of x. Observe also that for each $a \in \mathbb{C}$ with $|a| = 1$, we have $a(x \oplus y) = ax \oplus ay$ and $a(r \odot x) = r \odot ax$, $x, y \in \mathbb{D}$, $r \in \mathbb{R}$. In particular, $\{a \in \mathbb{C} : |a| = 1\}$ is a subgroup of $\mathrm{Aut}(\mathbb{D}, \oplus)$, the group of automorphisms of the structure $(\mathbb{D}, \oplus)$.

We have noted that $\oplus$ is both noncommutative and nonassociative. We can "repair" the noncommutativity as follows. For $x, y \in \mathbb{D}$ we define

$$\mathrm{gyr}[x, y] = \frac{1 + x\bar{y}}{1 + \bar{x}y}, \tag{5}$$

calling it the *gyration* determined by x and y; note that $|\mathrm{gyr}[x, y]| = 1$. Instead of commutativity, $(\mathbb{D}, \oplus)$ satisfies the following *gyrocommutative law*:

$$x \oplus y = \mathrm{gyr}[x, y](y \oplus x), \tag{6}$$

$x, y \in \mathbb{D}$. Strikingly, the gyration that we introduce to repair the noncommutativity of $\oplus$ also repairs its nonassociativity, and possesses other elegant properties as well. This leads us to our main definition. We draw the reader's attention to the formal similarities between the following and the usual definition of a vector space.

DEFINITION. A gyrovector space $(P, \oplus)$ is a nonempty set P together with a binary operation $\oplus : P \times P \to P$ and an operation $\odot : \mathbb{R} \to P \to P$ satisfying the following axioms.

(G1) *There exists* $0 \in P$ *such that for all* $x \in P$, $0 \oplus x = x \oplus 0 = x$. (Identity)

(G2) *For each* $x \in P$, *there exists* $-x \in P$ *such that* $-x \oplus x = x \oplus -x = 0$. (Inverses)

For each $x, y \in P$, the mapping $\mathrm{gyr}[x, y] : P \to P$ defined by

$$\mathrm{gyr}[x, y]z = -(x \oplus y) \oplus (x \oplus (y \oplus z)) \tag{7}$$

for $z \in P$ satisfies the following properties. For all $x, y, z \in P$, $r, r_1, r_2 \in \mathbb{R}$,

(G3)	$x \oplus (y \oplus z) = (x \oplus y) \oplus \mathrm{gyr}[x, y]z$	(Left gyroassociative law)
(G4)	$\mathrm{gyr}[x, y] \in \mathrm{Aut}(P, \oplus)$	(Gyroautomorphism)
(G5)	$\mathrm{gyr}[x \oplus y, y] = \mathrm{gyr}[x, y]$	(Left Loop property)
(G6)	$x \oplus y = \mathrm{gyr}[x, y](y \oplus x)$	(Gyrocommutative law)
(G7)	$(r_1 + r_2) \odot x = (r_1 \odot x) \oplus (r_2 \odot x)$	
(G8)	$(r_1 r_2) \odot x = r_1 \odot (r_2 \odot x)$	
(G9)	$1 \odot x = x$	
(G10)	$\mathrm{gyr}[x, y](r \odot z) = r \odot \mathrm{gyr}[x, y]z$	
(G11)	$\mathrm{gyr}[r_1 \odot x, r_2 \odot x] = \mathrm{id}_P$	

The mapping $\mathrm{gyr}[x, y] : P \to P$ given by (7) is called the *gyroautomorphism* generated by $x, y \in P$, and its action is called the *gyration* generated by $x, y \in P$. The structure $(\mathbb{D}, \oplus, \odot)$ itself is a gyrovector space, as can be shown by direct computations, and the notion of a gyration as an action is consistent with our identification in $(\mathbb{D}, \oplus, \odot)$ of $\mathrm{gyr}[x, y]$ with a unimodular complex number.

A *gyrogroup* $(P, \oplus)$ is a structure satisfying (G1) through (G6), and a *gyrocommutative gyrogroup* also satisfies (G7). Any group is a gyrogroup with $\mathrm{gyr}[x, y] \equiv \mathrm{id}$ for all x, y. Conversely, a gyrogroup in which every gyration is trivial is a group. Similar remarks apply to abelian groups and gyrocommutative gyrogroups, but the same cannot be said for gyrovector spaces. While a vector space is obviously a gyrovector space, there exist abelian groups with an action of $\mathbb{R}$ satisfying (G7) through (G9) (and trivially, (G10) and (G11)) but which are not vector spaces [**11**]. Thus a gyrovector space with trivial gyrations is not necessarily a vector space. What is missing is an as yet unknown *gyrodistributive law* connecting the operations $\odot$ and $\oplus$. In $(\mathbb{D}, \oplus, \odot)$ and other gyrovector spaces, the expressions $r \odot (x \oplus y)$ and $(r \odot x) \oplus (r \odot y)$ are not, in general, equal.

In a brief digression, we now explain our "gyroterminology." The first structure to be recognized as a gyrocommutative gyrogroup, and later, as a gyrovector space, is what we now call the *Einstein* gyrogroup $(\mathbb{B}, \oplus)$, where $\mathbb{B} = \{\mathbf{v} \in \mathbb{R}^3 : \|\mathbf{v}\| < 1\}$ is the set of all relativistically admissible velocities in $\mathbb{R}^3$ (in units where the speed of light is 1), and the operation $\oplus$ is the relativistic addition law for velocity vectors derived by Einstein in 1905 [**3**]. This is the generalization of (1) to not necessarily parallel vectors; see, for example, [**13**],[**15**] for the exact formulas. In this context, a gyration is a rotation known as *Thomas precession*. Its significance for nonparallel velocities was first discovered in the theory of electron spin by L. H. Thomas in late 1925; see [**12**] for details of the story. For a macroscopic example, if two gyroscopes in the same initial state diverge and are then brought together for comparison, their axes differ as a result of undergoing different Thomas precessions [**15**]. What we now call the gyro-structure of the Einstein gyrogroup was first observed in 1988 in [**13**]. "Gyroterminology" itself began in [**15**] from the recognition that the *physically* significant Thomas precession and its abstraction into Thomas gyration was the key to the *algebraic* generalizations of group theoretic concepts—such as gyroassociativity, gyrocommutativity, and so on—which make up the structure of $(\mathbb{B}, \oplus)$. (See also reference 36 in [**15**].)

Our purpose here is not to analyze the algebraic structure $(\mathbb{D}, \oplus, \odot)$ in detail, but rather to touch on those aspects which are relevant to geometry. A detailed axiomatic study of gyrogroups can be found in [**18**]. Perhaps the most technical axiom is the loop property (G5), which is the key in proving the following [**18**].

(L) Given $a, b \in P$, the unique solution of $a \oplus x = b$ is $x = -a \oplus b$.

(R) Given $a, b \in P$, the unique solution of $x \oplus a = b$ is $x = b \ominus \mathrm{gyr}[b, a]a$.

Here $x \ominus y$ abbreviates $x \oplus (-y)$. Properties (L) and (R) imply that every gyrogroup is a *loop* [**9**], which is where (G5) gets its name. (For more on the relationship between gyrogroups and other loops, see [**18**],[**19**] and the references therein.)

The solution given in (R) motivates the definition of a *dual operation* $\boxplus$ in a gyrogroup $(P, \oplus)$ by the equation

$$x \boxplus y = x \oplus \mathrm{gyr}[x, -y]y \tag{8}$$

for $x, y \in P$. Thus the solution of $x \oplus a = b$ in (R) can be written as $x = b \boxplus (-a) = b \boxminus a$, so that $(b \boxminus a) \boxplus a = b$. Of course, if a gyrogroup $(P, \oplus)$ is a group, then the dual operation $\boxplus$ agrees with the original operation $\oplus$. The explicit form of the dual operation in $(\mathbb{D}, \oplus)$ is

$$x \boxplus y = \frac{\left(1 - |y|^2\right)x + \left(1 - |x|^2\right)y}{1 - |x|^2|y|^2} \tag{9}$$

for $x, y \in \mathbb{D}$. This operation is clearly commutative, and in fact, a gyrogroup $(P, \oplus)$ is gyrocommutative if and only if its dual $(P, \boxplus)$ is commutative [**18**].

With the additional notation, we can summarize properties (L) and (R) by the following *cancellation laws*:

$$-x \oplus (x \oplus y) = y \tag{10}$$
$$(x \oplus y) \boxminus y = x \tag{11}$$
$$(x \boxplus y) \ominus y = x \tag{12}$$

for $x, y \in P$. We will use these in our study of the geometry of $\mathbb{D}$. Equation (10) is equivalent to (L), which is also equivalent to

$$\mathrm{gyr}[-x, x] = \mathrm{id} \tag{13}$$

for $x, y \in P$. Equation (11) is a restatement of (R) using (8); a proof of (12) can be found in **[18]**.

Norms and metrics

We now define the *Poincaré norm* $\|\cdot\|: \mathbb{D} \to [0, 1)$ by

$$\|x\| = |x| \tag{14}$$

for $x \in \mathbb{D}$. This norm satisfies the following properties **[16]**: for all $x, y, z \in \mathbb{D}$, $r \in \mathbb{R}$,

$$\|x\| \geq 0,\ \|x\| = 0 \Rightarrow x = 0 \tag{P1}$$
$$\|r \odot x\| = |r| \odot \|x\| \tag{P2}$$
$$\|x \oplus y\| \leq \|x\| \oplus \|y\| \tag{P3}$$
$$\|\mathrm{gyr}[x, y]z\| = \|z\| \tag{P4}$$

(The right hand sides of (P2) and (P3) are computed in $((-1, 1), \oplus, \odot)$.) Note the formal similarities between the properties (Pj) and properties of norms in vector spaces. In addition, the useful *scaling identity*

$$\frac{r \odot x}{\|r \odot x\|} = \frac{x}{\|x\|} \tag{15}$$

holds for all $r > 0$, $x \in \mathbb{D}$.

The identity $\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2x \cdot y$ in a Euclidean vector space, where $\cdot$ denotes the dot product, has an interesting analog for the Poincaré norm in $\mathbb{D}$, namely

$$\|x \oplus y\|^2 = \|x\|^2 \oplus \|y\|^2 \oplus \frac{\frac{1}{2}(2 \odot x) \cdot (2 \odot y)}{1 + \frac{1}{2}(2 \odot x) \cdot (2 \odot y)} \tag{16}$$

for $x, y \in \mathbb{D}$ **[19]**. Here the dot product $\cdot$ is given by $a \cdot b = \mathrm{Re}(a\bar{b})$, where $\mathrm{Re}(z)$ denotes the real part of z. If x and y are orthogonal, then (16) specializes to the *Hyperbolic Pythagorean Theorem*:

$$\|x \oplus y\|^2 = \|x\|^2 \oplus \|y\|^2. \tag{17}$$

In addition, since $((-1, 1), \oplus)$ is an abelian group, (16) implies the *Hyperbolic Polarization Identity*

$$\|x \oplus y\|^2 \ominus \|x \ominus y\|^2 = (2 \odot x) \cdot (2 \odot y), \tag{18}$$

$x, y \in \mathbb{D}$, an obvious analog of the polarization identity $\|x + y\|^2 - \|x - y\|^2 = 4x \cdot y$ in a Euclidean vector space.

From the Poincaré norm, we define the *Poincaré metric* [**4**] by

$$d_{\oplus}(x,y)=\|x\ominus y\|, \tag{19}$$

$x, y\in\mathbb{D}$. The nondegeneracy condition $d_{\oplus}(x,y)=0\Rightarrow x=y$ follows from (P1), and the symmetry condition $d_{\oplus}(x,y)=d_{\oplus}(y,x)$ follows from gyrocommutativity (G6) and (P4). The Poincaré metric is *gyroautomorphism invariant* and (left) *gyrotranslation invariant*, which are the properties

$$d_{\oplus}(z,w)=d_{\oplus}(\mathrm{gyr}[x,y]z,\mathrm{gyr}[x,y]w) \tag{20}$$

$$d_{\oplus}(y,z)=d_{\oplus}(x\oplus y,x\oplus z) \tag{21}$$

respectively, for $x,y,z,w\in\mathbb{D}$. Equation (20) follows from (P4), while (21) follows from (P4) and the following identity, which holds in any gyrocommutative gyrogroup [**16**]:

$$-(x\oplus a)\oplus(x\oplus b)=\mathrm{gyr}[x,a](-a\oplus b), \tag{22}$$

$x, a, b\in\mathbb{D}$. Finally, we have the triangle inequality

$$d_{\oplus}(x,z)\le d_{\oplus}(x,y)\oplus d_{\oplus}(y,z)\le d_{\oplus}(x,y)+d_{\oplus}(y,z) \tag{23}$$

for $x,y,z\in\mathbb{D}$, which follows from (21), (P3), (G6), and (P4) [**16**].

Recalling that the gyrogroup $(\mathbb{D},\oplus)$ has the dual operation $\boxplus$ given by (9), we infer that $\mathbb{D}$ has a *dual metric* defined by

$$d_{\boxplus}(x,y)=\|x\boxminus y\|, \tag{24}$$

$x, y\in\mathbb{D}$. The properties of $d_{\boxplus}$ are less transparent than those of $d_{\oplus}$. While $d_{\boxplus}$ is gyroautomorphism invariant, it is not invariant under gyrotranslations, using either $\oplus$ or $\boxplus$. It does satisfy an interesting "gyrotriangle inequality"

$$d_{\boxplus}(x,y)\boxplus d_{\boxplus}(y,z)\ge d_{\boxplus}(x,\mathrm{gyr}[x\boxminus y,y\boxminus z]z) \tag{25}$$

for $x,y,z\in\mathbb{D}$ [**18**], but this does not by itself imply the usual triangle inequality, which must be proven separately. Indeed, the geometric meaning of (25) has yet to be discovered.

Gyrogeometry

We now turn to the geometry of the gyrovector space $(\mathbb{D},\oplus,\odot)$. We begin considering those classes of curves in $\mathbb{D}$ which are analogs to lines in vector spaces. If $a, b\in\mathbb{C}$ are given points, then the formula

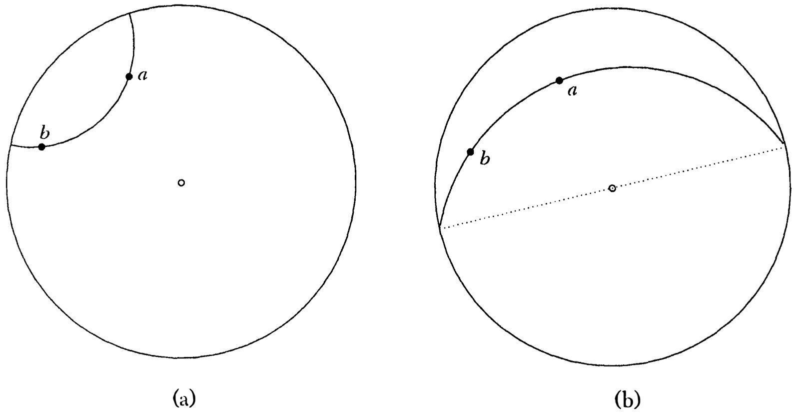
$$r(t)=a+t(-a+b)=t(b-a)+a, \tag{26}$$

$t\in\mathbb{R}$, parametrizes the unique line passing through a (when $t=0$) and b (when $t=1$). This formulation stresses that different rearrangements of the parametrization lead to the same curve. In $(\mathbb{D},\oplus,\odot)$, there are *three* distinct analogs of (26) corresponding to the three cancellation laws (10), (11) and (12). We will consider two of these here—the two we understand best—and save a few remarks about the third for the end of the paper. Fix $a, b\in\mathbb{D}$ and consider the curves given by

$$l(t)=a\oplus t\odot(-a\oplus b) \tag{27}$$

$$m(t)=t\odot(b\boxminus a)\oplus a \tag{28}$$

for $t \in \mathbb{R}$. We have $l(0) = m(0) = a$. Using (G9) and (10) in (27), and (G9) and (11) in (28), we see that $l(1) = m(1) = b$. We will refer to the curve (27) as being a *gyroline* through a and b; see FIGURE 1(a). We will refer to the curve (27) as being a *dual gyroline* through a and b; see FIGURE 1(b). For given $a, b \in \mathbb{D}$, it is easy to show using gyro-structure that there is a unique gyroline and a unique dual gyroline passing through a and b. These algebraic proofs generalize to other settings. However, in $\mathbb{D}$, just as enlightening is the geometric observation that as curves in $\mathbb{C}$, gyrolines and dual gyrolines turn out to be special types of circular arcs through a and b; we will discuss this further below. From this the uniqueness follows readily.



(a) (b)

FIGURE 1

(a) The gyroline $l(t) = a \oplus t \odot (-a \oplus b)$
(b) The dual gyroline $m(t) = t \odot (b \boxminus a) \oplus a$

Interestingly, one can characterize when finite subsets of points lie on a common gyroline or dual gyroline with just the gyrations. In particular, we have the following pair of *gyrotransitive laws*.

(T1) A set $\{a_1, a_2, \dots, a_n\} \subset \mathbb{D}$ lies on a common gyroline if and only if

$$\mathrm{gyr}[a_1, -a_2]\mathrm{gyr}[a_2, -a_3] \cdots \mathrm{gyr}[a_{n-1}, -a_n] = \mathrm{gyr}[a_1, -a_n].$$

(T2) A set $\{b_1, b_2, \dots, b_n\} \subset \mathbb{D}$ lies on a common dual gyroline if and only if

$$\mathrm{gyr}[b_1, b_2]\mathrm{gyr}[b_2, b_3] \cdots \mathrm{gyr}[b_{n-1}, b_n] = \mathrm{gyr}[b_1, b_n].$$

See **[17]** for (T2); the proof of (T1) is similar.

We have the following "betweenness" properties of gyrolines and dual gyrolines **[19]**. For $q < r < s$,

$$d_\oplus(l(q), l(r)) \oplus d_\oplus(l(r), l(s)) = d_\oplus(l(q), l(s)) \tag{29}$$

$$d_\boxplus(m(q), m(r)) \boxplus d_\boxplus(m(r), m(s)) = d_\boxplus(m(q), m(s)) \tag{30}$$

(Recall that $\boxplus = \oplus$ in $((-1,1), \oplus, \odot)$.) Equation (29) is a straightforward consequence of gyrotranslation invariance (21), (G7) and (P2). Equation (30), on the other hand, can be shown using the following important property of dual gyrolines. For each $a, c \in \mathbb{D}$, gyrotranslating a by multiples of c on the left describes a group action of $\mathbb{R}$ on $\mathbb{D}$:

$$s \odot c \oplus (t \odot c \oplus a) = (s \odot c \oplus t \odot c) \oplus a = (s+t) \odot c \oplus a, \tag{31}$$

using (G3), (G11) and (G7). Not so straightforward, and hence omitted, are the proofs of converses: if three points satisfy (29), resp. (30), then there exists a gyroline, resp. dual gyroline, containing them.

As FIGURE 1(a) and (29) suggest, gyrolines in $\mathbb{D}$ are the "lines" of the *Poincaré model* of hyperbolic geometry. These are the *geodesics* (curves of shortest arclength) in the metric space $(\mathbb{D}, d_{\oplus})$ **[8]**. The gyroline through a and b is an arc of the unique Euclidean circle passing through a and b (or a line segment if a and b are collinear with the origin) which approaches the boundary of $\mathbb{D}$ orthogonally. (This is because $l(t)$ is the image under a Möbius transformation of the line segment given by $t \odot (-a \oplus b)$ **[8]**.)

FIGURE 1(b) shows that a dual gyroline is an arc of a circle approaching the boundary of $\mathbb{D}$ at the endpoints of a gyroline through the origin which we call the *supporting diameter*. If $m(t) = t \odot b \oplus a$, $t \in \mathbb{R}$, is a dual gyroline, then its supporting diameter is given by $p(t) = t \odot b$. Despite the algebraic elegance of (30), dual gyrolines (28) turn out *not* to be geodesics in the metric space $(\mathbb{D}, d_{\boxplus})$. In fact, it is an open problem to understand the complete geometric significance of (30). Nevertheless, we will show later that dual gyrolines do in fact play a role in hyperbolic geometry, and they also form the lines of a geometry of their own, although it is not exactly the geometry of $(\mathbb{D}, d_{\boxplus})$. Incidentally, another open problem is to describe the actual geodesics in $(\mathbb{D}, d_{\boxplus})$ in terms of the gyro-structure of $\mathbb{D}$.

We now show how the gyro-structure of $\mathbb{D}$ illuminates a well-known result in hyperbolic geometry: given a line l and a point a not on l, there are infinitely many lines through a which are parallel to (i.e., do not intersect) l (**[5]**, **[10]**). To simplify computations, we illustrate this in the special case of a gyroline $l(t) = t \odot b$ through the origin, and a point a not on l. For $r \in \mathbb{R}$, define $\phi_r = \mathrm{gyr}[-a, r \odot b]$, and extend this continuously by $\phi_{\infty} = (|b| - a\bar{b})/(|b| - \bar{a}b)$ and $\phi_{-\infty} = (|b| + a\bar{b})/(|b| + \bar{a}b)$. Then for any $r \in [-\infty, \infty]$, the gyroline given by

$$l_r(t) = a \oplus t \odot \phi_r b, \tag{32}$$

$t \in \mathbb{R}$, passes through a (when $t = 0$) and is parallel to $l(t) = t \odot b$, FIGURE 2(a). Conversely, any gyroline passing through a and parallel to l turns out to be given by (32) for some $r \in [-\infty, \infty]$. The two extreme gyrolines l_{∞} and $l_{-\infty}$ satisfy $\lim_{t \to \infty} l_{\infty}(t) = b/|b|$ and $\lim_{t \to -\infty} l_{-\infty}(t) = -b/|b|$, respectively, and are said to be *asymptotically divergent* to l **[5]**.

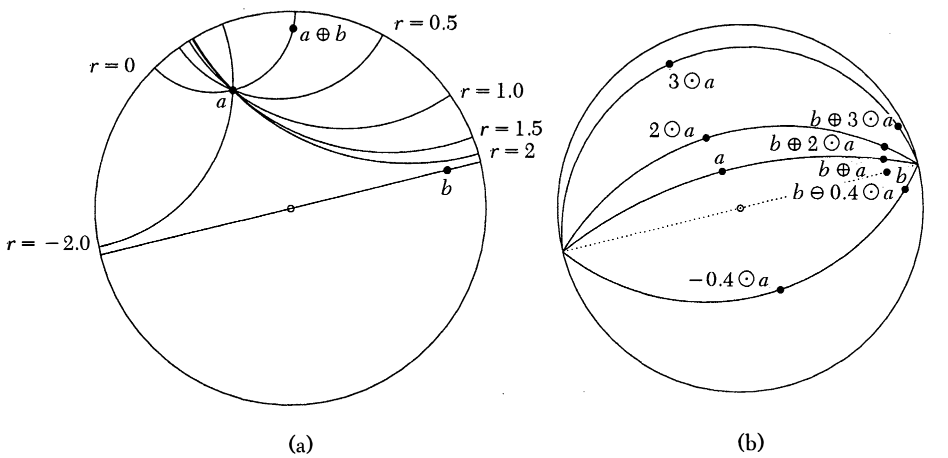


FIGURE 2
(a) Several gyrolines parallel to a given gyroline
(b) Several mutually parallel dual gyrolines

In FIGURE 2(b), two dual gyrolines are parallel (nonintersecting) if and only if they share the same supporting diameter. Thus given a dual gyroline m and a point a not on m, there is *exactly one* dual gyroline through a which is parallel to m. Therefore, while gyrolines satisfy the *hyperbolic* parallel postulate, dual gyrolines satisfy the *Euclidean* parallel postulate. There is a good reason for this, as we will discuss later.

Given points $a, b \in \mathbb{D}$, we will identify the element $-a \oplus b \in \mathbb{D}$ with the *gyrovector* consisting of the oriented segment of the gyroline joining a to b, i.e., with $l(t) = a \oplus t \odot (-a \oplus b)$ for $0 \leq t \leq 1$. The length of the gyrovector $-a \oplus b$ is defined to be the distance $d_{\oplus}(a, b)$ between its endpoints. If $-a \oplus b = -c \oplus d$, then we consider the corresponding gyrovectors to be equivalent. In particular, $-a \oplus b$ is equivalent to a unique gyrovector emanating from the origin. With this interpretation, (27) can be viewed as the unique gyroline passing through a in the direction of the gyrovector $-a \oplus b$. Equivalent gyrovectors obviously have the same length, but unlike the Euclidean case, they are *not* necessarily parallel; we leave as an exercise the task of finding a pair of equivalent, intersecting gyrovectors.

Given $a, b \in \mathbb{D}$, we identify the element $b \boxminus a \in \mathbb{D}$ with the *dual gyrovector* consisting of the oriented segment of the dual gyroline joining a to b, i.e., $m(t) = t \odot (b \boxminus a) \oplus a$, $0 \leq t \leq 1$. If $d \boxminus c = b \boxminus a$, then we say that the corresponding dual gyrovectors are equivalent. In particular, $b \boxminus a$ is equivalent to a unique gyrovector emanating from the origin. With this interpretation, (28) can be viewed as the unique dual gyroline passing through a in the direction of the gyrovector $b \boxminus a$. Equivalent dual gyrovectors share the same supporting diameter, and thus are parallel. Equivalent dual gyrovectors also have the same norm, and thus the same dual distance between their endpoints.

Next we will define angles between gyrovectors. As a preliminary to this, we observe that for $a, b \in \mathbb{D}$, the point $\frac{-a \oplus b}{\| -a \oplus b\|}$ on the unit circle in $\mathbb{C}$ depends only on the *gyroray* $l(t) = a \oplus t \odot (-a \oplus b)$, $t \geq 0$, emanating from a and passing through b. Indeed, if $c \in l$ is given by $c = l(s)$ for some $s > 0$, then (10) implies $-a \oplus c = s \odot (-a \oplus b)$, and thus

$$\frac{-a \oplus c}{\|-a \oplus c\|} = \frac{s \odot (-a \oplus b)}{\|s \odot (-a \oplus b)\|} = \frac{-a \oplus b}{\|-a \oplus b\|}$$

by the scaling identity (15). Now let $-a \oplus b_1$ and $-a \oplus b_2$ be gyrovectors emanating from the point a. We define the *angle* α between $-a \oplus b_1$ and $-a \oplus b_2$, FIGURE 3(a), by

$$\cos \alpha = \frac{-a \oplus b_1}{\|-a \oplus b_1\|} \cdot \frac{-a \oplus b_2}{\|-a \oplus b_2\|}. \tag{33}$$

By the preceding discussion, (33) depends only on the two gyrorays emanating from a and passing through b_1 and b_2, respectively. Thus we may also speak of the angle between the gyrorays. Although (33) depends only on the intrinsic gyro-structure of $\mathbb{D}$, it turns out that the angle measure is equal to that defined by the tangent vectors $l_i'(0)$ at a. This is easy to show by just computing the tangent vectors explicitly, but it is perhaps intuitively clear from the observation that since b_1 and b_2 can be taken to be "infinitesimally close" to a, the tangent vectors $l_i'(0)$ are just multiples of $\frac{-a \oplus b_i}{\| -a \oplus b_i\|}$, $i = 1, 2$.

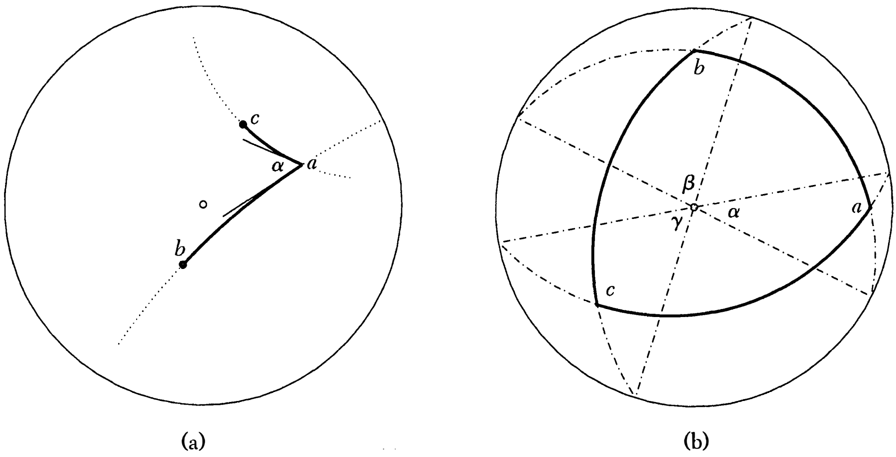


FIGURE 3
(a) The angle between gyrovectors
(b) The dual angle between dual gyrovectors

Dualizing this discussion, we define the *dual angle* between the dual gyrovectors $b_1 \boxminus a$ and $b_2 \boxminus a$, FIGURE 3(b), by

$$\cos\alpha = \frac{b_1 \boxminus a}{\|b_1 \boxminus a\|} \cdot \frac{b_2 \boxminus a}{\|b_2 \boxminus a\|}. \tag{34}$$

By an argument similar to the previous one, it can be shown that this definition does not depend on the choice of b_1 and b_2 on the dual gyrorays given by $(b_i \boxminus a) \oplus t \odot a$ for $t > 0$, $i = 1, 2$ **[19]**. Indeed, we see that the dual angle between dual gyrovectors is simply the angle between their supporting diameters.

Next we show how another well-known result in hyperbolic geometry is immediate from the gyro-structure of $\mathbb{D}$: that the sum of the angles in a hyperbolic triangle is less than π. Let $a, b, c \in \mathbb{D}$ be points that do not lie on the same gyroline, and let α, β, γ denote the corresponding angles of the hyperbolic triangle with vertices a, b, c, FIGURE 4. As one can check directly using (5) and (33) (see also **[18]**, eqns. (7.19)–(7.20)), we have

$$e^{i(\pi-\alpha-\beta-\gamma)} = \mathrm{gyr}[a, -b]\mathrm{gyr}[b, -c]\mathrm{gyr}[c, -a]. \tag{35}$$

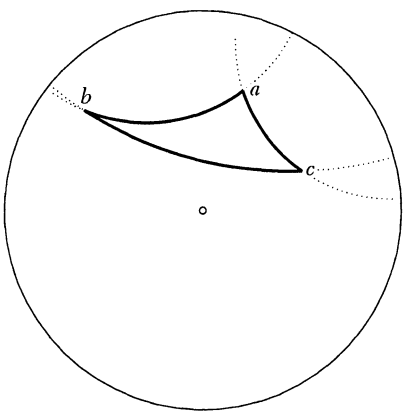


FIGURE 4
A hyperbolic triangle

Since a, b, and c do not lie on a common gyroline, it follows from (T1) and (13) that

$$\mathrm{gyr}[a, -b]\mathrm{gyr}[b, -c]\mathrm{gyr}[c, -a] \neq \mathrm{gyr}[a, -a] = 1. \tag{36}$$

Together, (35) and (36) imply that $\cos(\pi - \alpha - \beta - \gamma) < 1$ or $\pi > \alpha + \beta + \gamma$, as claimed.

Once again, we dualize the preceding result. Let $a, b, c \in \mathbb{D}$ be points which do not lie on the same dual gyroline, and let α, β, γ denote the corresponding dual angles of the *dual triangle* with vertices a, b, c, Figure 3(b). Since the dual angles (34) are given by the angles between the supporting diameters, it follows that $\alpha + \beta + \gamma = \pi$. Thus unlike hyperbolic triangles, dual triangles satisfy the *Euclidean* sum-of-angles property.

There are some results in hyperbolic geometry which are clarified by the gyro-structure of $\mathbb{D}$, but for which there do not seem to be matching dual results. Here we consider two examples: the hyperbolic law of cosines and the hyperbolic law of sines. Let $A = -c \oplus b$, $B = -c \oplus a$ and $C = -a \oplus b$ denote the gyrovectors giving the sides of the hyperbolic triangle with vertices a, b, c, and as before let α, β, γ denote the corresponding angles. By (22) and (G6), we have $A \ominus B = \mathrm{gyr}[-c, b]\mathrm{gyr}[b, -a]C$. Using this, (P4), (16) and (33), we obtain the following form of the *Hyperbolic Law of Cosines*:

$$\|C\|^2 = \|A\|^2 \oplus \|B\|^2 \ominus \frac{\frac{1}{2}(2 \odot \|A\|)(2 \odot \|B\|)\cos\gamma}{1 - \frac{1}{2}(2 \odot \|A\|)(2 \odot \|B\|)\cos\gamma}. \tag{37}$$

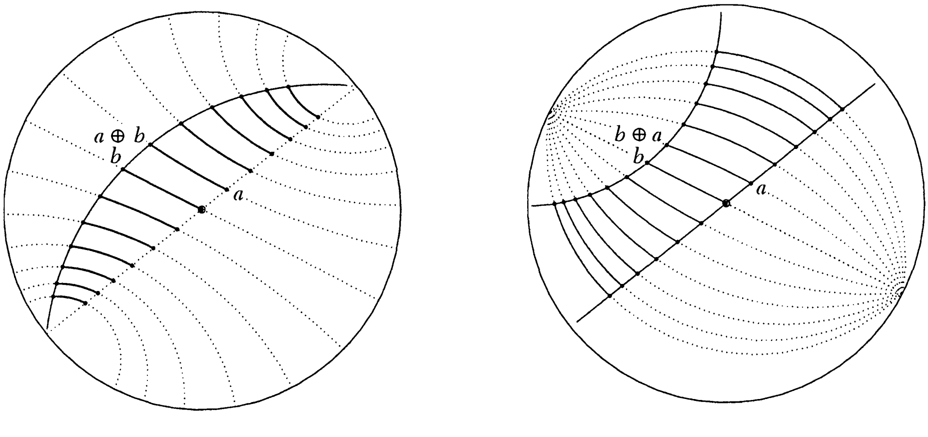
Somewhat more tedious calculations lead to the following *Hyperbolic Law of Sines*:

$$\left(1 - \|A\|^2\right)\frac{\sin\alpha}{\|A\|} = \left(1 - \|B\|^2\right)\frac{\sin\beta}{\|B\|} = \left(1 - \|C\|^2\right)\frac{\sin\gamma}{\|C\|}. \tag{38}$$

As a corollary to (37), we have the *Pythagorean theorem for hyperbolic right triangles*: If A and B are orthogonal, then $\|C\|^2 = \|A\|^2 \oplus \|B\|^2$ **[19][20]**.

So far we have been treating dual gyrolines as being objects which are dual, in some sense, to hyperbolic geometry. However, they actually do have an interpretation within hyperbolic geometry itself, and once again, the gyro-structure of $\mathbb{D}$ highlights this interpretation. Let $m(t) = t \odot b \oplus a$, $t \in \mathbb{R}$, be a dual gyroline with supporting diameter $p(t) = t \odot b$. For each $t \in \mathbb{R}$, consider the gyrovector joining $p(t)$ to $m(t)$. By (10), this is $-p(t) \oplus (p(t) \oplus a) = a$. Thus all such gyrovectors are equivalent. The angle between each such gyrovector and the supporting diameter is given by $\cos\alpha = \frac{b \cdot a}{\|b\|\,\|a\|}$, which is independent of t. Therefore, we have our interpretation: *A dual gyroline is the locus of points formed by a family of equivalent gyrovectors emanating from a given gyroline passing through the origin*, Figure 5(a). Since each gyrovector in the family has the same length and has the same angle with the given gyroline through the origin, dual gyrolines are known in hyperbolic geometry as *equidistant curves* **[5]**. In Euclidean geometry, lines and equidistant curves coincide. Here we see how the gyro-structure of $\mathbb{D}$ naturally reveals the distinction between the two types of curves in hyperbolic geometry.

Once again, the preceding discussion can be dualized. Let $l(t) = a \oplus t \odot b$ be a gyroline with "supporting diameter" given by $q(t) = t \odot b$. For each $t \in \mathbb{R}$, consider the dual gyrovector joining $q(t)$ to $l(t)$. This is $(a \oplus q(t)) \boxminus q(t) = a$, using (11), and thus all such dual gyrovectors are equivalent. The dual angle between any such gyrovector and the gyroline q is given by $\cos\alpha = (b \cdot a)/(\|b\|\,\|a\|)$. This gives us a dual interpretation of gyrolines: *A gyroline is the locus of points formed by a family of*



(a) (b)

FIGURE 5
(a) Dual gyrolines as equidistant curves in hyperbolic geometry
(b) Gyrolines as equidistant to dual gyrolines

equivalent dual gyrovectors emanating from a given gyroline passing through the origin. Here as before, each dual gyrovector in the family has the same dual distance between its endpoints and forms the same dual angle with the given gyroline through the origin, FIGURE 5(b).

We would be remiss if we did not mention that there are standard results in hyperbolic geometry for which there do not seem to be good "gyro-proofs." For example, the medians of a hyperbolic triangle are concurrent **[5]**. A proof of this using the parameterization (27) in a way which generalized the vector space proof would require the elusive gyrodistributive law discussed in §3.

We have already noted two Euclidean properties arising from the dual gyrolines: (i) dual gyrolines satisfy Euclidean parallelism, and (ii) the sum of the angles in a dual triangle is π. There is, in fact, an explanation for these Euclidean features in what otherwise appears to be a hyperbolic world. There is a well-known model for Euclidean geometry *within* hyperbolic geometry in which the lines are exactly those curves that are equidistant to lines through the origin **[10]**. We have seen that these are exactly the dual gyrolines (28). Thus the Euclidean properties we found through the gyro-structure reflect the role that dual gyrolines play in this Euclidean model. Interestingly, interpreted as Euclidean objects, dual triangles certainly have concurrent medians, but interpreted as objects in the metric space $(\mathbb{D}, d_{\boxplus})$, the median concurrence property does not hold **[19]**.

As discussed at the beginning of this section, the gyrolines (27) and dual gyrolines (28) correspond to the cancellation laws (10) and (11), respectively. There is a *third* type of gyroline analogous to (26), which corresponds to the third cancellation law (12). Fix $a, b \in \mathbb{D}$, and set

$$n(t) = t \odot (b \ominus a) \boxplus a. \tag{39}$$

See FIGURE 6. We have $n(0) = a$ and, by (12), $n(1) = b$. As a curve in the plane, the gyroline $n(t)$ is a hyperbola with asymptotes crossing at $2 \odot a$ and passing through, respectively, $(b \ominus a)/\|b \ominus a\|$ and $(a \ominus b)/\|a \ominus b\|$, the endpoints of the diameter $t \odot (b \ominus a)$. That the gyroline $n(t)$ arises so naturally out of the gyro-structure of the disk $\mathbb{D}$ suggests it might have significance within hyperbolic geometry. We do not know what that significance might be, and we conclude with an invitation to the reader to consider this problem.

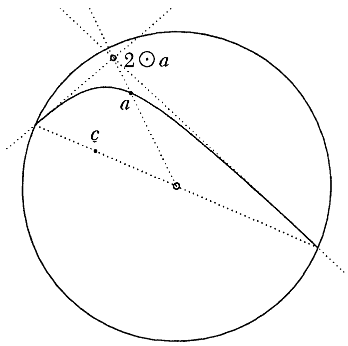


FIGURE 6
The gyroline $n(t) = t \odot (b \ominus a) \boxplus a$

Acknowledgment. The authors would like to thank the referees for helpful suggestions, and Oliver Jones for useful comments. One of us (AU) is pleased to acknowledge financial support from ND EPSCoR through NSF grant #OSE-9452892.

REFERENCES

1. C. F. Blakemore, R. Lange, and R. Tucci, A differential equation in group theory, this MAGAZINE 55 (1982), 220–221.
2. M. A. Carchidi, Generating exotic-looking vector spaces, *College Math. J.* 29 (1998), 304–308.
3. A. Einstein, Zur Elektrodynamik Bewegter Körper (On the Electrodynamics of Moving Bodies), *Ann. Physik* (Leipzig) 17 (1905), 891–921. For an English translation, see H. M. Schwartz, *Amer. J. Phys.* 45 (1977), 18, or H. A. Lorentz, A. Einstein, H. Minkowski and H. Weyl, *The Principle of Relativity*, (trans. W. Perrett and G. B. Jeffrey), Dover, New York, NY, 1952, 37–65.
4. S. D. Fisher, The Möbius group and invariant spaces of analytic functions, *Amer. Math. Monthly* 95 (1988), 514–527.
5. M. J. Greenberg, *Euclidean and Non-Euclidean Geometries: Development and History*, 3rd Ed., W. H. Freeman Co., New York, NY, 1993.
6. S. B. Jackson and D. Greenspan, Hyperbolic analytic geometry, this MAGAZINE, 20 (1955), 251–269.
7. D. Kalman and P. Turner, Algebraic structures with exotic operations, *Int. J. Math. Ed. in Science and Tech.* 10(2) (1979), 173–174.
8. S. Krantz, *Complex Analysis: The Geometric Viewpoint*, Carus Mathematical Monograph 23, MAA, Washington, DC, 1990.
9. H. O. Pflugfelder, *Quasigroups and Loops: Introduction*, Heldermann-Verlag, Berlin, Germany, 1990.
10. A. Ramsay and R. D. Richtmyer, *Introduction to Hyperbolic Geometry*, Springer-Verlag, New York, NY, 1995.
11. J. Rickart, private email response to a post to *sci.math.*
12. S. Tomonaga, *The Story of Spin* (trans. T. Oka), Univ. of Chicago Press, Chicago, IL, 1997.
13. A. A. Ungar, Thomas rotation and the parametrization of the Lorentz transformation group, *Found. Phys. Lett.*, 1 (1988), 57–89.
14. A. A. Ungar, The relativistic noncommutative nonassociative group of velocities and the Thomas rotation, *Results Math.* 16 (1989), 168–179.
15. A. A. Ungar, Thomas precession and its associated grouplike structure, *Amer. J. Phys.* 59 (1991), 342–354.
16. A. A. Ungar, The holomorphic automorphism group of the complex unit disk, *Aequationes Math.* 47 (1994), 240–254.
17. A. A. Ungar, Midpoints in gyrogroups, *Found. Phys.* 26 (1996), 1277–1328.
18. A. A. Ungar, Thomas precession: its underlying gyrogroup axioms and their use in hyperbolic geometry and relativistic physics, *Found. Phys.* 27 (1997), 881–951.
19. A. A. Ungar, From Pythagoras to Einstein: the hyperbolic Pythagorean theorem, *Found. Phys.* 28 (1998), 1283–1321.
20. A. A. Ungar, The hyperbolic Pythagorean theorem in the Poincaré disc model of hyperbolic geometry, *Amer. Math. Monthly* 106 (1999), 759–762.

PIPCIRs—Polynomials Whose Inflection Points Coincide with Their Interior Roots

RACHEL BELINSKY
Morris Brown College
Atlanta, GA 30314-4140

PETER ANDREWS
Eastern Illinois University
Charleston, IL 61920

Introduction

As a consequence of Rolle's theorem, one of the great results of first semester calculus, between any two real roots of a polynomial there is a critical point. So, if a polynomial of degree n has n distinct, real roots, then between each pair of successive roots there is a critical point, and between each pair of successive critical points there is an inflection point.

Several recent articles ([**1**], [**2**], [**3**], [**7**]) looked at real polynomials of degree n with n distinct, real roots and at the question of what restrictions there might be on where, in the $n-1$ intervals between the roots, the critical points can fall. For polynomials like this, we could also turn things around a bit and think of the $n-1$ critical points as dividing the real numbers into $n-2$ bounded intervals, each containing exactly one root (we will call these the interior roots) and one inflection point. In [**2**], Anderson discovered an intriguing collection of polynomials, of degree n, with n distinct, real roots, all of whose $n-2$ inflection points fall exactly on the $n-2$ interior roots. If (without losing any generality) you make ± 1 the exterior roots and normalize by insisting that the polynomial be monic, then there is exactly one such polynomial, $\widetilde{Q}_n(x)$, for each degree $n \geq 3$. Anderson obtained the following results:

RESULT 1:. *For each $n \geq 3$ there is a unique monic polynomial $\tilde{q}_{n-2}$ of degree $n-2$ such that the degree n polynomial $\widetilde{Q}_n$ defined by*

$$\widetilde{Q}_n(x) = (x^2 - 1)\ \tilde{q}_{n-2}(x)$$

satisfies the condition

$$\widetilde{Q}_n''(x) = (n)(n-1)\tilde{q}_{n-2}(x).$$

RESULT 2:. *The polynomials $\tilde{q}_n$ have n distinct real roots, all in the interval* $(-1, 1)$.

Together, these results imply that the polynomials $\widetilde{Q}_n$ have the maximum number of real roots and inflection points, and that all their inflection points coincide exactly with their interior roots.

FIGURE 1 shows some of these special polynomials. The apparent similarity of these graphs can be somewhat deceptive. Notice how the values of $\widetilde{Q}_n(x)$, $-1 \leq x \leq 1$, get dramatically smaller as n increases. Normalizing the even degree polynomials by

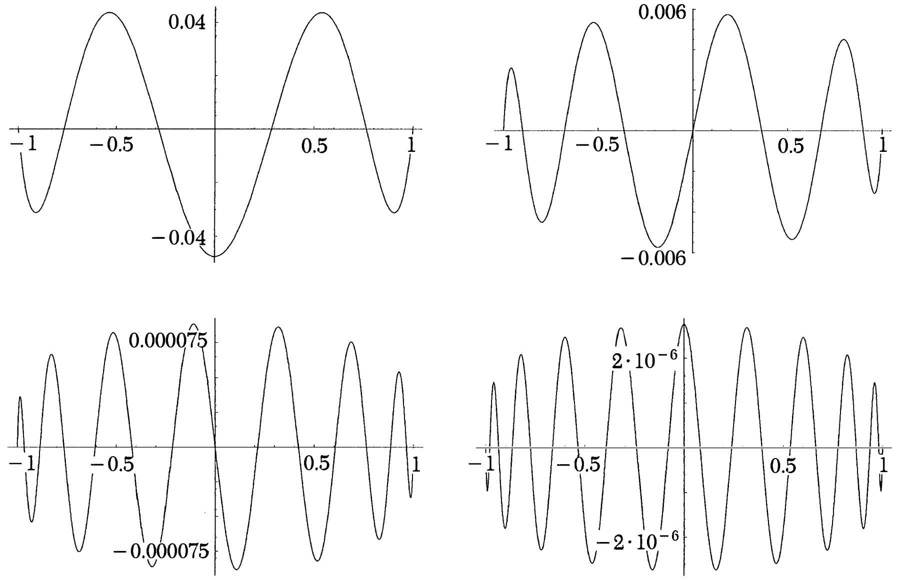


FIGURE 1

$\widetilde{Q}_n(x)$. Clockwise, from the top left, $n = 6, 9, 15, 20$

plotting $\widetilde{Q}_{2m}(x)/\widetilde{Q}_{2m}(0)$, as in Figure 2, reveals an amazingly consistent pattern: all of the graphs appear (Figure 3) to be contained inside the same envelope. Anderson concludes **[2]** as follows:

The author wagers that the asymptotic limit is given by $\sqrt[4]{1-x^2}$ ... It is left as a challenge to the reader to prove or disprove this, and to find an asymptotic limit for odd degree polynomials as well.

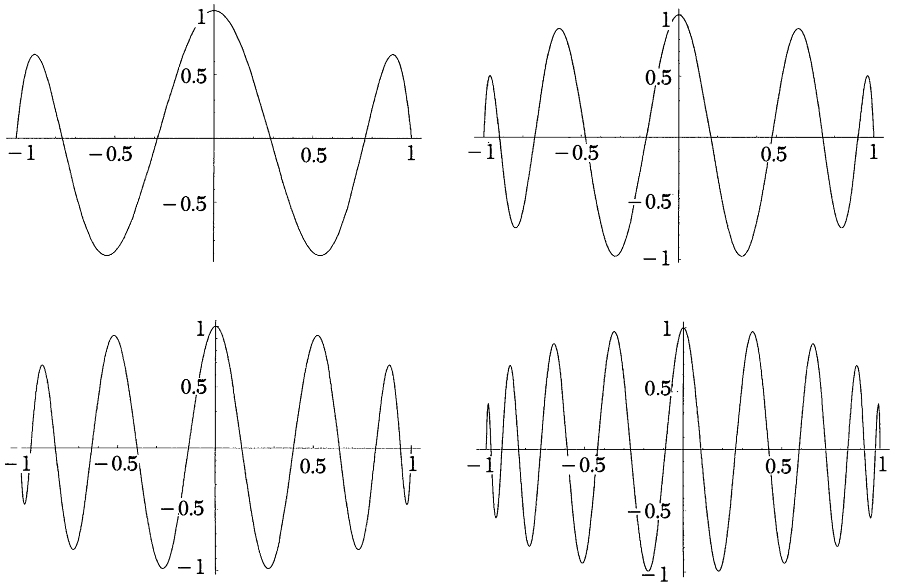


FIGURE 2

$\widetilde{Q}_n(x)/\widetilde{Q}_n(0)$. Clockwise, from top left, $n = 6, 10, 12, 18$

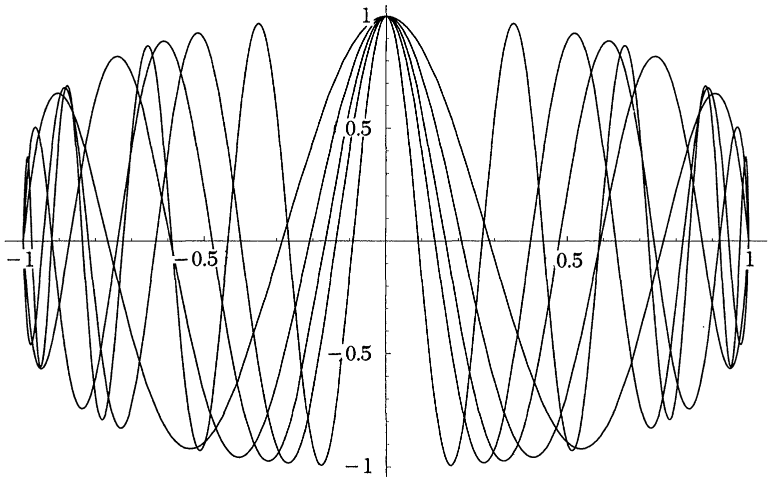


FIGURE 3
The graphs from FIGURE 2, all on the same axes

Careful plotting indicates just how close Anderson's proposed envelope gets, even for relatively small values of n. FIGURE 4 shows $\widetilde{Q}_8(x)/\widetilde{Q}_8(0)$ and $\pm\sqrt[4]{1-x^2}$ along with a zoom near the first positive critical point of $\widetilde{Q}_8(x)/\widetilde{Q}_8(0)$; one sees that $\widetilde{Q}_8(x)/\widetilde{Q}_8(0)$ pokes barely outside the purported envelope.

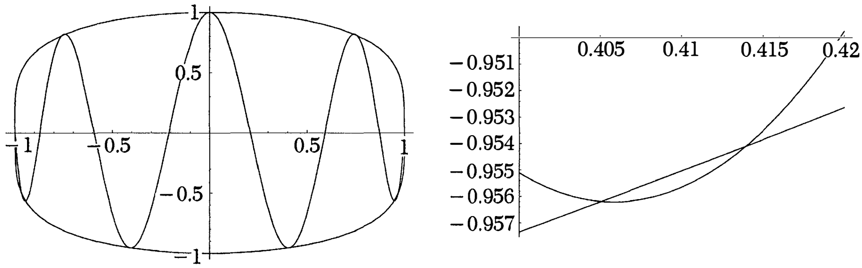


FIGURE 4
(a): $\widetilde{Q}_8(x)/\widetilde{Q}_8(0)$ and $\pm\sqrt[4]{1-x^2}$; (b): the same graph zoomed in on the "valley" near $x = 0.41$

In this note we take up Anderson's challenge, proving $\pm\sqrt[4]{1-x^2}$ to be, indeed, the asymptotic envelope for $\widetilde{Q}_{2m}(x)/\widetilde{Q}_{2m}(0)$. We also show $\pm\sqrt[4]{1-x^2}$ to be the asymptotic envelope for the suitably normalized $\widetilde{Q}_{2m+1}$. The key to our approach is noting the connections between these polynomials, whose inflection points coincide with their interior roots (PIPCIRs, for short), and the classical Legendre and Jacobi polynomials.

1. PIPCIRs and classical orthogonal polynomials

A PIPCIR, p_n, with exterior roots at ± 1, satisfies the equation

$$p_n'' = n(n-1)\frac{p_n}{x^2-1},$$

for $n \geq 3$. This can be re-written as

$$(1-x^2)p_n'' + n(n-1)p_n = 0. \tag{1}$$

Differentiating (1) yields

$$-2xp_n'' + (1-x^2)p_n''' + n(n-1)p_n' = 0.$$

Letting $y_{n-1} = p_n'$, we get the differential equation

$$(1-x^2)y_{n-1}'' - 2xy_{n-1}' + n(n-1)y_{n-1} = 0, \tag{2}$$

which is *Legendre's differential equation*. Every solution to this equation that is bounded on $[-1,1]$ is a multiple of the Legendre polynomial P_n.

Properties of the Legendre polynomials can be found in references on orthogonal polynomials, **[6]** being among the classics. Each P_n is a degree-n polynomial with n distinct, real roots, all in the interval $(-1,1)$, and is normalized so $P_n(1) = 1$. One of the many definitions of these polynomials is given by *Rodrigues' formula*:

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n.$$

The first few examples are as follows:

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \tfrac{3}{2}x^2 - \tfrac{1}{2}, \quad P_3(x) = \tfrac{5}{2}x^3 - \tfrac{3}{2}x.$$

Orthogonality of the Legendre polynomials on the interval $[-1,1]$ means that

$$\int_{-1}^{1} P_i(x)P_j(x)\,dx = 0, \quad i \neq j.$$

In particular, since $P_0(x) = 1$,

$$\int_{-1}^{1} P_n(x)\,dx = 0, \quad n > 0. \tag{3}$$

Now back to the PIPCIRs. Since we defined $y_{n-1} = p_n'$, we see that $p_n'(x) = cP_{n-1}(x)$, for some constant c. One particular choice for p_n would then be $Q_n(x) = \int_1^x P_{n-1}(t)\,dt$, or, to emphasize that we are mostly interested in $x \in [-1,1]$,

$$Q_n(x) = -\int_x^1 P_{n-1}(t)\,dt, \quad n > 1.$$

Thus, $Q_n'(x) = P_{n-1}(x)$ and $Q_n''(x) = P_{n-1}'(x)$. Moreover, $Q_n(1) = 0$ and $Q_n(-1) = -\int_{-1}^{1} P_{n-1}(t)\,dt = 0$, $n > 1$, by (3). Thus, Q_n, is a degree n polynomial with roots at ± 1 and with $n-1$ critical points at the $n-1$ roots of P_{n-1}, all of which lie in $(-1,1)$. No real roots of Q_n can lie outside of the interval $[-1,1]$.

Since the Legendre polynomials solve (2), we have $(1-x^2)P_n''(x) - 2xP_n'(x) + n(n+1)P_n(x) = 0$, which can be re-written as

$$\frac{d}{dx}\left[(1-x^2)Q_{n+1}''(x)\right] + n(n+1)P_n(x) = 0.$$

Integrating from 1 to x yields

$$(1-x^2)Q''_{n+1}(x)-n(n+1)\int_x^1 P_n(t)\,dt=0,$$

or

$$(1-x^2)Q''_{n+1}(x)+n(n+1)Q_{n+1}(x)=0. \tag{4}$$

Thus, the polynomials Q_n satisfy equation (1) and any inflection points that Q_n might have must coincide with its interior roots.

Now we just need to make sure Q_n has enough real roots. Since Q_n is a polynomial with roots at ± 1, it can be factored as

$$Q_n(x)=(1-x^2)q_{n-2}(x), \tag{5}$$

where q_{n-2} is a degree $n-2$ polynomial, all of whose roots must lie in $(-1,1)$. We can compute $q(1)$ with L'Hôpital's Rule:

$$q_n(1)=\lim_{x\to 1}\frac{Q_{n+2}(x)}{1-x^2}=\lim_{x\to 1}\frac{P_{n+1}(x)}{-2x}=-\frac{1}{2}.$$

Substituting (5) in (4), we get

$$(1-x^2)q''_n-4xq'_n+n(n+3)q_n=0, \tag{6}$$

another classic differential equation.

The *Jacobi polynomials* $P_n^{(\alpha,\beta)}$ are solutions to

$$(1-x^2)y''+[\beta-\alpha-(\alpha+\beta+2)x]y'+n(n+\alpha+\beta+1)y=0, \tag{7}$$

bounded on the interval $[-1,1]$. The polynomials $P_n^{(\alpha,\beta)}$ have n distinct, real roots, all in $(-1,1)$, and are normalized so that $P_n^{(\alpha,\beta)}(1)=\binom{n+\alpha}{n}$.

When $\alpha=1=\beta$, (7) becomes (6), so that any solution to (6) that is bounded on $[-1,1]$ (such as q_n) must be a constant multiple of $P_n^{(1,1)}$. Moreover, since $q_n(1)=-1/2$ and $P_n^{(1,1)}(1)=n+1$, we have

$$q_n(x)=\frac{-1}{2(n+1)}P_n^{(1,1)}(x). \tag{8}$$

We see also that q_n has n distinct, real roots, all in $(-1,1)$. We have finally shown that the polynomials Q_n, $n\geq 3$, are all PIPCIRs. Results 1 and 2 from **[2]** follow since each q_n (and hence Q_n) is unique up to a constant multiple. The polynomials Q_n are constant multiples of the $\widetilde{Q}_n$ from **[2]**.

Explicit formulas for the Q_n can be obtained by directly integrating the Legendre polynomials (see **[5]**, p. 120):

$$Q_n(x)=\sum_{k=0}^{\lfloor \frac{n}{2}\rfloor}\frac{(-1)^k(2n-2k-3)!!}{(2k)!!(n-2k)!}x^{n-2k}, \quad n\geq 2. \tag{9}$$

(The double factorial, $n!!$, is defined as $n(n-2)(n-4)\cdots(4)(2)$ if n is even and $n(n-2)(n-4)\cdots(3)(1)$ if n is odd. For completeness, $0!!=1$.)

If n is even, (9) gives us

$$Q_n(0)=\frac{(-1)^{n/2}(n-3)!!}{n!!}. \tag{10}$$

If n is odd, there is no constant term in (9) and $Q_n(0) = 0$. Since $Q_2(x) = (x^2 - 1)/2$ has no interior roots and no inflection points, it could be considered as a PIPCIR as well. This leaves the following table of the first few PIPCIRs, should you be looking for a classroom example or test question!

$$Q_2(x) = \frac{x^2 - 1}{2} \qquad q_0(x) = -\frac{1}{2}$$

$$Q_3(x) = \frac{x^3 - x}{2} \qquad q_1(x) = -\frac{x}{2}$$

$$Q_4(x) = \frac{5x^4 - 6x^2 + 1}{8} \qquad q_2(x) = \frac{-5x^2 + 1}{8}$$

$$Q_5(x) = \frac{7x^5 - 10x^3 + 3x}{8} \qquad q_3(x) = \frac{-7x^3 + 3x}{8}$$

$$Q_6(x) = \frac{21x^6 - 35x^4 + 15x^2 - 1}{16} \qquad q_4(x) = \frac{-21x^4 + 14x^2 - 1}{16}$$

2. Asymptotics and envelopes

Much has been written about finding approximations for the classical orthogonal polynomials. For example, (Theorem 8.21.8) in **[6]** describes an asymptotic formula for $P_n^{(\alpha,\beta)}(\cos\theta)$, valid for all α and β and for $0 < \theta < \pi$. When $\alpha = \beta = 1$, this reduces to

$$P_n^{(1,1)}(\cos\theta) = \frac{1}{\sqrt{n\pi}}\left(\sin\frac{\theta}{2}\cos\frac{\theta}{2}\right)^{-\frac{3}{2}}\cos\left(\left(n + \frac{3}{2}\right)\theta - \frac{3\pi}{4}\right) + \mathcal{O}(n^{-\frac{3}{2}}).$$

(Since $f(n) = \mathcal{O}(g(n))$ means that $\lim_{n\to\infty}\frac{f(n)}{g(n)}$ is bounded, this formula implies that the error term goes to 0 at least as fast as (a constant multiple of) $n^{-\frac{3}{2}}$.)

If we let $x = \cos\theta$, then

$$\sin\frac{\theta}{2}\cos\frac{\theta}{2} = \frac{1}{2}\sin\theta = \frac{1}{2}\sqrt{1 - x^2},$$

and

$$P_n^{(1,1)}(x) = \sqrt{\frac{8}{n\pi}}\,(1 - x^2)^{-\frac{3}{4}}\cos\left(\left(n + \frac{3}{2}\right)\arccos x - \frac{3\pi}{4}\right) + \mathcal{O}(n^{-\frac{3}{2}}), \quad -1 < x < 1.$$

Combining this with (8) and (5), we get

$$\begin{aligned} Q_n(x) &= -\frac{\sqrt[4]{1 - x^2}}{2(n-1)}\sqrt{\frac{8}{\pi(n-2)}}\cos\left(\left(n - \frac{1}{2}\right)\arccos x - \frac{3\pi}{4}\right) + \mathcal{O}(n^{-\frac{5}{2}}) \\ &= -\sqrt{\frac{2}{\pi(n-2)(n-1)^2}}\,\sqrt[4]{1 - x^2}\cos\left(\left(n - \frac{1}{2}\right)\arccos x - \frac{3\pi}{4}\right) + \mathcal{O}(n^{-\frac{5}{2}}) \\ &= \sqrt{\frac{2}{\pi(n-2)(n-1)^2}}\,A_n(x) + \mathcal{O}(n^{-\frac{5}{2}}), \end{aligned} \tag{11}$$

where

$$A_n(x) = -\sqrt[4]{1-x^2}\cos((n-1/2)\arccos x - 3\pi/4). \tag{12}$$

So, for $x \in (-1, 1)$,

$$Q_n(x) \approx \sqrt{\frac{2}{\pi(n-2)(n-1)^2}}\, A_n(x).$$

(Here, and subsequently, we will use $X_n \approx Y_n$ to denote $\lim_{n\to\infty} X_n/Y_n = 1$.) This provides us with an obvious normalization. Define

$$\overline{K}_n = \sqrt{\frac{2}{\pi(n-2)(n-1)^2}} \quad \text{and} \quad \overline{Q}_n = \frac{1}{\overline{K}_n} Q_n.$$

Then $\overline{Q}_n(x) \approx A_n(x)$ and, if x_0 is a point where $\cos((n-1/2)\arccos x_0 - 3\pi/4) = 1$, then, (12) implies that

$$\overline{Q}_n(x_0) \approx A_n(x_0) = -\sqrt[4]{1-x_0^2}.$$

This establishes the envelope for suitably normalized Q_n in both even and odd cases. In the even case, however, the normalizing factor, $\overline{K}_n$, is not the same as the more obvious choice, $\widehat{K}_n = |Q_n(0)|$, suggested earlier in this article.

In order to compare these two normalizations, define $\widehat{Q}_n(x) = Q_n(x)/\widehat{K}_n = Q_n(x)/|Q_n(0)|$. Equation (10) gives a formula for $Q_n(0)$ when n is even. Recalling Stirling's formula (see, e.g., [4], p. 484) $n! \approx \sqrt{2\pi n}\,(\frac{n}{e})^n$, we see that, for n even,

$$n!! = 2^{n/2}\left(\frac{n}{2}\right)! \approx \sqrt{\pi n}\left(\frac{n}{e}\right)^{\frac{n}{2}},$$

and for n odd,

$$n!! = \frac{(n+1)!}{(n+1)!!} \approx \sqrt{2}\left(\frac{n+1}{e}\right)^{\frac{n+1}{2}}.$$

This means

$$\widehat{K}_n = |Q_n(0)| = \frac{(n-3)!!}{n!!} = \frac{(n-1)!!}{(n-1)n!!} \approx \sqrt{\frac{2}{\pi n(n-1)^2}},$$

and $\widehat{K}_{2m}/\overline{K}_{2m} \approx \sqrt{\frac{2m-2}{2m}} \to 1$ as $m \to \infty$. Thus, $\overline{Q}_{2m}(x) \approx \widehat{Q}_{2m}(x)$.

When n is odd, $Q_n(0) = 0$. There may not seem to be any more obvious normalizing factor than $\overline{K}_n$ in this case. However, considering (11) and (12), we could look for the highest peak on the graph of Q_n, at $\hat{x}_n$, the smallest positive x for which $\cos((n-1/2)\arccos x - 3\pi/4) = 1$ and then try $\widehat{K}_n = Q_n(\hat{x}_n)$. It is easy to see $\hat{x}_n = \cos\left(\frac{\pi(8k+3)}{2(2n-1)}\right)$, where $k = \left\lfloor \frac{n-2}{4} \right\rfloor$. Since $\sqrt[4]{1-\hat{x}_n^2} \to 1$ as $n \to \infty$, $\widehat{K}_n = Q_n(\hat{x}_n) \approx \overline{K}_n$, for n odd (see (11) and (12)). Hence, $\overline{Q}_{2m+1}(x) \approx \widehat{Q}_{2m+1}(x)$. FIGURE 5 shows some examples of $\overline{Q}_n$ and $\widehat{Q}_n$ plotted with the asymptotic envelope $\pm\sqrt[4]{1-x^2}$.

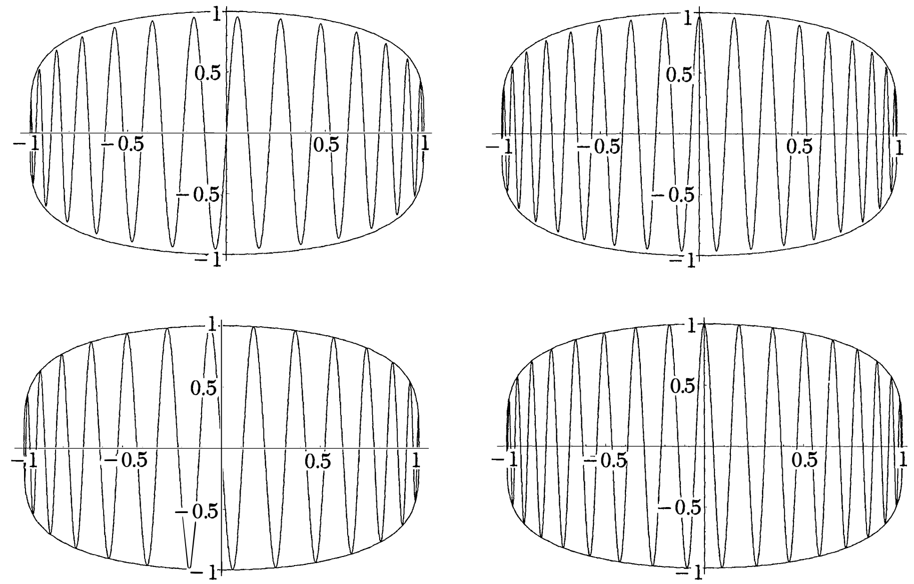


FIGURE 5

Clockwise, from the top left, $\overline{Q}_{29}(x)$, $\overline{Q}_{36}(x)$, $\widehat{Q}_{36}(x)$, and $\widehat{Q}_{29}(x)$, along with $\pm\sqrt[4]{1-x^2}$

Recalling $q_{n-2}(x) = Q_n(x)/(1-x^2)$, we can define

$$\hat{q}_{2k}(x) = \frac{q_{2k}(x)}{q_{2k}(0)} \quad \text{and} \quad \hat{q}_{2k+1}(x) = \frac{q_{2k+1}(x)}{q_{2k+1}(\hat{x}_{2k+3})}.$$

Thus, $(1-x^2)^{-3/4}$ is an asymptotic envelope for $\hat{q}_n(x)$ on $(-1,1)$. FIGURE 6 shows some sample graphs of $\hat{q}_n$ plotted along with this envelope.

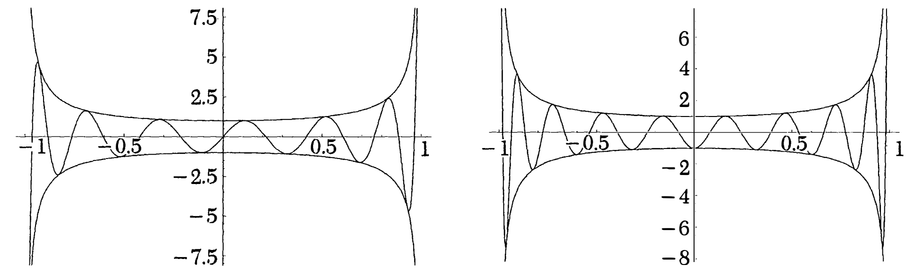


FIGURE 6

Graphs of $\hat{q}_{13}(x)$ and $\hat{q}_{18}(x)$, along with $\pm\sqrt[4]{(1-x^2)^{-3}}$

REFERENCES

1. B. Anderson, Polynomial root dragging, *Amer. Math. Monthly* 100 (1993), 864–866.
2. B. Anderson, Where the inflection points of a polynomial may lie, this MAGAZINE 70 (1997), 32–39.
3. P. Andrews, Where not to find the critical points of a polynomial—variation on a Putnam theme, *Amer. Math. Monthly* 102 (1995), 155–157.
4. W. Kaplan, *Advanced Calculus, Fourth Ed.*, Addison-Wesley, Reading, MA, 1991.
5. O. Farrell and B. Ross, *Solved Problems: Gamma and Beta Functions, Legendre Polynomials, Bessel Functions*, Macmillan, New York, NY, 1963.
6. G. Szegö, *Orthogonal Polynomials*, American Mathematical Society, Providence, RI, 1939.
7. P. Walker, Separation of the zeros of polynomials, *Amer. Math. Monthly* 100 (1993), 272–273.

NOTES

Beyond Monge's Theorem: A Generalization of the Pythagorean Theorem

EISSO J. ATZEMA
University of Maine
Orono, ME 04469-5752

It is a well-known fact from linear algebra that the volume of an n-dimensional parallelepiped in $\mathbb{R}^n$ spanned by vectors $\mathbf{A}_1, \ldots, \mathbf{A}_n$ is given by the (absolute value of the) determinant of the matrix with the vectors $\mathbf{A}_1, \ldots, \mathbf{A}_n$ as rows or columns. Although perhaps not all students realize this, it is this property that is behind those pesky integral transformations in (advanced) calculus that one has to go through to determine the volume of almost any object in 3-space that is more complicated than a sphere or a cylinder.

But what if, rather than a parallelepiped that has the same dimensions as the space it lives in, we have a parallelepiped that is spanned by fewer vectors? In the linear space spanned by the vectors that define such a parallelepiped, this parallelepiped will have a volume as well, but there does not seem an easy way to determine such lower-dimensional volume. In other words, if instead of our n-dimensional parallelepiped living in $\mathbb{R}^n$, we have a p-dimensional parallelepiped in $\mathbb{R}^n$, how can the (p-dimensional) volume of that parallelepiped be determined? Let us consider some examples to make it clear what we are looking for.

If we have just one vector in $\mathbb{R}^n$ $(n > 1)$, our parallelepiped will be a line segment, the (1-dimensional) "volume" of which is usually referred to as length. In this case, for any point $\mathbf{A} = (a_1, \ldots, a_n)$ in $\mathbb{R}^n$, the length of the corresponding vector $\mathbf{A}$ is defined by the relation

$$\operatorname{length}(\mathbf{A})^2 = a_1^2 + \cdots + a_n^2,$$

which is the Pythagorean theorem. It follows directly how to compute the length of an arbitrary line segment.

For two vectors in $\mathbb{R}^n$, the parallelepiped they span is more commonly known as a parallelogram and its (2-dimensional) "volume" is known as area. If the vectors live in $\mathbb{R}^3$, the determination of this area is a fairly straightforward matter. In fact, for two points $\mathbf{A} = (a_1, a_2, a_3)$, $\mathbf{B} = (b_1, b_2, b_3) \in \mathbb{R}^3$, the area of the parallelogram $\mathbf{AB}$ spanned by the vectors $\mathbf{A}, \mathbf{B}$ is given by the length of the cross product $\mathbf{A} \times \mathbf{B}$:

$$\operatorname{area}(\mathbf{AB})^2 = (b_3 a_2 - a_3 b_2)^2 + (b_1 a_3 - a_1 b_3)^2 + (b_2 a_1 - a_2 b_1)^2.$$

The three summands on the right-hand side can be viewed as squares of determinants of matrices that have the projection vectors of $\mathbf{A}$ and $\mathbf{B}$ onto the coordinate planes for their rows (or columns). In other words, these summands can be viewed as expressions for the square of the areas of the projections of the parallelogram on the three coordinate planes! It follows that the area of any plane object in space is determined by the areas of the projections of that object on the three coordinate planes. This result is sometimes referred to as Monge's theorem in honor of the French mathematician Gaspard Monge who discussed this result extensively in his *Application de l'algèbre à la géométrie* of 1802, although the property was formulated before (see **[1]**, p. 520).

A pattern seems to be emerging. To clarify our idea, we need to be more precise about some of the terminology. For $\mathbb{R}^n$, let us define a *p-dimensional coordinate plane* as a p-dimensional subspace containing p coordinate axes. In our new terminology, the length of a line segment is determined by its projections on the n 1-dimensional coordinate planes of the space, while the area of a parallelogram is determined by the area of its projections on the three 2-dimensional coordinate planes. From these examples, it does not seem a very big leap to conjecturing that the volume of a p-dimensional parallelepiped in $\mathbb{R}^n$ is determined by the volumes of its projections on the $\binom{n}{p}$ p-dimensional coordinate planes in the same manner as for the examples above. In fact, to show that our conjecture is correct, all we need is the following little-taught identity:

THEOREM 1 [Cauchy, 1815]. *For $n \times p$-matrices $A = (\mathbf{A}_1, \dots, \mathbf{A}_p)$, $B = (\mathbf{B}_1, \dots, \mathbf{B}_p)$ with $\mathbf{A}_i, \mathbf{B}_j \in \mathbb{R}^n$, define $A_{i_1 \dots i_p}$, $B_{i_1 \dots i_p}$ as the $p \times p$ matrices formed by rows $i_1, \dots, i_p$ of A and B, respectively. Then*

$$\det(A^t B) = \sum \det\left(B_{i_1 \dots i_p}\right)\det\left(A_{i_1 \dots i_p}\right),$$

where the summation is over the $\binom{n}{p}$ combinations of p numbers out of n.

Proof. We have

$$A^t B = \begin{pmatrix} \mathbf{A}_1 \cdot \mathbf{B}_1 & \cdots & \mathbf{A}_1 \cdot \mathbf{B}_p \\ \vdots & & \vdots \\ \mathbf{A}_p \cdot \mathbf{B}_1 & \cdots & \mathbf{A}_p \cdot \mathbf{B}_p \end{pmatrix},$$

where the dots between the vectors denote dot products. Writing out the dot products and using the properties of determinants gives

$$\det(A^t B) = \sum \begin{vmatrix} a_{1i_1} b_{1i_1} & \cdots & a_{1i_p} b_{pi_p} \\ \vdots & & \vdots \\ a_{pi_1} b_{1i_1} & \cdots & a_{pi_p} b_{pi_p} \end{vmatrix},$$

where the summation is taken over all permutations $i_1, \dots, i_p$ of p numbers out of n. This sum splits into sums over those permutations that involve the same p numbers. Restricting to the permutations of the first p numbers among themselves, we find

$$\sum_{\sigma \in \text{perm}(p)} \text{sgn}(\sigma) \cdot b_{1\sigma(1)} \cdots b_{p\sigma(p)} \det\left(A_{1 \dots p}\right) = \det\left(B_{1 \dots p}\right)\det\left(A_{1 \dots p}\right).$$

A similar result holds for all other sums of permutations among the same p numbers. This proves the theorem. For Cauchy's original proof, see **[2]**; proofs similar to ours are given in **[3]**, 45–48 and **[4]**, 49–52.

Note that the determinant of any orthogonal matrix O equals ± 1. Therefore, the left-hand side $\det(A^t B)$ of Cauchy's identity of Theorem 1 is invariant under orthogonal transformations and, consequently, the right-hand side of that identity is invariant under such transformations as well. Often such invariant behavior turns out to be crucial for proving equalities. Sure enough, in the case of our conjecture, the invariance just observed provides the key to a proof as well. Indeed, notice that by setting $A = B$, the right-hand side of Cauchy's identity becomes exactly the kind of expression that we conjectured for the square of the volume of the parallelepiped spanned by vectors $\mathbf{A}_1, \dots, \mathbf{A}_p$. But since this expression is invariant under orthogonal transformations, all we need do is rotate our object to a position parallel to a p-dimensional coordinate plane. The right-hand side then simply reduces to the square of one single determinant which indeed is an expression for the square of the p-dimensional volume of the parallelepiped. This proves our conjecture and we have the following generalization of the Pythagorean theorem:

THEOREM 2. *For a p-dimensional parallelepiped D contained in $\mathbb{R}^n$, let $D_{i_1 \dots i_p}$ denote the projection of D on the p-dimensional coordinate plane containing the $x_{i_1}, \dots, x_{i_p}$-axes and let* $\operatorname{Vol}(D)$ *denote the p-dimensional volume of D and likewise for the volumes of the projections of D. Then*

$$\operatorname{Vol}(D)^2 = \sum \operatorname{Vol}\left(D_{i_1 \dots i_p}\right)^2,$$

where the summation is over the $\binom{n}{p}$ p-dimensional coordinate planes in $\mathbb{R}^n$.

Note that Theorem 2 is independent of the position of the parallelepiped. Not only is the theorem true for parallelepipeds that are spanned by a set of vectors, but it is also true for any parallelepiped obtained by shifting a parallelepiped spanned by a set of vectors. In fact, the statement of Theorem 2 is true for any object that can be formed by putting together any number of (shifted) p-dimensional parallelepipeds contained in a p-dimensional plane, i.e., for any object contained in a p-dimensional plane that is the union of a finite set of (shifted) p-dimensional parallelepipeds, no two of which have overlapping interiors. To prove this claim, let D_1 be the parallelepiped obtained by shifting the parallelepiped spanned by vectors $\mathbf{A}_1, \dots, \mathbf{A}_p$ by a vector $\mathbf{a}$. Similarly, the parallelepiped D_2 is obtained by shifting the parallelepiped spanned by vectors $\mathbf{B}_1, \dots, \mathbf{B}_p$ by a vector $\mathbf{b}$. Finally, let D be the union of D_1 and D_2, where we assume that the interiors of D_1 and D_2 are disjoint. It follows that $\operatorname{Vol}(D) = \operatorname{Vol}(D_1) + \operatorname{Vol}(D_2)$ with similar relationships for the projections of these volumes onto the coordinate planes. Using the notations of Theorem 1, note that not only are $\sum \det(A_{i_1 \dots i_p})^2$ and $\sum \det(B_{j_1 \dots j_p})^2$ invariant under orthogonal transformations, but so is also the expression $\sum \det(A_{j_1 \dots j_p}) \det(B_{j_1 \dots j_p})$. Consequently, the expression

$$\begin{aligned}\sum \det\left(A_{i_1 \dots i_p}\right)^2 &+ 2\det\left(A_{j_1 \dots j_p}\right)\det\left(B_{j_1 \dots j_p}\right) + \det\left(B_{j_1 \dots j_p}\right)^2 \\ &= \sum \left(\det\left(A_{j_1 \dots j_p}\right) + \det\left(B_{j_1 \dots j_p}\right)\right)^2\end{aligned}$$

is invariant under orthogonal transformations as well. By choosing the same orientation for $\mathbf{A}_1, \dots, \mathbf{A}_p$ and $\mathbf{B}_1, \dots, \mathbf{B}_p$, it follows that the expression

$$\sum \left(\operatorname{Vol}\left(D_{1, i_1 \dots i_p}\right) + \operatorname{Vol}\left(D_{2, i_1 \dots i_p}\right)\right)^2 = \sum \operatorname{Vol}\left(D_{i_1 \dots i_p}\right)^2$$

is also invariant under orthogonal transformations. By the same argument as used in the proof of Theorem 2, we conclude that Theorem 2 is true for the union of two parallelepipeds as well, provided their interiors do not intersect. A similar proof applies to the case of an object that is the union of an arbitrary number of parallelepipeds, provided their interiors do not intersect and that they all are in the same p-dimensional plane. Indeed, although a proof goes slightly beyond the scope of this note, it can be shown that Theorem 2 even applies to any "sufficiently nice" object D contained in a p-dimensional subspace.

Acknowledgment. The author wishes to thank Bob Franzosa, Phil Locke, and Judy O'Neal for their critical comments as well as the two anonymous referees for their helpful suggestions.

REFERENCES

1. Ch. Boyer (with U. Merzbach), *A History of Mathematics*, 2nd edition, Wiley, New York, NY, 1968.
2. A. Cauchy, Mémoire sur les fonctions qui ne peuvent obtenir que deux valeurs égales et de signes contraires par suite des transpositions opérées entre les variables qu'elles renferment, *Journal de l'Ecole Polytechnique* 17 (1815), 29–112 (= *Oeuvres*, T. 2, art.1).
3. H. Dorrie, *Determinanten*, R. Oldenbourg, München/Berlin, 1940.
4. Th. Hanus, *An Elementary Treatise on the Theory of Determinants, A Text Book for Colleges*, Ginn and Company, Boston, MA, 1888.

Quadrilaterals with Integer Sides

CHARLES JEPSEN
JENNIFER HEPPNER
DANIEL WILLMS
Grinnell College
Grinnell, IA 50112-1690

Given a positive integer n, how many incongruent triangles are there with integer sides and perimeter n? Denoting the number of such triangles by $T(n)$, one answer is given by

$$T(2n-3) = T(2n) = \left\{ \frac{n^2}{12} \right\},$$

where $\{x\}$ is the nearest integer to x. See **[1]**–**[5]** for different solution procedures. In the last section of **[4]**, the authors raise the problem of determining the number of incongruent quadrilaterals with a prescribed perimeter. Here restrictions are needed since a quadrilateral is not uniquely determined by its four sides. We solve this problem in the special cases where the quadrilaterals are:

(a) cyclic (i.e., the vertices lie on a circle);
(b) trapezoids.

In the latter case, we must exclude parallelograms (because there are infinitely many with the same side lengths). In the former case, we choose to exclude rectangles because then the formulas work out a little nicer. The number of rectangles is easy to count directly and can be included in the count if so desired.

Denote by $CQ(n)$ the number of (nonrectangular) cyclic quadrilaterals with integer sides and perimeter n and by $TR(n)$ the number of such trapezoids (that are not parallelograms). For example, $CQ(7) = 3$ and $TR(7) = 3$. (See FIGURE 1.) We obtain formulas for these counting numbers by finding relationships among them (equalities I–III below) and then evaluating $TR(2n)$ using a generating function. (See, e.g., [6] for a discussion of generating functions.)

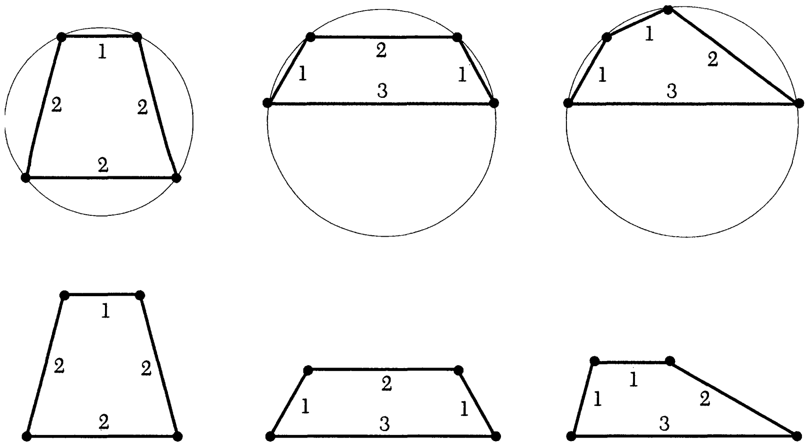


FIGURE 1

Our starting point is a useful description of quadrilaterals. The four sides of a quadrilateral can be given in clockwise order: a, b, c, d. Note that

$$(1)\ a+b+c>d \quad (2)\ a+b+d>c \quad (3)\ a+c+d>b \quad (4)\ b+c+d>a.$$

(It is convenient to use the same letter to denote both the side of a quadrilateral and its length.) By reflecting the quadrilateral about a side or a diagonal, we may assume

$$(5)\ c \geq a \quad (6)\ d \geq b \quad (7)\ a+d \geq b+c.$$

Not all of these conditions are needed. In fact, conditions (2), (3), (4), and (6) follow from (5) and (7). For example, $a+d \geq b+c$ implies $a+b+d > a+d \geq b+c > c$. Also, from $c \geq a$ and $a+d \geq b+c$, we get $d \geq b+c-a \geq b$. Indeed, since parallelograms are excluded, we see that $d > b$. (For $d = b$ implies $c = a$.)

We use (a, b, c, d) to designate a quadrilateral with sides a, b, c, d in clockwise order and satisfying conditions (1), (5), (7):

$$\begin{cases} a+b+c>d \\ c \geq a \\ a+d \geq b+c. \end{cases} \qquad (*)$$

Note that in the case where $a+d = b+c$, both (a, b, c, d) and (b, a, d, c) give the same quadrilateral. We choose (a, b, c, d) if $b \geq a$.

LEMMA. *Suppose a, b, c, d are positive integers satisfying $(*)$. Then:*

(i) *There is a unique cyclic quadrilateral designated by (a, b, c, d).*
(ii) *There is a trapezoid designated by (a, b, c, d) with parallel sides b and d if and only if $a+d > b+c$. If this condition holds, the trapezoid is unique.*

Proof. To prove (i), consider a line segment of length d with line segments a and c attached to its end points. There are two cases: $a+c > d$, $a+c \leq d$. We prove the first case; the second is similar. In the first case, a, c, d are the sides of a triangle. (The remaining triangle inequalities hold: $a+d \geq b+c > c$, $c+d > c \geq a$.) Beginning with the triangle, rotate line segment c clockwise so the angle θ between c and d increases. (See FIGURE 2.) Form the circle through the three endpoints of line segments a and d. The arc of this circle that lies above d contains an inscribed quadrilateral three of whose sides are a, c, d. Label the fourth side x. From elementary geometry, x is a function of θ given by

$$x^2 + 2(ax+cd)\cos\theta = c^2 + d^2 - a^2.$$

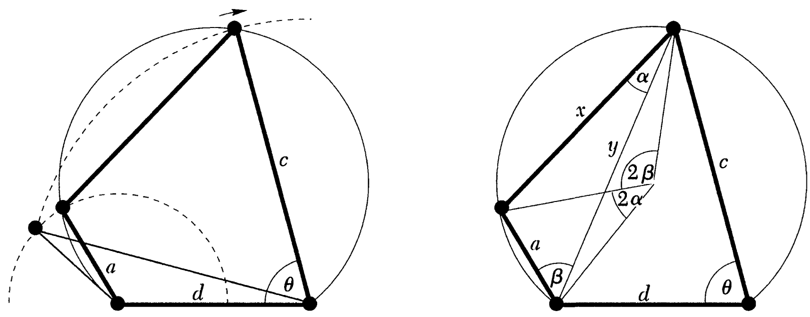


FIGURE 2

(With the labels in FIGURE 2, $\theta = \alpha + \beta$ so two applications of the Law of Cosines gives $y^2 = c^2 + d^2 - 2cd\cos\theta = a^2 + x^2 + 2ax\cos\theta$ and the above equality follows.) By implicit differentiation,

$$(x + a\cos\theta)\frac{dx}{d\theta} = (ax+cd)\sin\theta.$$

Multiply both sides by $2(ax+cd)$, expand and simplify the left hand side, and get

$$\left(ax^2 + 2cdx + a(c^2+d^2-a^2)\right)\frac{dx}{d\theta} = 2(ax+cd)^2\sin\theta.$$

This implies that $\frac{dx}{d\theta}$ is positive for θ between its initial value and π, so x increases with θ and assumes every value between 0 and $a+d+c$. Hence there is a unique angle θ so that $x = b$. This angle θ gives the unique cyclic quadrilateral (a, b, c, d).

To prove (ii), suppose $a+d > b+c$ so that $a+(d-b) > c$. There is a unique triangle with sides $a, d-b, c$. (Again the remaining two triangle inequalities hold.) Extending side $d-b$ to a side of length d and adding a parallel side of length b, we obtain the trapezoid (a, b, c, d). (See FIGURE 3.) The uniqueness of the trapezoid follows from the uniqueness of the triangle. The converse follows by reversing the steps in the above argument. Note that trapezoid (a, b, c, d) is a cyclic quadrilateral if and only if $c = a$. This completes the proof. ∎

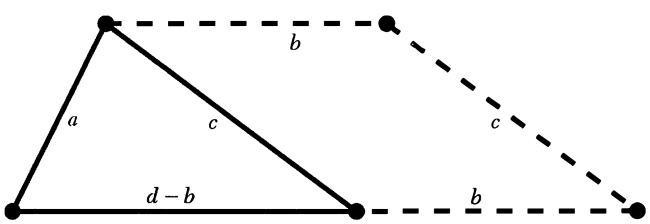


FIGURE 3

Statements (i) and (ii) imply that $CQ(n)$ and $TR(n)$ are well-defined, finite numbers. Moreover, they yield the relationships among the numbers given in I–III below. The first is immediate. A quadrilateral (a, b, c, d) with $a+d=b+c$ has even perimeter $a+b+c+d=2(a+d)$. So a quadrilateral of odd perimeter satisfies $a+d>b+c$. Hence:

I. $CQ(2n-1)=TR(2n-1)$.

REMARK. For even perimeter, the difference $CQ(2n)-TR(2n)$ is the number of quadrilaterals (a, b, c, d) with $c \geq a$, $b \geq a$, and $a+d=b+c$. This is the number of solutions of $a+d=n$ and $b+c=n$ where $a \leq b < d$. A direct count gives

$$CQ(2n)-TR(2n)=\begin{cases} \frac{1}{4}n(n-2), & \text{if } n \text{ is even} \\ \frac{1}{4}(n-1)^2, & \text{if } n \text{ is odd.} \end{cases} \qquad (**)$$

We derive this result in a different way below.

Our second relationship is the same one that holds for triangles:

II. $TR\,(2n-3)=TR\,(2n)$.

Proof. Let (a, b, c, d) be a trapezoid of perimeter $2n$ so that $a+b+c>d$, $c \geq a$, $a+d>b+c$. We show that $(a-1, b, c-1, d-1)$ is a trapezoid of perimeter $2n-3$. First note that $a>1$. For if $a=1$, then $a+d>b+c>d-a$ gives $1+d>b+c>d-1$ so that $b+c=d$. But this says $a+b+c+d=1+2d$, which is impossible since $a+b+c+d=2n$. From this, we have $c>1$, and $d>1$ because $d>b$. To show $(a-1, b, c-1, d-1)$ designates a trapezoid, we must show

$$(a-1)+b+(c-1)>d-1, \quad c-1 \geq a-1, \quad (a-1)+(d-1)>b+(c-1).$$

The second inequality is obvious. To show the first, note that $a+b+c>d$ implies $(a-1)+b+(c-1)>d-2$ so $(a-1)+b+(c-1) \geq d-1$. If $(a-1)+b+(c-1)=d-1$, then $a+b+c+d=1+2d$, again impossible. In a similar manner, $(a-1)+(d-1)>b+(c-1)$ follows from $a+d>b+c$.

Conversely, if (p, q, r, s) is a trapezoid of perimeter $2n-3$, it is easy to check that $(p+1, q, r+1, s+1)$ is a trapezoid of perimeter $2n$. These correspondences are inverses of each other, and we have equality II.

III. $CQ(2n-1)=CQ(2n)$.

Proof. Let (a, b, c, d) be a cyclic quadrilateral of perimeter $2n$ and consider two cases: either $a+d>b+c$ or $a+d=b+c$. If $a+d>b+c$, we find that $(a-1, b+1, c-1, d)$ is a quadrilateral of perimeter $2n-1$. The argument is very similar to one given above and is omitted. If $a+d=b+c=n$, we show that $(c-1, 1, d-1, 2n-c-d)$ is a cyclic quadrilateral of perimeter $2n-1$. Note that $c>1$. For if $c=1$, then $a=1$ and $d=b=n-1$. But this says the given quadrilateral is a rectangle, which is not the case. Similarly $d>1$. We must show

$$(c-1)+1+(d-1)>2n-c-d, \quad d-1 \geq c-1,$$
$$(c-1)+(2n-c-d)>1+(d-1).$$

In this case, $b \geq a$ so $d-c=b-a \geq 0$ and the second inequality holds. The first inequality is equivalent to $2c+2d>2n+1$ which in turn is equivalent to $c+d \geq n+1$. Now $c \geq a$ implies $c+d \geq a+d=n$. If $c+d=n$, then $a+d=b+c=n$ says that $a=c$ and $b=d$, which is not true because the given quadrilateral is not a rectangle. Hence $c+d \geq n+1$. Finally, the third inequality is equivalent to $2n>2d+1$ which is equivalent to $n \geq d+1$, and this is true because $n=a+d \geq 1+d$.

Conversely, let (p, q, r, s) be a cyclic quadrilateral of perimeter $2n-1$, and consider two cases: either $q > 1$ or $q = 1$. If $q > 1$, we check that $(p+1, q-1, r+1, s)$ designates a cyclic quadrilateral of perimeter $2n$. (The proof is similar to previous ones.) If $q = 1$, we show that $(n-r-1, n-p-1, p+1, r+1)$ is a cyclic quadrilateral of perimeter $2n$. The inequality $(n-r-1)+(n-p-1)+(p+1) > r+1$ is equivalent to $n > r+1$. But $p+s > q+r$ says $p+(2n-2-p-r) > 1+r$ so $2n > 2r+3 > 2r+2$ and hence $n > r+1$. The remaining inequalities are proved similarly. These two correspondences are defined so that they are inverses of each other. This gives equality III.

It follows from I, II, and III that to find $CQ(n)$ and $TR(n)$ for all n, it suffices to find $TR(2n)$. We do this using a generating function. $TR(2n)$ counts the number of solutions of $a+b+c+d = 2n$ where a, b, c, d are positive integers such that $a+b+c > d$, $c \geq a$, and $a+d > b+c$. Let $u = a+b+c-d$, $w = c-a$, $y = a+d-b-c$. Note that $u+y = 2a$ and $b+w+y = d$. Then

$$\begin{aligned}2n &= a+b+c+d = a+b+(w+a)+(b+w+y) = 2b+2a+2w+y\\ &= 2b+u+2w+2y.\end{aligned}$$

This says u is even, say $u = 2v$, so $b+v+w+y = n$. Also, since $u+y = 2v+y$ is even, y must be even.

It is easy to check that the number of solutions of $a+b+c+d = 2n$ where $a+b+c > d, c \geq a, a+d > b+c$ is the same as the number of solutions of $b+v+w+y = n$ where $b > 0$, $v > 0$, $w \geq 0$, $y > 0$, and y is even. Hence the generating function for $TR(2n)$ is

$$\begin{aligned}f(x) &= (x+x^2+\cdots)(x+x^2+\cdots)(1+x+x^2+\cdots)(x^2+x^4+\cdots)\\ &= x^4(1+x+x^2+\cdots)^3(1+x^2+x^4+\cdots)\\ &= \frac{x^4}{(1-x)^3(1-x^2)} = x^4\left(\frac{1}{(1-x)^4(1+x)}\right) = x^4 g(x).\end{aligned}$$

Resolving $g(x)$ into its partial fractions yields

$$g(x) = \frac{1/2}{(1-x)^4} + \frac{1/4}{(1-x)^3} + \frac{1/8}{(1-x)^2} + \frac{1/16}{1-x} + \frac{1/16}{1+x}.$$

Now $TR(2n)$ is the coefficient of x^{n-4} in $g(x)$. Using the binomial expansion on expressions of the form $(1-x)^{-k}$, we find this coefficient to be

$$\frac{1}{12}(n-3)(n-2)(n-1) + \frac{1}{8}(n-3)(n-2) + \frac{1}{8}(n-3) + \frac{1}{16} + \frac{1}{16}(-1)^{n-4},$$

which simplifies to

$$\frac{1}{24}(n-3)(n-1)(2n-1) + \frac{1}{16} + \frac{1}{16}(-1)^n.$$

If n is odd, the last two terms cancel. If n is even, we get

$$\frac{1}{24}(n-3)(n-1)(2n-1) + \frac{1}{8} = \frac{1}{24}n(n-2)(2n-5).$$

We have reached our conclusion:

$$TR(2n) = \begin{cases} \frac{1}{24}n(n-2)(2n-5), & \text{if } n \text{ is even} \\ \frac{1}{24}(n-3)(n-1)(2n-1), & \text{if } n \text{ is odd.} \end{cases}$$

REMARKS. The same generating function argument seems not to work to count $CQ(2n)$; the case where $a+d=b+c$ causes difficulty. Our results also imply equation $(**)$ above:

$$\begin{aligned} CQ(2n) - TR(2n) &= TR(2(n+1)) - TR(2n) \\ &= \begin{cases} \frac{1}{4}n(n-2), & \text{if } n \text{ is even} \\ \frac{1}{4}(n-1)^2, & \text{if } n \text{ is odd.} \end{cases} \end{aligned}$$

Explicit formulas for $CQ(n)$ and $TR(n)$ are as follows:

$$CQ(n) = \begin{cases} \frac{1}{96}(n+1)n(n-4), & \text{if } n \equiv 0 \pmod 4 \\ \frac{1}{96}(n+3)(n-1)(n-2), & \text{if } n \equiv 1 \pmod 4 \\ \frac{1}{96}(n+2)(n-2)(n-3), & \text{if } n \equiv 2 \pmod 4 \\ \frac{1}{96}(n+2)(n+1)(n-3), & \text{if } n \equiv 3 \pmod 4 \end{cases}$$

$$TR(n) = \begin{cases} \frac{1}{96}n(n-4)(n-5), & \text{if } n \equiv 0 \pmod 4 \\ \frac{1}{96}(n+3)(n-1)(n-2), & \text{if } n \equiv 1 \pmod 4 \\ \frac{1}{96}(n-1)(n-2)(n-6), & \text{if } n \equiv 2 \pmod 4 \\ \frac{1}{96}(n+2)(n+1)(n-3), & \text{if } n \equiv 3 \pmod 4 \end{cases}$$

Acknowledgment. This paper is the result of a research project carried out in Summer 1998 by the latter two authors under the direction of the first author. In that research project, the second author found two different derivations of the above formulas for $TR(n)$, and the third author counted $CQ(n)$ directly. These results led us to observe relations I–III and to find the proofs that yield the approach in this note.

REFERENCES

1. George Andrews, A note on partitions and triangles with integer sides, *Amer. Math. Monthly* 86 (1979), 477–478.
2. Michael Hirschhorn, Triangles with integer sides, revisited, this MAGAZINE 73 (2000), 53–56.
3. Ross Honsberger, *Mathematical Gems III*, Mathematical Association of America, Washington, DC, 1985, 39–47.
4. J. H. Jordan, R. Walch, and R. J. Wisner, Triangles with integer sides, *Amer. Math. Monthly* 86 (1979), 686–689.
5. N. Krier and B. Manvel, Counting integer triangles, this MAGAZINE 71 (1998), 291–295.
6. Herbert Wilf, *Generatingfunctionology*, 2nd edition, Academic Press, San Diego, CA, 1994.

Math Bite: Once in a While, Differentiation Is Multiplicative

Many students of calculus would be a lot happier if the Leibniz formulas

$$(fg)' = f'g + fg' \quad \text{and} \quad \left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2},$$

where f' denotes the derivative of f, could be replaced by the much simpler formulas

$$(fg)' = f'g' \tag{1}$$

and

$$\left(\frac{f}{g}\right)' = \frac{f'}{g'}. \tag{2}$$

Discovering exactly when the usually erroneous equations (1) and (2) are valid is a simple but neat exercise involving three separable differential equations. Paul Zorn called my attention to the article [**1**], in which the function f is fixed in (1), and the one-dimensional subspace of all corresponding g is determined. Fixing g in (2) and then determining f involves essentially the same calculation, while fixing f in (2) and finding g leads to the formula

$$g(x) = C\exp\left(\int \frac{f' \pm \sqrt{(f')^2 - 4ff'}}{2f}\,dx\right). \tag{3}$$

More concrete problems arise when these formulas are used to find companions for simple concrete choices of functions. For example, setting $f(x) = x^r$ in (3) produces

$$\left\{\frac{x^r}{Cx^{r/2}\left(\frac{r-\sqrt{r^2-4rx}}{r+\sqrt{r^2-4rx}}\right)^{\pm r/2}\exp\left(\pm\sqrt{r^2-4rx}\right)}\right\}'$$

$$= \frac{\{x^r\}'}{\left\{Cx^{r/2}\left(\frac{r-\sqrt{r^2-4rx}}{r+\sqrt{r^2-4rx}}\right)^{\pm r/2}\exp\left(\pm\sqrt{r^2-4rx}\right)\right\}'}.$$

Several other examples can be found in [**1**].

Another question is whether we can find a pair of functions $\{p,q\}$, neither identically zero, such that at least two of the relations

$$\text{(i)}\ \left(\frac{p}{q}\right)' = \frac{p'}{q'} \qquad \text{(ii)}\ \left(\frac{q}{p}\right)' = \frac{q'}{p'} \qquad \text{(iii)}\ (pq)' = p'q'$$

hold simultaneously. Any such pair solving (iii) satisfies neither (i) nor (ii), and there are very few simultaneous solutions to (i) and (ii), namely the pairs

$$\left\{ce^{\frac{1+i}{2}x},\ de^{\frac{1-i}{2}x}\right\},$$

where $i = \sqrt{-1}$ and c and d are arbitrary non-zero constants.

REMARKS. An easy related exercise for beginning calculus students is to find all pairs of *polynomials* (f, g) such that $(fg)' = f'g'$. There are many similar questions. For example, given f and g, it is easy to find all functions h such that $(fgh)' = f'g'h'$.

My interest in this question was motivated by Exercise 2 on page 545 of [**2**].

REFERENCES

1. L. G. Maharam and E. P. Shaughnessy, When does $(fg)' = f'g'$?, *College Math. J.* 7 (1976), 38–39.
2. J. Stewart, *Calculus*, 3^{rd} ed., Brooks/Cole, Pacific Grove, CA, 1995.

—J. MARSHALL ASH
DEPAUL UNIVERSITY
CHICAGO, IL 60614

Generalizing Van Aubel Using Duality

MICHAEL DE VILLIERS
University of Durban – Westville
South Africa

A recent paper by DeTemple and Harold [1] elegantly utilized the Finsler – Hadwiger theorem to prove Van Aubel's theorem, which states that the line segments, connecting the centers of squares constructed on the opposite sides of a quadrilateral, are congruent and lie on perpendicular lines. This result can easily be generalized by using a less well known "duality" between the concepts *angle* and *side* within Euclidean plane geometry.

Although similar to the general duality between *points* and *lines* in projective geometry, this "duality" is limited. Nevertheless, it occurs quite frequently and examples of these are explored fairly extensively in [2]. Obviously this "duality" does not apply to theorems related to or based on the Fifth Postulate (compare [7]). For example, the dual to the theorem "three corresponding sides of two triangles equal imply their congruency," namely, "three corresponding angles of two triangles equal imply their congruency," is not valid. (Note, however, that the dual is perfectly true in both non-Euclidean geometries).

The square is self-dual regarding these concepts as it has all angles and all sides congruent. The parallelogram is also self-dual since it has both opposite sides and opposite angles congruent. Similarly, the rectangle and rhombus are each other's duals as shown in the table below:

Rectangle	Rhombus
All *angles* congruent	All *sides* congruent
Center equidistant from *vertices*, hence has *circum*circle	Center equidistant from *sides*, hence has *in*circle
Axes of symmetry bisect opposite *sides*	Axes of symmetry bisect opposite *angles*

Furthermore, the *congruent* diagonals of the rectangle has as its dual the *perpendicular* diagonals of the rhombus and is illustrated by the following two elementary results:

(1) The midpoints of the sides of any quadrilateral with congruent diagonals form a rhombus.
(2) The midpoints of the sides of any quadrilateral with perpendicular diagonals form a rectangle.

The following two dual generalizations of Van Aubel's theorem are proved in [3] by generalizing the transformation approach in [1]. A vector proof and a slightly different transformation proof for the same generalizations are respectively given in [2] and [4].

THEOREM 1. *If similar rectangles with centers E, F, G and H are erected externally on the sides of quadrilateral $ABCD$ as shown in* FIGURE 1, *then the segments EG and FH lie on perpendicular lines. Further, if J, K, L and M are the midpoints of the dashed segments shown, then JL and KM are congruent segments, concurrent with the other two lines.*

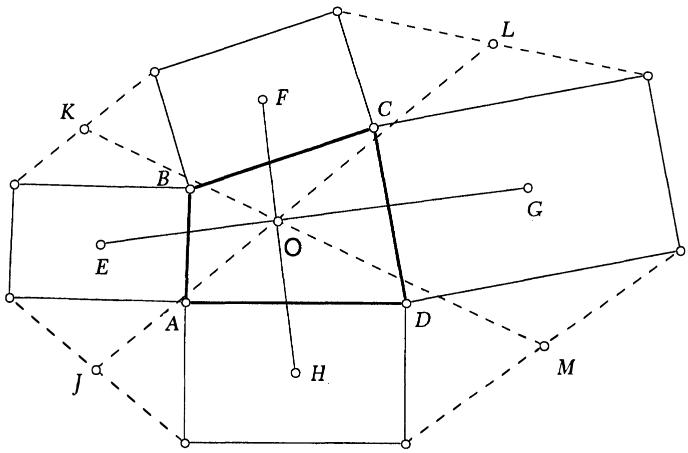


FIGURE 1

THEOREM 2. *If similar rhombi with centers E, F, G and H are erected externally on the sides of quadrilateral ABCD as shown in* FIGURE 2, *then the segments EG and FH are congruent. Further, if J, K, L and M are the midpoints of the dashed segments shown, then JL and KM lie on perpendicular lines.*

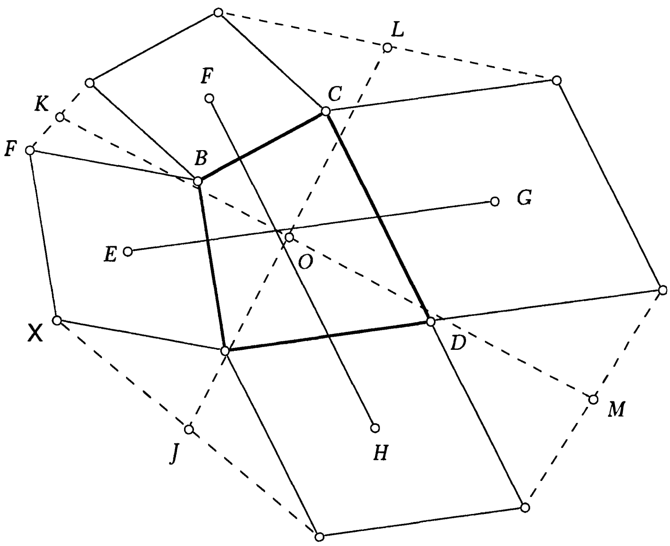


FIGURE 2

To Theorem 1 the following two properties can be added:

(a) *the ratio of EG and FH equals the ratio of the sides of the rectangles*
(b) *the angle of JL and KM equals the angle of the diagonals of the rectangles*

and to Theorem 2, the following corresponding duals:

(a) *the angle of EG and FH equals the angle of the sides of the rhombi*
(b) *the ratio of JL and KM equals the ratio of the diagonals of the rhombi.*

By combining Theorems 1 and 2, we obtain Van Aubel's theorem, just as the squares are the intersection of the rectangles and rhombi. (For example, for squares it gives us segments JL and KM, and EG and FH, respectively congruent and lying on perpendicular lines, as well as concurrent in a single point. In addition, it also follows that all eight angles at the point of intersection are congruent.)

The latter four properties are also contained in the following self-dual generalization, which can be proved by using vectors, by complex algebra, or by generalizing the transformation approach used in [**3**]:

THEOREM 3. *If similar parallelograms with centers E, F, G and H are erected externally on the sides of quadrilateral $ABCD$ as shown in* FIGURE 3, *then* $\frac{FH}{EG} = \frac{XY}{YB}$, *and the angle of EG and FH equals the angle of the sides of the parallelograms. Further, if J, K, L and M are the midpoints of the dashed segments shown, then* $\frac{KM}{JL} = \frac{YA}{XB}$, *and the angle of JL and KM equals the angle of the diagonals of the parallelograms.*

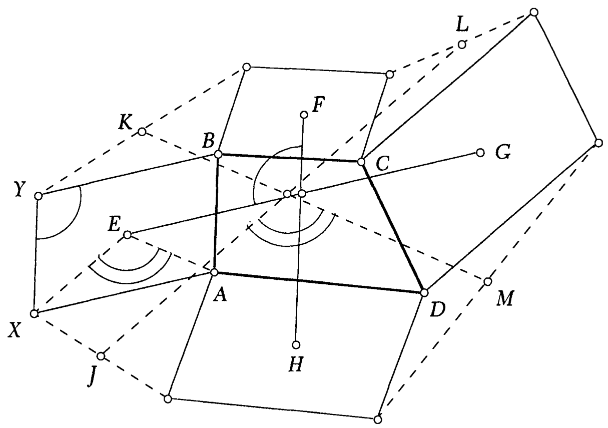


FIGURE 3

The latter theorem can be further generalized into the following two dual theorems using ideas of Friedrich Bachmann which are extensively developed in [**5**] and [**6**], and which also provide a powerful technique and notation, giving automatic proofs for problems of this kind.

THEOREM 4. *If similar parallelograms are erected externally on the sides of quadrilateral $ABCD$ and similar triangles*

$$XP_0A,\ AP_0'B,\ QP_1B,\ BP_1'C,\ RP_2C,\ CP_2'D,\ SP_3D,\ DP_3'A$$

are constructed as shown in FIGURE 4, *and E, F, G and H are the respective midpoints of the segments P_iP_{i+1}' for $i = 0, 1, 2, 3$, then* $\frac{FH}{EG} = \frac{XY}{YB}$, *and the angle of EG and FH equals the angle of the sides of the parallelograms.*

THEOREM 5. *If similar parallelograms with centers E, F, G, and H are erected externally on the sides of quadrilateral $ABCD$ and I_i are the midpoints of the dashed segments as shown in* FIGURE 5, *parallelograms are constructed with I_i as centers as well as similar triangles TP_0E, $EP_0'F$, QP_1F, $FP_1'G$, RP_2G, $GP_2'H$, SP_3H, $HP_3'E$, and K, L, M and J are the respective midpoints of the segments P_iP_{i+1}' for $i = 0, 1, 2, 3$, then* $\frac{KM}{JL} = \frac{YA}{XB}$, *and the angle of JL and KM equals the angle of the diagonals of the parallelograms.*

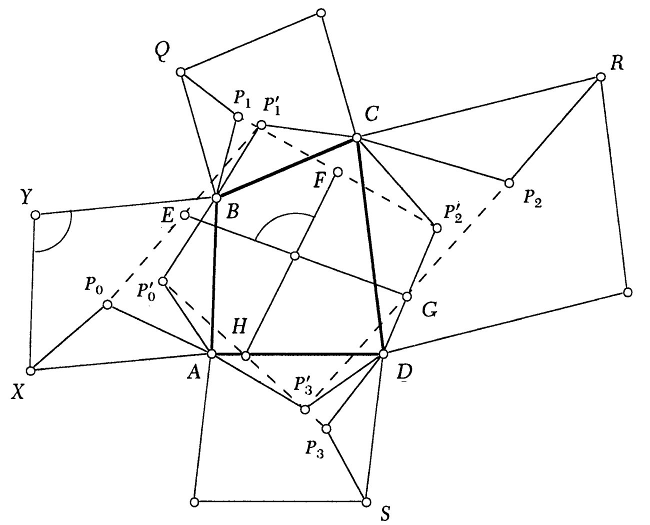


FIGURE 4

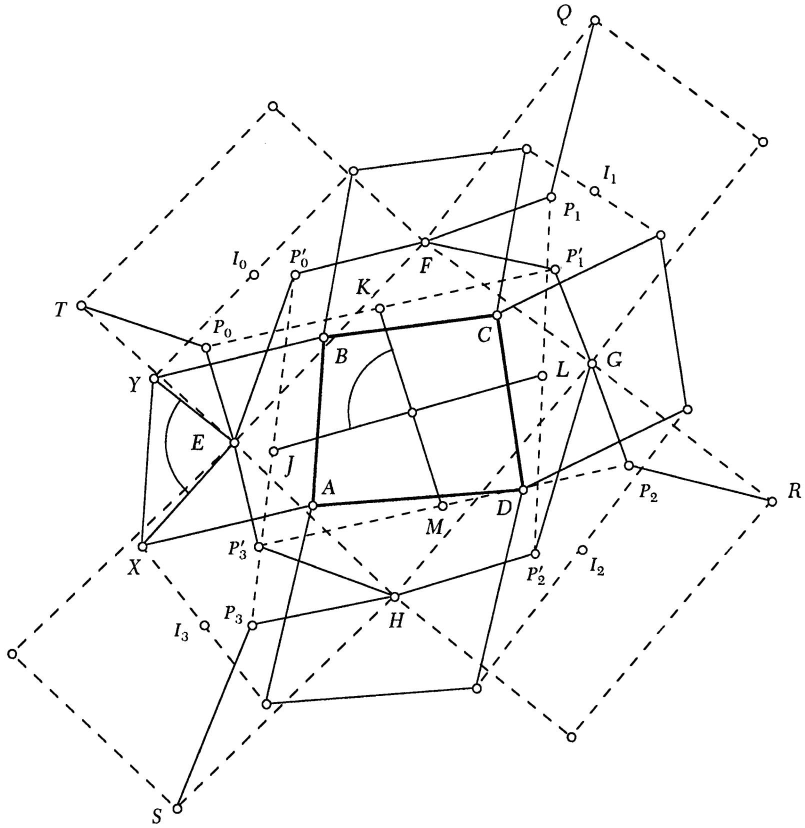


FIGURE 5

Acknowledgment. I am indebted to Hessel Pot from Woerden in the Netherlands who in a personal communication to me in 1997 pointed out the additional properties to Theorems 1 and 2, as well as Theorem 3. Thanks also to Chris Fisher, University of Regina, Canada, whose technique and own generalization of Van Aubel's theorem (first communicated to me via e-mail in 1998) in combination with Theorem 3, led to Theorems 4 and 5.

REFERENCES

1. D. DeTemple, S. Harold, A round-up of square problems, this MAGAZINE 16 (1996), 15–27.
2. M. de Villiers, *Some Adventures in Euclidean Geometry*, University of Durban-Westville, Durban, South Africa, 1996.
3. M. de Villiers, Dual generalizations of Van Aubel's theorem, *The Mathematical Gazette*, November (1998), 405–412.
4. M. de Villiers, The Role of Proof in Investigative, Computer-based Geometry: Some Personal Reflections, In D. Schattschneider, J. King (Eds.), *Geometry Turned On*, Math. Assoc. of America, Washington DC, 1997.
5. J. C. Fisher, D. Ruoff, J. Shilleto, Polygons and Polynomials, In C. Davis, B. Grunbaum, F. A. Sherk, *The Geometric Vein*, Springer-Verlag, New York, NY, 1982.
6. J. C. Fisher, D. Ruoff, J. Shilleto, Perpendicular polygons, *Amer. Math. Monthly*, 92 (Jan. 1985), 23–37.
7. E. G. Kogbetliantz, *Fundamentals of Mathematics from an Advanced Viewpoint*, Vols. 3&4, Gordon & Breach Science Publishers, New York, NY, 1969.

The Abundancy Ratio, a Measure of Perfection

PAUL A. WEINER
Saint Mary's University of Minnesota
Winona, MN 55987-1399

Introduction The concept of perfect numbers originated with the ancient Greeks. In modern terminology, a positive integer n is *perfect* if $\sigma(n)=2n$, where $\sigma(n)$ denotes the sum of the positive divisors of n. Euclid showed that an even number is perfect if it has the form $2^{p-1}(2^p-1)$, where both p and 2^p-1 are primes. Euler proved the converse: every even perfect number must be of Euclid's type. Therefore finding an even perfect number is equivalent to finding a prime p such that 2^p-1 is prime. Numbers of the form 2^p-1 where p is prime are now called *Mersenne numbers* in honor of the French monk Marin Mersenne (1588–1648) who studied them in 1644.

It has been conjectured that there are an infinite number of Mersenne primes, and hence an infinite number of even perfect numbers, but this conjecture remains one of the great unsolved problems of number theory. As of March 2000 only 38 Mersenne primes were known, the largest being $2^{6972593}-1$.

Another celebrated unanswered question is whether there are any *odd* perfect numbers. It is known [1] that there are none less than 10^{300}, which helps to explain why none have been found.

A history of perfect numbers, Mersenne primes and related topics may be found at the web site `www.utm.edu/research/primes/mersenne.shtml`.

The abundancy ratio The search for odd perfect numbers has prompted investigations concerning *abundant* numbers, those for which $\sigma(n) > 2n$, and *deficient* numbers, those for which $\sigma(n) < 2n$. This suggests studying the ratio $\frac{\sigma(n)}{n}$, sometimes called the *abundancy ratio* of n. An integer is abundant, deficient, or perfect if its abundancy ratio is respectively greater than, less than, or equal to 2.

The abundancy ratio is also the sum of the reciprocals of the divisors of n. To see why, note that as d runs through the divisors of n, so does n/d, and we have

$$\frac{\sigma(n)}{n} = \frac{1}{n}\sum_{d|n} d = \frac{1}{n}\sum_{d|n}\frac{n}{d} = \sum_{d|n}\frac{1}{d}. \tag{1}$$

Observing that if $m \mid n$ then every divisor of m is also a divisor of n, we immediately see from (1) that

$$m \mid n \quad \text{implies} \quad \frac{\sigma(m)}{m} \le \frac{\sigma(n)}{n}, \quad \text{with equality only if } m = n. \tag{2}$$

We can also use (1) to show that the abundancy ratio takes on arbitrarily large values. Because every integer from 1 to n is a divisor of $n!$, we have

$$\frac{\sigma(n!)}{n!} = \sum_{d|n!}\frac{1}{d} \ge \sum_{k=1}^{n}\frac{1}{k},$$

so $\sigma(n!)/n!$ dominates the nth partial sum of the divergent harmonic series.

The abundancy ratio is a multiplicative function of n, so it is completely determined by its values at prime powers p^a. The divisors of p^a are in geometric progression, hence (1) implies

$$\frac{\sigma(p^a)}{p^a} = \sum_{r=0}^{a} p^{-r} = \frac{1-p^{-a-1}}{1-p^{-1}} = \frac{p}{p-1} - \frac{p^{-a}}{p-1},$$

giving the upper bound

$$\frac{\sigma(p^a)}{p^a} < \frac{p}{p-1}.$$

Therefore for any integer $n > 1$, we have

$$\frac{\sigma(n)}{n} < \prod_{p|n}\frac{p}{p-1} = \prod_{p|n}\left(1 + \frac{1}{p-1}\right). \tag{3}$$

This can be used to prove Euler's theorem that the sum of the reciprocals of the primes diverges. First note that the infinite product extended over all primes p

$$\prod_{p}\left(1 + \frac{1}{p-1}\right)$$

is divergent, for if it converged the value would by (3) provide a finite upper bound for all abundancy ratios, contradicting the fact that $\sigma(n)/n$ takes on arbitrarily large values. But it is known that an infinite product $\prod(1 + a_n)$ with positive a_n converges if and only if the series Σa_n converges. Therefore the series $\Sigma_p \frac{1}{p-1}$ diverges. But $\frac{2}{p} \ge \frac{1}{p-1}$ for $p \ge 2$. So $\Sigma_p \frac{2}{p}$ diverges, and so then does $\Sigma_p \frac{1}{p}$.

Distribution of values of the abundancy ratio Since $\sigma(n)/n$ is a positive rational ≥ 1, it is natural to ask how these rationals are distributed. Laatsch [**2**] has shown that the set of abundancy ratios $\sigma(n)/n$ for $n \geq 1$ is dense in the interval $[1, \infty)$. The next theorem shows that not all rationals in this interval are abundancy ratios.

THEOREM 1. *If k is relatively prime to m, and $m < k < \sigma(m)$, then k/m is not the abundancy ratio of any integer.*

Proof. Assume $\frac{k}{m} = \frac{\sigma(n)}{n}$ for some n. Then $m\sigma(n) = kn$, so $m|kn$, hence $m|n$ because $(k, m) = 1$. But by (2), $\frac{\sigma(m)}{m} \leq \frac{\sigma(n)}{n} = \frac{k}{m}$, contradicting the assumption $k < \sigma(m)$.

COROLLARY. (a) *For $m > 1$, the rational $(m+1)/m$ is an abundancy ratio if and only if m is prime.*

(b) *For m prime, $(m+1)/m$ is the abundancy ratio only of m.*

Proof. (a) If m is prime then $\frac{\sigma(m)}{m} = \frac{m+1}{m}$. Conversely, if m is composite, then $m < m+1 < \sigma(m)$, so by Theorem 1, $(m+1)/m$ is not an abundancy ratio.

The proof of (b) is left as an exercise for the reader.

The next theorem may be regarded as a complement to Laatsch's result.

THEOREM 2. *The set of rationals that are not abundancy ratios is dense in $[1, \infty)$.*

The proof will use the following lemma.

LEMMA. *Let m be a positive integer. If p is prime with $p > 2m$, then among any $2m$ consecutive integers, there is at least one integer relatively prime to pm.*

Proof of Lemma. Let S be any set of $2m$ consecutive integers. If $p > 2m$ there is at most one multiple of p in S. But S contains at least two integers relatively prime to m, one of which is relatively prime to p and, therefore, also to pm.

Proof of Theorem 2. Choose any real $x \geq 1$, and any $\epsilon > 0$. We will exhibit a rational in the interval $(x - \epsilon, x + \epsilon)$ that is not an abundancy ratio. By Laatsch's theorem, choose $m > 1$ so the abundancy ratio $\sigma(m)/m$ is in the interval $(x - \frac{\epsilon}{2}, x + \frac{\epsilon}{2})$. For every prime $p > 2m$ we have

$$x - \frac{\epsilon}{2} < \frac{\sigma(m)}{m} < \frac{\sigma(pm)}{pm} = \left(1 + \frac{1}{p}\right)\frac{\sigma(m)}{m} < \left(1 + \frac{1}{p}\right)\left(x + \frac{\epsilon}{2}\right).$$

If we also require $p > \frac{2x+\epsilon}{\epsilon}$, then $(1 + \frac{1}{p})(x + \frac{\epsilon}{2}) < x + \epsilon$, and we have

$$x - \frac{\epsilon}{2} < \frac{\sigma(pm)}{pm} < x + \epsilon. \tag{4}$$

By the lemma, we know that $\sigma(pm) - k$ is relatively prime to pm for some k with $1 \leq k \leq 2m$. For such k we also have

$$\sigma(pm) - k \geq \sigma(pm) - 2m \geq (p+1)(m+1) - 2m > pm$$

because $p > 2m$. Therefore, by Theorem 1, $\frac{\sigma(pm)-k}{pm}$ is not an abundancy ratio. And by (4) we have

$$\frac{\sigma(pm)-k}{pm} \geq \frac{\sigma(pm)-2m}{pm} = \frac{\sigma(pm)}{pm} - \frac{2}{p} > x - \frac{\epsilon}{2} - \frac{2}{p}.$$

If $p \geq \frac{4}{\epsilon}$, we have $x - \frac{\epsilon}{2} - \frac{2}{p} \geq x - \epsilon$, giving us $\frac{\sigma(pm)-k}{pm} > x - \epsilon$. Therefore all the foregoing inequalities are satisfied if we choose $p > \max\left\{2m, \frac{2x+\epsilon}{\epsilon}, \frac{4}{\epsilon}\right\}$, and we find that

$$x - \epsilon < \frac{\sigma(pm)-k}{pm} < \frac{\sigma(pm)}{pm} < x + \epsilon,$$

by (4). This provides a rational, $\frac{\sigma(pm)-k}{pm}$, not an abundancy ratio, within ϵ of x, and completes the proof.

A connection with odd perfect numbers The next result reveals a surprising connection with odd perfect numbers.

THEOREM 3. *If* $\frac{\sigma(n)}{n} = \frac{5}{3}$ *for some* n, *then* $5n$ *is an odd perfect number.*

Proof. For the given n we have $3\sigma(n) = 5n$, so $3 \mid n$. If n is even, then $6 \mid n$, and so by (2), $\frac{\sigma(n)}{n} \geq \frac{\sigma(6)}{6} = 2$, contradicting $\frac{\sigma(n)}{n} = \frac{5}{3}$. Thus n is odd, so $5n$ is also odd, and hence $\sigma(n)$ is odd. From the multiplicative property of $\sigma(n)$ it is easy to show that if n and $\sigma(n)$ are both odd, then n must be a square. Therefore $3^2 \mid n$. Does $5 \mid n$? If so, then $3^2 \cdot 5 \mid n$, and (2) implies

$$\frac{\sigma(n)}{n} \geq \frac{\sigma(3^2 \cdot 5)}{3^2 \cdot 5} = \frac{26}{15} > \frac{5}{3},$$

contradicting $\frac{\sigma(n)}{n} = \frac{5}{3}$. Therefore $(5, n) = 1$, so

$$\frac{\sigma(5n)}{5n} = \frac{\sigma(5)\sigma(n)}{5n} = \frac{6}{5} \cdot \frac{5}{3} = 2,$$

which means that $5n$ is an odd perfect number.

Acknowledgment. I would like to thank Professor Tom M. Apostol for his help in the preparation of this paper. Also thanks to the referees who provided many helpful suggestions.

REFERENCES

1. R. P. Brent, G. L. Cohen, and H. J. J. te Riele, Improved techniques for lower bounds for odd perfect numbers, *Math. Comp.* 57 (1991), 857–868.
2. R. Laatsch, Measuring the abundancy of integers, this MAGAZINE 59 (1986), 84–92.

Cantor, Schröder, and Bernstein in Orbit

BERTHOLD SCHWEIZER
University of Massachusetts
Amherst, MA 01003-4515

One form of the Cantor–Schröder–Bernstein Theorem states that if X and Y are given non-empty, disjoint sets, f a one-to-one mapping from X onto a subset of Y and g a one-to-one mapping from Y onto a subset of X, then there exists a one-to-one mapping ϕ from X onto Y.

There are two commonly given proofs of this theorem. The first [**5**, **6**], which traces sets, is easy to follow but not very intuitive; and so a beginning student gains little insight from it. The second, which traces points, is also easy to follow and gives more insight. However, the way it is usually presented [**3**, **6**] suffers from a defect. This defect is linguistic rather than mathematical but can nevertheless lead to confusion and misunderstanding. It lies in the use of such terms as "parent," "ancestor," and "descendant." Thus, for example, if there are points x in X and y in Y such that $f(x)=y$ and $g(y)=x$, then x is a parent of y and y is a parent of x, and each of x and y is its own ancestor over and over again—a rather counterintuitive state of affairs. In addition, the diagrams that accompany the proof as given, e.g., in [**3**] and [**6**] shed little light on the situation.

The principal purpose of this note is to present a picture (given some years ago in [**10**]) which makes the point-tracing argument very clear, i.e., a proof (almost) without words of the Cantor–Schröder–Bernstein Theorem. To this end, for any x in X, define the *orbit* of x to be the set of points

$$\{\dots g^{-1}f^{-1}g^{-1}(x), f^{-1}g^{-1}(x), g^{-1}(x), x, f(x), gf(x), fgf(x)\dots\},$$

where f^{-1} and g^{-1} are the inverse functions of f and g, respectively, and where juxtaposition denotes composition of functions. This set may be viewed as a directed graph in which, for any a, b in the set, there is a directed edge from a to b (denoted by $a \to b$) if and only if either $b=f(a)$ or $b=g(a)$. Since f and g are single-valued and one-to-one, two orbits are either identical or distinct, i.e., the set of all orbits is a partition of $X \cup Y$. Next, a little reflection shows that these orbits fall into four distinct classes, which may be illustrated as follows:

I. $x \to y \to x \to y \to \cdots$

II. $y \to x \to y \to x \to \cdots$

III. $\cdots \to x \to y \to x \to y \to \cdots$

IV.
$$\begin{tikzcd} x \arrow[r] & y \arrow[r] & x \arrow[r] & y \arrow[r] & x \arrow[r] & y \arrow[d] \\ y \arrow[u] & & & & & x \arrow[d] \\ x \arrow[u] & y \arrow[l] & \cdots\ \cdots \arrow[l] & \cdots & x \arrow[l] & y \arrow[l] \end{tikzcd}$$

where the x's and y's symbolize distinct points of X and Y, with $y=f(x)$ and $x=g(y)$, respectively. Since every point of $X \cup Y$ belongs to a unique orbit of one of the above types, the desired mapping $\phi: X \to Y$ may now be defined as follows:

(a) If x belongs to an orbit of Type I, III, or IV, map it onto its immediate successor, i.e., let $\phi(x)=f(x)$.

(b) If x belongs to an orbit of Type II, map it onto its immediate predecessor, i.e., let $\phi(x)=g^{-1}(x)$.

The mapping ϕ is clearly one-to-one from X onto Y, and this establishes the theorem.

Notes

1. A stronger version of the Cantor–Schröder–Bernstein Theorem, due to Banach [**1**] (see also [**6**]), states that there are partitions $X = X_1 \cup X_2$ and $Y = Y_1 \cup Y_2$ such that $f(X_1) = Y_1$ and $g(Y_2) = X_2$. This result can also be obtained directly from the "picture-proof": simply let X_1 be the set of all points x in X that belong to orbits of Type I, III, and IV, and X_2 the set of all points x in X that belong to orbits of Type II.
2. The defect referred to above is the fact that in the proofs given in [**3**] and [**6**] the orbits of Types III and IV are lumped into one class. Thus each point in an orbit of Type IV must be viewed as its own ancestor (or descendant) countably often. Once this is understood, the proofs are of course correct.
3. The fact that the set of orbits yields a partition is a consequence of the fact that f and g are mappings, i.e., it does not require that f and g be one-to-one. It is also not hard to show directly that "belonging to the same orbit" is an equivalence relation. (See [**9**] for an elaboration of these points.)
4. The mapping ϕ is not unique. For example, a different mapping λ may be defined as follows:
   (a′) If x belongs to an orbit of Type I, map it onto its immediate successor, i.e., let $\lambda(x) = f(x)$.
   (b′) If x belongs to an orbit of Type II, III, or IV, map it onto its immediate predecessor, i.e., let $\lambda(x) = g^{-1}(x)$.
5. Another graph-theoretic proof, employing the notion of a bipartite graph, and due in essence to D. König [**7**], is given in [**2**]. Here the authors clearly point out the existence of cycles, i.e., orbits of Type IV. But they do not provide a diagram.
6. The Cantor–Schröder–Bernstein Theorem was conjectured by Cantor, not in the form given above, but in terms of cardinal numbers. Specifically, Cantor conjectured (indeed, was convinced) that if X and Y are given non-empty sets such that the cardinal number of X is less than or equal to the cardinal number of Y and the cardinal number of Y is less than or equal to the cardinal number of X, then these two cardinal numbers are equal. E. Schröder announced a proof of this Equivalence Theorem in 1896, but his argument (published in 1898) was later found to be faulty [**8**]. In 1897, F. Bernstein presented a correct—set-tracing—proof in Cantor's seminar. His proof, duly acknowledged, first appeared in E. Borel's book [**4**]; see [**5**] and [**8**] for further details and references.

REFERENCES

1. S. Banach, Un théorème sur les transformations biunivoque, *Fund. Math.* 6 (1924), 236–239.
2. E. Bertram and P. Horak, Some applications of graph theory to other parts of mathematics, *Math. Intelligencer* 3, No. 3 (1999), 6–11.
3. G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, Macmillan, New York, NY, 1977.
4. E. Borel, *Leçons sur la Théorie des Fonctions*, Gauthier-Villars, Paris, France, 1898.
5. A. Fraenkel, *Abstract Set Theory*, North-Holland, Amsterdam, Holland, 1953.
6. V. Klee and J. R. Reay, A surprising but easily proved geometric decomposition theorem, this MAGAZINE 71 (1998), 3–11.
7. D. König, *Theorie der endlichen und unendlichen Graphen*, Chelsea Publishing, New York, NY, 1950.
8. A. Korselt, Über einen Beweis des Äquivalenzsatzes, *Math. Ann.* 70 (1911), 294–296.
9. R. E. Rice, B. Schweizer, and A. Sklar, When is $f(f(z)) = az^2 + bz + c$?, *Amer. Math. Monthly* 87 (1980), 252–263.
10. B. Schweizer, Bemerkung, *Aequationes Math.* 11 (1974), 312.

A Matrix Proof of Newton's Identities

DAN KALMAN
American University
Washington, DC 20016-8050

Newton's identities relate sums of powers of roots of a polynomial with the coefficients of the polynomial. They are generally encountered in discussions of symmetric functions (see [**4**, **9**]): a polynomial's coefficients are symmetric functions of the roots, as is the sum of the k^{th} powers of those roots.

Newton's identities also have a natural expression in the context of matrix algebra, where the trace of the k^{th} power of a matrix is the sum of the k^{th} powers of the eigenvalues. In this setting, Newton's identities can be derived as a simple consequence of the Cayley–Hamilton theorem. Presenting that derivation is the purpose of this note.

There are a variety of derivations for Newton's identities in the literature. Berlekamp's derivation [**2**] using generating function methods is short and elegant, and Mead presents a very interesting argument [**7**] using a novel notation. In yet another approach [**1**], Baker uses differentiation to obtain a nice recursion. Eidswick's derivation [**3**] uses a related application of logarithmic differentiation. All of these proofs are elementary and understandable, but they involve manipulations or concepts that might make them a bit forbidding to students. In contrast, the proof presented here uses only methods that would be readily accessible to most linear algebra students.

Interestingly, the matrix interpretation of Newton's identities is familiar in the linear algebra literature, providing a means of computing the characteristic polynomial of a matrix in terms of the traces of the powers of the matrix ([**1**, **8**]). However, using the matrix setting to *derive* Newton's identities doesn't seem to be well known.

Let $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ have roots r_j, $j = 1, \ldots, n$. Define

$$s_k \equiv \sum_{j=1}^{n} r_j^k.$$

Newton's identities are

$$s_k + a_{n-1}s_{k-1} + \cdots + a_0 s_{k-n} = 0 \quad (k > n)$$
$$s_k + a_{n-1}s_{k-1} + \cdots + a_{n-k+1}s_1 = -ka_{n-k} \quad (1 \le k \le n)$$

Now let C be an $n \times n$ matrix with characteristic polynomial equal to p. For example, C might be

$$\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix},$$

the *companion matrix* of p ([**6**]). Then the roots of p are the eigenvalues of C, and more generally, the k^{th} powers of the roots of p are the eigenvalues of C^k. Accordingly, we observe that s_k is the *trace* of C^k, written $\text{tr}(C^k)$. Recall that the trace of a matrix is at once the sum of the eigenvalues and the sum of the diagonal entries. In particular, the trace operation is linear: $\mathrm{tr}(\alpha A + \beta B) = \alpha\,\mathrm{tr}(A) + \beta\,\mathrm{tr}(B)$.

Now for $k > n$, using the trace formulation, Newton's identity becomes

$$\mathrm{tr}(C^k) + a_{n-1}\mathrm{tr}(C^{k-1}) + \cdots + a_0\mathrm{tr}(C^{k-n}) = 0$$

and since the trace function is linear, we can rewrite this as

$$\mathrm{tr}\left(C^k + a_{n-1}C^{k-1} + \cdots + a_0 C^{k-n}\right) = 0, \quad \text{or} \quad \mathrm{tr}\left(C^{k-n}p(C)\right) = 0.$$

Thus, the $k > n$ case follows immediately from the Cayley–Hamilton theorem, which says that $p(C) = 0$.

For $1 \le k \le n$, the trace version of Newton's identity is

$$\mathrm{tr}(C^k) + a_{n-1}\mathrm{tr}(C^{k-1}) + \cdots + a_{n-k+1}\mathrm{tr}(C) = -ka_{n-k}$$

which can again be rewritten as

$$\mathrm{tr}\left(C^k + a_{n-1}C^{k-1} + \cdots + a_{n-k+1}C\right) = -ka_{n-k}.$$

For reasons that will be clear later, we modify this slightly, to

$$\mathrm{tr}\left(C^k + a_{n-1}C^{k-1} + \cdots + a_{n-k+1}C + a_{n-k}I\right) = (n-k)a_{n-k}. \tag{1}$$

This identity can also be derived from the Cayley–Hamilton theorem, in a slightly different way. As is well known, a real number r is a root of a real polynomial $p(x)$ if and only if $(x - r)$ is a factor of $p(x)$, and the complimentary factor can be determined using synthetic division. This situation can be mimicked exactly using matrices: let $X = xI$, and divide $p(X)$ by $X - C$ using synthetic division. Since $p(C) = 0$, the division terminates without remainder, providing the factorization

$$\begin{aligned} p(X) = (X - C)\,\big[X^{n-1} &+ (C + a_{n-1}I)X^{n-2} + \left(C^2 + a_{n-1}C + a_{n-2}I\right)X^{n-3} \\ &+ \cdots + \left(C^{n-1} + a_{n-1}C^{n-2} + \cdots + a_1 I\right)I\big] \end{aligned}$$

(see [**5**]).

To relate this to equation (1), we will want to introduce the trace operation. Unfortunately, the trace does not relate well to matrix products, so it is necessary to eliminate the factor of $(X - C)$ on the right. Fortunately, as long as x is not an eigenvalue of C, we know that $(xI - C) = (X - C)$ is non-singular, so we can write

$$\begin{aligned} (X - C)^{-1}p(X) = X^{n-1} &+ (C + a_{n-1}I)X^{n-2} + \left(C^2 + a_{n-1}C + a_{n-2}I\right)X^{n-3} + \cdots \\ &+ \left(C^{n-1} + a_{n-1}C^{n-2} + \cdots + a_1 I\right)I. \end{aligned}$$

Taking the trace of each side then leads to

$$\begin{aligned} \mathrm{tr}\left[(X - C)^{-1}p(X)\right] = nx^{n-1} &+ \mathrm{tr}(C + a_{n-1}I)x^{n-2} + \cdots \\ &+ \mathrm{tr}\left(C^{n-1} + a_{n-1}C^{n-2} + \cdots + a_1 I\right) \end{aligned} \tag{2}$$

because $\mathrm{tr}(I) = n$ and $\mathrm{tr}(X^k A) = \mathrm{tr}(x^k IA) = x^k\mathrm{tr}(A)$ for any matrix A.

We will next show that the left side of this equation is none other than $p'(x)$. Then, comparing coefficients on either side will complete the proof. Indeed, equating the coefficient of x^{n-k-1} in $p'(x)$ with the corresponding coefficient on the right side of equation (2) gives

$$(n-k)a_{n-k} = \mathrm{tr}\left(C^k + a_{n-1}C^{k-1} + \cdots + a_{n-k+1}C + a_{n-k}I\right)$$

which is exactly the same as equation (1).

So, consider $A=(X-C)^{-1}p(X)$. Observe that $p(X)=p(xI)=p(x)I$, so we can equally well write $A=p(x)(xI-C)^{-1}$. This shows that

$$\mathrm{tr}(A)=p(x)\mathrm{tr}(xI-C)^{-1}.$$

Now the trace of any matrix is the sum of its eigenvalues (with multiplicities as in the characteristic polynomial). And the eigenvalues of $(xI-C)^{-1}$ are simply the fractions $1/(x-r_1)$, $1/(x-r_2),\cdots,1/(x-r_n)$. This shows

$$\mathrm{tr}(A)=p(x)\left(\frac{1}{x-r_1}+\frac{1}{x-r_2}+\cdots+\frac{1}{x-r_n}\right)$$

which is immediately recognizable as the derivative $p'(x)$ (using the fact that $p(x)=(x-r_1)(x-r_2)\cdots(x-r_n)$). This completes the proof.

REFERENCES

1. George A. Baker, Jr., A new derivation of Newton's identities and their application to the calculation of the eigenvalues of a matrix, *Journal of the Society for Industrial and Applied Mathematics* 7 (1959), 143–148.
2. Elwyn R. Berlekamp, *Algebraic Coding Theory*, Revised 1984 Edition, Aegean Park Press, Laguna Hills, CA, 1984, page 212.
3. J. A. Eidswick, A proof of Newton's power sum formulas, *American Mathematical Monthly* 75 (1968), 396–397.
4. I. N. Herstein, *Topics in Algebra*, Ginn, Waltham, MA, 1964, page 208.
5. Dan Kalman, An extension of the remainder theorem, *Delta* 8 (1978), 77–80.
6. David C. Lay, *Linear Algebra and Its Applications*, 2nd ed., Addison-Wesley, Reading, MA, 1997, page 365.
7. D. G. Mead, Newton's identities, *American Mathematical Monthly* 99 (1992), 749–751.
8. D. W. Robinson, A matrix application of Newton's identities, *American Mathematical Monthly* 68 (1961), 367–369.
9. B. L. van der Waerden, *Algebra*, volume 1, Frederick Ungar, New York, NY, 1970, page 101.

Boxes for Isoperimetric Triangles

JOHN E. WETZEL
University of Illinois at Urbana-Champaign
Urbana, IL 61801

Introduction A rectangular region *covers* a family of curves if it contains a congruent copy of each curve in the family. We call such a region a *box* for the family. In this note we answer two questions concerning boxes for the family $\mathscr{T}$ of all triangles of perimeter two:

(1) Among all boxes for $\mathscr{T}$, which has least area?
(2) Among all boxes for $\mathscr{T}$ of prescribed shape, which has least area?

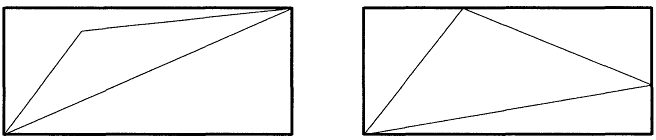
Some results are known about *triangular* covers for $\mathscr{T}$. In **[8]** we found the side of the smallest equilateral triangle that can cover $\mathscr{T}$, but the smallest triangular covers for $\mathscr{T}$ of other shapes remain unknown. With Füredi in **[5]**, we found the smallest triangle (without regard to shape) that can cover $\mathscr{T}$, and we showed somewhat surprisingly that this smallest triangle, whose area is about 0.2895, is the smallest possible convex cover for $\mathscr{T}$. Why study covers for $\mathscr{T}$? One reason is to gain insight into covers for the family of *all* closed curves of length two.

Problems of finding the smallest convex set (perhaps of prescribed shape) that can cover each curve of a prescribed type are called "worm problems," after a well-known unsolved problem in combinatorial geometry posed in the mid-1960s by L. Moser: Find the convex cover of least area for the family of all arcs of length one. Few such problems have been solved. For a glimpse at this and related curve covering problems, see Croft, Falconer, and Guy [**3**] and the website Finch [**4**].

Triangles inscribed in rectangles A triangle T is *inscribed* in a rectangle R (and the rectangle is *circumscribed* about the triangle) if $T \subset R$ and each side of R contains a vertex of T. There are just two possible configurations: two vertices of T lie at opposite corners of R and the third vertex lies elsewhere in R (Figure 1a), or one vertex of T is at a corner of R and another vertex lies on each of the two opposite sides of R (Figure 1b). In either case, at least one vertex of the triangle lies at a corner of the rectangle. Suppose that a triangle ABC of perimeter two is inscribed in a $u \times v$ rectangle, and vertex A lies at a corner of R. If B or C lies at the opposite corner of R it is clear that $\sqrt{u^2+v^2} < 1$. Otherwise, B and C lie on the sides of R opposite A. Reflect the rectangle and inscribed triangle across the side of R containing the vertex B, and then reflect the image rectangle and triangle across the line containing the image D of C (Figure 2). Then in the notation of the figure, we see that

$$\sqrt{u^2+v^2} = \frac{1}{2}AE < \frac{1}{2}ABDE = \frac{1}{2}(a+b+c) = 1. \tag{1}$$

(The earliest reference of which I am aware for this well-known inequality for the diagonal of a circumscribed rectangle is Jones and Schaer [**6**, fact (2), p. 5], where a proof is given using Minkowski's inequality; see also Chakerian and Klamkin [**2**, Theorem 3]. For the extension to orthotopes in $\mathbb{R}^d$, see Schaer and Wetzel [**6**, Lemma 1].)



(a) (b)

FIGURE 1
Triangles inscribed in a rectangle.

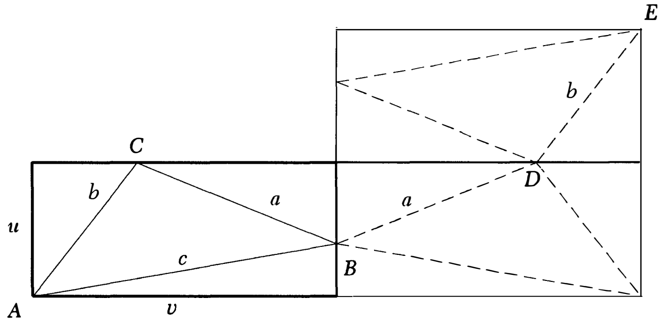


FIGURE 2
Reflection proof of (1).

The breadth of a triangle Suppose X is a convex set in the plane. For each direction θ, $0 \leq \theta \leq 2\pi$, let $w(\theta)$ be the width of X in the direction θ, i.e., the distance between the two parallel support lines of X with angle of inclination θ. It is known and not difficult to see that $w(\theta)$ is a continuous function of θ. The maximum of $w(\theta)$ is called the *diameter* of X. The minimum of $w(\theta)$ is called the *breadth* (or *thickness*) of X.

In the case of a triangle, the diameter is the length of the longest side, and the breadth is the length of the shortest altitude, the altitude to the longest side.

How broad can a triangle of perimeter two be? A little thought suggests that the shortest altitude of such a triangle is as large as possible when the triangle is equilateral. Since the equilateral triangle with side $2/3$ has altitude $1/\sqrt{3}$, it seems likely that the breadth of a triangle with perimeter two is at most $1/\sqrt{3}$.

This fact is an easy consequence of Heron's formula and the arithmetic-geometric means inequality. Indeed, arrange the notation so that $a = BC$ is the longest side of ABC; then the breadth of ABC is the altitude h_a to BC. Writing Δ for the area of ABC, we see since $a \geq \frac{2}{3}$ that

$$h_a \leq \frac{3}{2}ah_a = 3\Delta = 3\sqrt{(1)(1-a)(1-b)(1-c)}$$

$$\leq 3\left(\frac{(1-a)+(1-b)+(1-c)}{3}\right)^{3/2} = \frac{1}{\sqrt{3}}.$$

(See Bottema, et. al. [**1**, 6.1, p. 42]. The central inequality in this argument, that $3\Delta\sqrt{3} \leq s^2$ (where s is the semiperimeter), is the isoperimetric inequality for triangles; it asserts that the equilateral triangle has maximum area among triangles of the same perimeter. The inequality also follows immediately from Santaló's inequality $h_a + h_b + h_c \leq s\sqrt{3}$; see [**1**, 6.1, p. 60].)

Boxes for triangles of perimeter two The smallest rectangle that can accommodate every closed curve of length two has sides $2/\pi$ and $\sqrt{\pi^2 - 4}/\pi$ and area about 0.49095 (see Chakerian and Klamkin [**2**], Schaer and Wetzel [**7**]). For triangles of perimeter two one would expect to do a little better.

THEOREM 1. *The rectangle of least area that contains a congruent copy of every triangle of perimeter two has sides of lengths* $1/\sqrt{3}$ *and* $\sqrt{2}/\sqrt{3}$ *and area* $\sqrt{2}/3 \approx 0.47140$.

Proof. We show first that a rectangle with these dimensions is a cover for $\mathcal{T}$. Let ABC be such a triangle, and let $w(\theta)$ be its width in the direction θ. The maximum and minimum of $w(\theta)$ for $0 \leq \theta \leq 2\pi$ occur at some angles θ_1 and θ_2; and $w(\theta_1) \geq 2/3$ and $w(\theta_2) \leq 1/\sqrt{3}$. It follows from the intermediate value theorem that there is a direction θ so that $w(\theta) = 1/\sqrt{3}$. The rectangle circumscribed about ABC with one side in the direction θ has a side of length $u = 1/\sqrt{3}$ (FIGURE 3). Let x be the length of the other side. Then $\sqrt{x^2 + u^2} \leq 1$, so that $x \leq \sqrt{1-u^2} = \sqrt{2}/\sqrt{3}$. So ABC fits in the rectangle whose sides are $1/\sqrt{3}$ and $\sqrt{2}/\sqrt{3}$. To complete the argument, we must show that no rectangle of smaller area covers $\mathcal{T}$. Suppose that an $x \times y$ rectangle $(x \leq y)$ contains a congruent copy of every triangle of perimeter two. Then $x \geq 1/\sqrt{3}$, because an equilateral triangle with side $2/3$ must be accommodated; and $\sqrt{x^2 + y^2} \geq 1$, because flat triangles whose longest side is nearly one must be accommodated. We are to conclude that $xy \geq \sqrt{2}/3$. If $x > 1/\sqrt{3}$ and $\sqrt{x^2 + y^2} > 1$, we replace the $x \times y$ rectangle by a smaller similar $x' \times y'$ rectangle for which at least one equality holds: either $x' = 1/\sqrt{3}$ and $\sqrt{x'^2 + y'^2} \geq 1$, or $x' \geq 1/\sqrt{3}$ and $\sqrt{x'^2 + y'^2} = 1$. In the former case, $y' \geq \sqrt{1 - x'^2} = \sqrt{2}/\sqrt{3}$, so that $xy \geq x'y' \geq \sqrt{2}/3$. In the latter case, to show that $xy \geq x'y' = x'\sqrt{1 - x'^2} \geq \sqrt{2}/3$ for $1/\sqrt{3} \leq x' \leq 1/\sqrt{2}$ is a calculus exercise the details of which we leave to the reader. ■

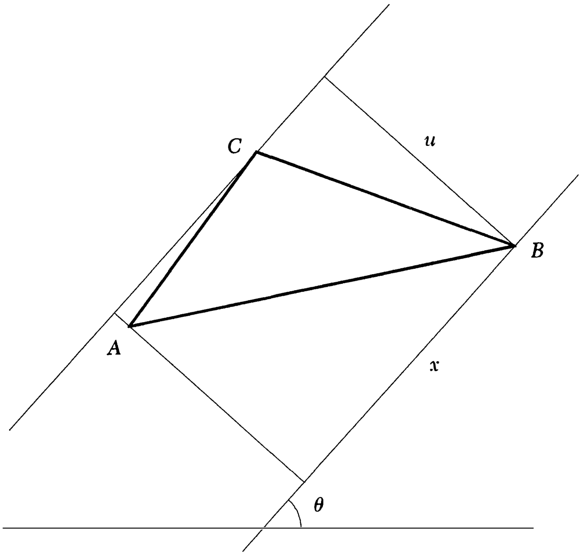


FIGURE 3
A direction with width $u = 1/\sqrt{3}$.

Finally, we determine the smallest box for $\mathscr{T}$ of prescribed shape. For long and thin rectangles it is plain that only the width matters; but for rectangles that are more rotund, the diagonals also must have length at least one.

THEOREM 2. *Suppose a $u \times v$ rectangle R_0 is given, with $u \leq v$. The smallest rectangle similar to R_0 that contains a congruent copy of every triangle of perimeter two has sides*

(a) $1/\sqrt{3}$ *and* $v/(u\sqrt{3})$ *if* $v \geq \sqrt{2}u$;
(b) $u/\sqrt{u^2 + v^2}$ *and* $v/\sqrt{u^2 + v^2}$ *if* $v \leq \sqrt{2}u$.

Proof. Suppose first that $v \geq \sqrt{2}u$; and let R be a rectangle with sides $x = 1/\sqrt{3}$ and $y = v/(u\sqrt{3})$. Then R is similar to R_0, $x \leq y$, and $y = v/(u\sqrt{3}) \geq \sqrt{2}/\sqrt{3}$. So R is a superset of the rectangle of Theorem 1 and consequently is a cover for $\mathscr{T}$. No smaller rectangle similar to R can have this property, because the equilateral triangle with perimeter two must be accommodated. This proves (a).

Suppose finally that $v \leq \sqrt{2}u$, and let R be a rectangle with sides $x = u/\sqrt{u^2 + v^2}$ and $y = v/\sqrt{u^2 + v^2}$. Since $\sqrt{x^2 + y^2} = 1$, no smaller rectangle similar to R_0 can cover $\mathscr{T}$. It remains to show that R is a cover for $\mathscr{T}$. Note first that $y \geq x \geq 1/\sqrt{3}$. Now suppose a triangle ABC with perimeter two is given, with longest side $a = BC$. If $x \leq a$, then there is a direction θ_0 in which ABC has width $w(\theta_0) = x$ (as in the proof of Theorem 1), and it follows that ABC fits in the rectangle with sides x and $\sqrt{1 - x^2} = y$. If $x \geq a$, then $y \geq x \geq 1/\sqrt{3} \geq h_a$, and since both $\angle B$ and $\angle C$ are acute, ABC can be covered by the rectangle with sides x and y, viz., R. This proves (b). ■

In particular, the smallest *square* cover for $\mathscr{T}$ has side $1/\sqrt{2}$ and area $1/2$.

Further questions There are many closely related interesting problems that might yield to the calculus and a little insight. Here are a few:

1. Find the smallest triangle similar to a given triangle that can cover every triangle of perimeter two, or every rectangle of perimeter two.
2. It is easy to see that every triangle of perimeter two has a circumscribed rectangle whose area is no larger than $2\sqrt{3}/9$, and since the equilateral triangle with perimeter two lies in no smaller rectangle, this constant is sharp. How *large* a circumscribed rectangle can one guarantee?
3. Find analogues of Theorem 1 and Theorem 2 in $\mathbb{R}^3$, in $\mathbb{R}^d$.

REFERENCES

1. O. Bottema, R. Ž. Djordjević, R. R. Janić, D. S. Mitrinović, and P. M. Vasić, *Geometric Inequalities*, Wolters-Noordhoff, Groningen, The Netherlands, 1969.
2. G. D. Chakerian and M. S. Klamkin, "Minimal covers for closed curves," this MAGAZINE 46 (1973), 55–63.
3. Hallard T. Croft, Kenneth J. Falconer, and Richard K. Guy, *Unsolved Problems in Geometry*, Springer-Verlag, New York, NY, 1991.
4. Steven Finch, "Moser's worm constant," URL: `www.mathsoft.com/asolve/constant/worm/`, 2000.
5. Zoltan Füredi and John E. Wetzel, "The smallest convex cover for triangles of perimeter two," to appear in *Geom. Dedicata.*
6. J. P. Jones and J. Schaer, *The Worm Problem*, University of Calgary, Department of Mathematics, Statistics, and Computing Science, Research Paper No. 100, Calgary, Alberta, Canada, 1970.
7. Jonathan Schaer and John E. Wetzel, "Boxes for curves of constant length," *Israel J. Math.* 12, 1972, 257–265.
8. John E. Wetzel, "The smallest equilateral cover for triangles of perimeter two," this MAGAZINE 70 (1997), 125–130.

Scooping the Loop Snooper

an elementary proof of the undecidability of the halting problem

No program can say what another will do.
Now, I won't just assert that, I'll prove it to you:
I will prove that although you might work till you drop,
you can't predict whether a program will stop.

Imagine we have a procedure called P
that will snoop in the source code of programs to see
there aren't infinite loops that go round and around;
and P prints the word "Fine!" if no looping is found.

You feed in your code, and the input it needs,
and then P takes them both and it studies and reads
and computes whether things will all end as they should
(as opposed to going loopy the way that they could).

Well, the truth is that P cannot possibly be,
because if you wrote it and gave it to me,
I could use it to set up a logical bind
that would shatter your reason and scramble your mind.

Here's the trick I would use—and it's simple to do.
I'd define a procedure—we'll name the thing Q—
that would take any program and call P (of course!)
to tell if it looped, by reading the source;

And if so, Q would simply print "Loop!" and then stop;
but if no, Q would go right back up to the top,
and start off again, looping endlessly back,
till the universe dies and is frozen and black.

And this program called Q wouldn't stay on the shelf;
I would run it, and (fiendishly) feed it *itself*.
What behavior results when I do this with Q?
When it reads its own source code, just what will it do?

If P warns of loops, Q will print "Loop!" and quit;
yet P is supposed to speak truly of it.
So if Q's going to quit, then P should say, "Fine!"—
which will make Q go back to its very first line!

No matter what P would have done, Q will scoop it:
Q uses P's output to make P look stupid.
If P gets things right then it lies in its tooth;
and if it speaks falsely, it's telling the truth!

I've created a paradox, neat as can be—
and simply by using your putative P.
When you assumed P you stepped into a snare;
Your assumptions have led you right into my lair.

So, how to escape from this logical mess?
I don't have to tell you; I'm sure you can guess.
By *reductio*, there cannot possibly be
a procedure that acts like the mythical P.

You can never discover mechanical means
for predicting the acts of computing machines.
It's something that cannot be done. So we users
must find our own bugs; our computers are losers!

—Geoffrey K. Pullum
Stevenson College
University of California
Santa Cruz, CA 95064

PROBLEMS

GEORGE T. GILBERT, *Editor*
Texas Christian University

ZE-LI DOU, KEN RICHARDSON, and SUSAN G. STAPLES, *Assistant Editors*
Texas Christian University

Proposals

To be considered for publication, solutions should be received by March 1, 2001.

1603. *Proposed by Ho-joo Lee, student, Kwangwoon University, Seoul, South Korea.*

Find all integer solutions to $x^5 + y^5 = (x+y)^3$.

1604. *Proposed by Răzvan Tudoran, University of Timişora, Timişora, Romania.*

Let g be a differentiable function on the nonnegative reals such that $g(0) \in [0, 1]$ and $\lim_{x\to\infty} g(x) = \infty$. Let f be defined on the nonnegative reals and satisfy $f(0) > g(0)$ and, for some positive k and r and all nonnegative x and y,

$$|f(x) - f(y)| \le k|g(x) - g(y)|^r.$$

Prove that there exists nonnegative c such that $f(c) = [g(c)]^{\lfloor r \rfloor + 1}$.

1605. *Proposed by Chi Hin Lau, student, University of Hong Kong, Hong Kong, China.*

In $\triangle ABC$, $\angle A = 60°$ and P is a point in its plane such that $PA = 6$, $PB = 7$, and $PC = 10$. Find the maximum possible area of $\triangle ABC$.

1606. *Proposed by Anthony A. Ruffa, Naval Undersea Warfare Center Division, Newport, Rhode Island.*

For x real and nonzero, show that

$$\frac{\sin x}{x} = \cos^2(x/2) + \sum_{n=1}^{\infty} \sin^2(x/2^{n+1}) \prod_{m=1}^{n} \cos(x/2^m).$$

We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals must, in general, be accompanied by solutions and by any bibliographical information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution.

Solutions should be written in a style appropriate for this MAGAZINE. *Each solution should begin on a separate sheet containing the solver's name and full address.*

Solutions and new proposals should be mailed to Elgin Johnston, Problems Editor, Department of Mathematics, Iowa State University, Ames, IA 50011, or mailed electronically (ideally as a LATEX file) to `johnston@math.iastate.edu`. *Readers who use e-mail should also provide an e-mail address.*

1607. *Proposed by Hassan A. Shah Ali, Tehran, Iran.*

Let n, k, and m be positive integers satisfying

$$m > \binom{n}{k} - \binom{\lfloor n/2 \rfloor}{k} - \binom{\lceil n/2 \rceil}{k}.$$

Let A be a set with $|A| = n$ and let $A_1, \ldots, A_m$ be distinct k-subsets of A. Prove that if $a_1 \in A_1, \ldots, a_m \in A_m$, then there exists $i \in \{1, \ldots, m\}$ such that $A_i \subset \{a_1, \ldots, a_m\}$.

Quickies

Answers to the Quickies are on pages 329.

Q903. *Proposed by Erwin Just, Professor Emeritus, Bronx Community College, Bronx, New York.*

Let $*$ be an associative binary operation on a set S and n a positive integer. If $x * y = y * x^n$ for all x and y in S, must $*$ be commutative?

Q904. *Proposed by Norman Schaumberger, Professor Emeritus, Bronx Community College, Bronx, New York.*

For $x > 2$, prove that

$$\ln\frac{x}{x-1} \le \sum_{j=0}^{\infty} \frac{1}{x^{2^j}} \le \ln\frac{x-1}{x-2}.$$

Solutions

Maximizing a Product of Integers with Fixed Sum **October 1999**

1579. *Proposed by T. S. Michael and William P. Wardlaw, U. S. Naval Academy, Annapolis, Maryland.*

For each positive integer n, find a set of positive integers whose sum is n and whose product is as large as possible.

(If repetitions are allowed with $n = 1979$, we obtain problem A-1 from the 1979 Putnam Competition.)

Solution by SMSU Problem Solving Group, Southwest Missouri State University, Springfield, Missouri.

For $n = 1, 2, 3, 4$, the set yielding the maximal product is just $\{n\}$. For each positive integer $n > 4$, we will show that the set $\{2, 3, \ldots, g-1, g+1, \ldots, k\}$, $3 \le g \le k-1$, of integers adding to n gives a maximum product unless n is 1, 2, or 3 less than the triangular number $k(k+1)/2$, in which case the sets $\{2, 3, \ldots, k\}$, $\{3, \ldots, k-1, k+1\}$, and $\{3, \ldots, k\}$, respectively, yield a maximum product.

Let $n = a_1 + a_2 + \cdots + a_k > 4$ with $a_i < a_{i+1}$. Because $2(n-2) > n$, we may assume $k \ge 2$. If $a_1 = 1$, then we may replace the sum with $a_2 + \cdots + a_{k-1} + (a_k + 1)$

to obtain a larger product. Also, if $a_1 > 3$, a larger product results if we replace the sum with $(a_1 - 1)/2 + (a_1 + 1)/2 + a_2 + \cdots + a_k$ if a_1 is odd or $a_1/2 + (a_1/2 + 1) + (a_2 - 1) + a_3 + \cdots + a_k$ if a_1 is even. Thus $a_1 = 2$ or $a_1 = 3$ for the maximal product.

Now if $a_{i+1} - a_i > 2$, then replace a_i with $a_i + 1$ and a_{i+1} with $a_{i+1} - 1$, with larger product. If $a_{i+1} = a_i + 2$ and $a_{j+1} = a_j + 2$, $i < j$, then replace a_i with $a_i + 1$ and a_{j+1} with $a_{j+1} - 1$ with larger product. Thus the maximum product occurs for a set of consecutive integers with perhaps one integer in the set omitted. We conclude that the maximal product occurs for a set of one of the four forms $\{2, 3, \ldots, k\}$, $\{2, 3, \ldots, g-1, g+1, \ldots, k\}$, $\{3, \ldots, k\}$, or $\{3, \ldots, g-1, g+1, \ldots, k\}$.

Observe that at most one set of the first two forms and at most one set of the latter two forms can yield a sum of n. For any $n > 1$, there exist unique nonnegative integers k and g, $g \le k - 1$, such that

$$n = k(k+1)/2 - 1 - g = 2 + 3 + \cdots + k - g.$$

If $3 \le g \le k-1$, then n can be written as $n = 2 + 3 + \cdots + (g-1) + (g+1) + \cdots + k$. If $g > 5$, then $n = 3 + \cdots + (g-3) + (g-1) + \cdots + k$ also. Neither type of set not containing 2 is possible if $g \le 5$. Because $2(g-2) > g$ for $g > 5$, the sum $n = 2 + 3 + \cdots + (g-1) + (g+1) + \cdots + k$ yields the maximum product. If $g = 0$, 1, or 2, then n can be expressed uniquely in one of the four possible forms as $n = 2 + 3 + \cdots + k$, $n = 3 + \cdots + (k-1) + (k+1)$, or $n = 3 + 4 + \cdots + k$, respectively.

Comment. Achilleas Sinefakopoulos reports that this problem appeared as question 1 in Hungary's 1964 Schweitzer Contest, a collection of which are gathered in Gábor J. Székely, editor, *Contests in Higher Mathematics*, Springer (1996), 4. He also points out that the Putnam question with the sum 1976 also appears on the 1976 International Mathematical Olympiad.

Also solved by Robert A. Agnew, Michel Bataille (France), J. C. Binz (Switzerland), David M. Bloom, Jean Bogaert (Belgium), Jeffrey Clark, Con Amore Problem Group (Denmark), Knut Dale (Norway), Daniele Donini (Italy), Tracy Dawn Hamilton, Victor Y. Kutsenok, Volkhard Schindler (Germany), Ben Schmidt, Harry Sedinger, W. R. Smythe, Westmont College Problem Solving Group, Li Zhu, Harald Ziehms (Germany), and the proposers. There was one incorrect solution.

A Variant of Nim **October 1999**

1580. *Proposed by Jerrold W. Grossman, Oakland University, Rochester, Michigan.*

Consider the following variation of the game of Nim. A position consists of piles of stones, with $n_i \ge 1$ stones in pile i. Two players alternately move by choosing one of the piles, permanently removing one or more stones from that pile, and, optionally, redistributing some (or all) of the remaining stones in that pile to one or more of the other remaining piles. (Once a pile is gone, no stones can be added to it.) The player who removes the last stone wins. Find a strategy for winning this game; in particular, determine which vectors of positive integers $(n_1, n_2, \ldots, n_k)$ allow the first player to win and which vectors allow the second player to win.

Solution by Hoe-Teck Wee, student, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Without loss of generality, we may assume $n_1 \le n_2 \le \cdots \le n_k$ throughout. Then, let

$$\mathscr{A} = \{(n_1, n_2, \ldots, n_k) \mid k \text{ even}, n_1 = n_2, n_3 = n_4, \ldots, n_{k-1} = n_k\},$$

and let $\mathscr{B}$ be the set of such vectors not in $\mathscr{A}$. We say that a position lies in $\mathscr{A}$ (respectively $\mathscr{B}$) if the vector representing the position lies in $\mathscr{A}$ (respectively $\mathscr{B}$). The vector $v = (n_1, n_2, \dots, n_k)$ allows the first player to win if and only if $v \in \mathscr{B}$.

To prove this claim, we shall prove the following:

(i) Given any position $v \in \mathscr{B}$, there exists a move that reduces it to a position in $\mathscr{A}$ or wins the game.
(ii) Given any position $v \in \mathscr{A}$, any move will reduce it to one in $\mathscr{B}$, or equivalently, there does not exist a move that either reduces a position in $\mathscr{A}$ to one also in $\mathscr{A}$ or wins the game.

It follows that a winning strategy for the game from a position in $\mathscr{B}$ is to play a move that reduces the position to one in $\mathscr{A}$. It also follows that the vectors in $\mathscr{B}$ allow the first player to win, and those in $\mathscr{A}$ allow the second player to win.

To prove (i), let $v = (n_1, n_2, \dots, n_k) \in \mathscr{B}$. First, assume $k = 2m+1$ is odd. If $m = 0$, the player can win the game by removing the single pile of stones. If $m > 0$, we have $n_1 \geq 1$, $n_{2i+1} \geq n_{2i}$, $i = 1, \dots, m$, so

$$n_{2m+1} + n_{2m-1} + \cdots + n_3 + n_1 \geq n_{2m} + \cdots + n_2 + 1.$$

Thus, we could permanently remove

$$n_{2m+1} - (n_{2m} - n_{2m-1}) - \cdots - (n_2 - n_1) \geq 1$$

stones from the kth pile and redistribute all the remaining stones, adding $n_{2i} - n_{2i-1}$ stones to the $(2i-1)$st pile, $i = 1, \dots, m$, so that we obtain the position $(n_2, n_2, \dots, n_{2m}, n_{2m}) \in \mathscr{A}$.

Now assume $k = 2m$ is even. We have $n_{2i} \geq n_{2i-1}$, $i = 1, \dots, m$, with at least one inequality strict since v is not in $\mathscr{A}$. Thus,

$$(n_2 - n_1) + \cdots + (n_{2m} - n_{2m-1}) > 0.$$

Therefore, we may permanently remove

$$(n_{2m} - n_1) - (n_3 - n_2) - \cdots - (n_{2m-1} - n_{2m-2}) > 0$$

stones from the kth pile, and redistribute another $(n_{2i+1} - n_{2i})$ stones to the $(2i)$th pile, $i = 1, \dots, m-1$, so that we have n_{2i+1} stones in both the $(2i)$th and $(2i+1)$st piles, $i = 1, \dots, m-1$, and n_1 stones in the kth pile. This leaves

$$(n_1, n_1, n_3, n_3, \dots, n_{2m-1}, n_{2m-1}) \in \mathscr{A}.$$

To prove (ii), let $v = (n_1, n_2, \dots, n_{2m}) \in \mathscr{A}$. Clearly the player cannot win the game at this stage; assume that there exists some move that reduces v to $v' = (n'_1, n'_2, \dots, n'_{2m}) \in \mathscr{A}$. With any move, exactly one pile decreases in size and the remaining piles either remain the same size or become bigger. Hence,

$$n'_{2i-1} = n'_{2i} \geq n_{2i-1} = n_{2i},\ i = 1, 2, \dots, m,$$

so we have

$$n'_1 + \cdots + n'_{2m} \geq n_1 + \cdots + n_{2m},$$

which contradicts the fact that at least one stone must be removed. Thus any move reduces a position in $\mathscr{A}$ to one in $\mathscr{B}$.

Also solved by J. C. Binz (Switzerland), Marty Getz and Dixon Jones, Joel D. Haywood, José H. Nieto and Julio Subocz (Venezuela), Michael Reid, and the proposer.

Ceva Transitivity **October 1999**

1581. *Proposed by Herbert Gülicher, Westfälische Wilhelms–Universität, Münster, Germany.*

Consider $\Delta P_1P_2P_3$ and points Q_1, Q_2, Q_3 in the interior of sides P_2P_3, P_1P_3, P_1P_2, respectively, such that P_1Q_1, P_2Q_2, and P_3Q_3 are concurrent (i.e., $\Delta Q_1Q_2Q_3$ is a cevian triangle of $\Delta P_1P_2P_3$). Let R_1, R_2, R_3 be in the interior of sides Q_2Q_3, Q_1Q_3, Q_1Q_2, respectively. Prove that the lines P_1R_1, P_2R_2, and P_3R_3 are concurrent if and only if the lines Q_1R_1, Q_2R_2, and Q_3R_3 are concurrent.

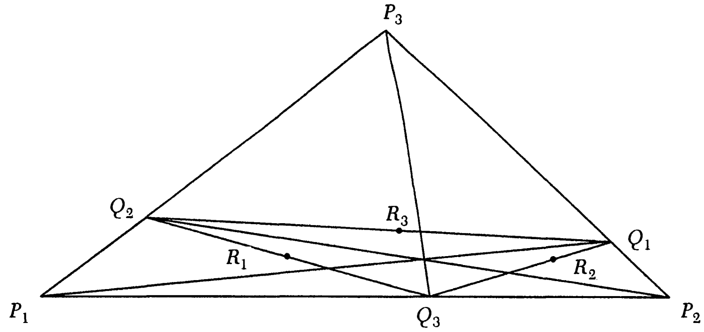


FIGURE 1

Solution by Daniele Donini, Bertinoro, Italy.

We interpret all subscripts as taken modulo 3. For $i = 1, 2, 3$, let S_i be the intersection of P_iQ_i and $Q_{i-1}Q_{i+1}$ and T_i be the intersection of P_iR_i and $P_{i-1}P_{i+1}$. By Ceva's theorem

$$\frac{P_3Q_1 \cdot P_1Q_2 \cdot P_2Q_3}{Q_1P_2 \cdot Q_2P_3 \cdot Q_3P_1} = 1 \quad \text{and} \quad \frac{Q_2S_1 \cdot Q_3S_2 \cdot Q_1S_3}{S_1Q_3 \cdot S_2Q_1 \cdot S_3Q_2} = 1.$$

By the invariance of cross-ratios of a quadruple of collinear points under projection, we have

$$\frac{P_{i-1}Q_i \cdot T_iP_{i+1}}{Q_iP_{i+1} \cdot P_{i-1}T_i} = \frac{Q_{i+1}S_i \cdot R_iQ_{i-1}}{S_iQ_{i-1} \cdot Q_{i+1}R_i}$$

for $i = 1, 2, 3$. The product of these three equalities gives

$$\frac{P_3Q_1 \cdot P_1Q_2 \cdot P_2Q_3}{Q_1P_2 \cdot Q_2P_3 \cdot Q_3P_1} \cdot \frac{T_1P_2 \cdot T_2P_3 \cdot T_3P_1}{P_3T_1 \cdot P_1T_2 \cdot P_2T_3} = \frac{Q_2S_1 \cdot Q_3S_2 \cdot Q_1S_3}{S_1Q_3 \cdot S_2Q_1 \cdot S_3Q_2} \cdot \frac{R_1Q_3 \cdot R_2Q_1 \cdot R_3Q_2}{Q_2R_1 \cdot Q_3R_2 \cdot Q_1R_3},$$

or

$$\frac{T_1P_2 \cdot T_2P_3 \cdot T_3P_1}{P_3T_1 \cdot P_1T_2 \cdot P_2T_3} = \frac{R_1Q_3 \cdot R_2Q_1 \cdot R_3Q_2}{Q_2R_1 \cdot Q_3R_2 \cdot Q_1R_3}.$$

By Ceva's theorem, these two expressions equal 1 if and only if the lines P_1T_1, P_2T_2, and P_3T_3 (i.e., the lines P_1R_1, P_2R_2, and P_3R_3) are concurrent if and only if the lines Q_1R_1, Q_2R_2, and Q_3R_3 are concurrent.

Comment. Several solvers pointed out that the Q_i and R_i need not be interior points.

Also solved by Michel Bataille (France), Jordi Dou (Spain), John G. Heuver, Geoffrey A. Kandall, Victor Y. Kutsenok, Neela Lakshmanan, Richard E. Pfiefer, Volkhard Schindler (Germany), and the proposer.

Winning a Series by Two Games **October 1999**

1582. *Proposed by Western Maryland College Problems Group, Westminster, Maryland.*

Let teams A and B play a series of games. Each game has three possible outcomes: A wins with probability p, B wins with probability q, or they tie with probability $r = 1 - p - q$. The series ends when one team has won two more games than the other, that team being declared the winner of the series.

(a) Find the probability that A wins the series.
(b) Let X be the number of games in the series. Find the probability function for X and its expected value.

I. Solution by SUNY Oswego Problem Group, SUNY Oswego, Oswego, New York.

(a) To determine who wins the series, we need only consider those games in which there was a winner. The probability that A wins such a game is $p/(p+q)$, which will be abbreviated p'. Similarly, the probability of B winning such a game is $q/(p+q) = q'$. Since the series ends when one team has two more wins than the other, there must be an even number of non-tied games. In addition, at each earlier stage with an even number of non-tied games, there are an equal number which were won by each team. Thus, if A wins after $2n+2$ non-tied games, each of the first n pairs of non-tied games will contain one win by each team. Therefore, the probability that A wins after exactly $2n+2$ non-tied games is $2^n(p')^{n+2}(q')^n$. Therefore, the probability that A wins the series is

$$\sum_{n=0}^{\infty} 2^n(p')^{n+2}(q')^n = \frac{(p')^2}{1-2p'q'} = \frac{p^2}{p^2+q^2}.$$

(b) Suppose the series ends after x games, of which $2n+2$ are not ties. Then, first of all, the xth game is not a tie and of the first $x-1$ games, $x-2n-2$ are ties and $2n+1$ are non-ties. The probability of this occurring is

$$\binom{x-1}{2n+1} r^{x-2n-2}(p+q)^{2n+2}.$$

In addition, the non-tied games must form a winning sequence for one of the teams, which would occur with probability $(2p'q')^n\,[(p')^2+(q')^2]$. Putting these together, we see that the probability of the series ending after x games which include $2n+2$ non-ties is

$$\begin{aligned}&\binom{x-1}{2n+1} r^{x-2n-2}(p+q)^{2n+2}(2p'q')^n\left[(p')^2+(q')^2\right]\\ &\quad= \binom{x-1}{2n+1} r^{x-2n-2}(2pq)^n(p^2+q^2).\end{aligned}$$

To get $P(X=x)$, we use the binomial theorem to sum these terms for the different values of n, obtaining

$$\begin{aligned}P(X=x) &= \sum_{2n+2\le x} \binom{x-1}{2n+1} r^{x-2n-2}(2pq)^n(p^2+q^2)\\ &= \frac{(p^2+q^2)}{\sqrt{2pq}} \sum_{2n+1\le x-1} \binom{x-1}{2n+1}\left(\sqrt{2pq}\right)^{2n+1} r^{(x-1)-(2n+1)}\\ &= \frac{(p^2+q^2)}{2\sqrt{2pq}}\left[\left(r+\sqrt{2pq}\right)^{x-1} - \left(r-\sqrt{2pq}\right)^{x-1}\right].\end{aligned}$$

The expected value of X is

$$E(X)=\sum_{x=1}^{\infty} xP(X=x)=\frac{(p^2+q^2)}{2\sqrt{2pq}}\sum_{x=1}^{\infty} x\left[\left(r+\sqrt{2pq}\right)^{x-1}-\left(r-\sqrt{2pq}\right)^{x-1}\right]$$

$$=\frac{(p^2+q^2)}{2\sqrt{2pq}}\left(\frac{1}{\left[1-\left(r+\sqrt{2pq}\right)\right]^2}-\frac{1}{\left[1-\left(r-\sqrt{2pq}\right)\right]^2}\right)=\frac{2(p+q)}{p^2+q^2}.$$

II. Solution by David M. Bloom, Brooklyn College of CUNY, Brooklyn, New York.

We generalize the probability that A wins the series and its expected length to a series that ends when one team leads the other by n games.

(a) A wins the series with probability $p^n/(p^n+q^n)$. To prove this, let P_j be the probability that A wins the series if A already leads B by j games, $-n\le j\le n$. Then, if $p'=p/(p+q)$ and $q'=q/(p+q)$, we have the recurrence

$$P_j=p'P_{j+1}+q'P_{j-1},\quad -n<j<n.$$

The roots of the characteristic equation are 1 and q/p. Therefore, if $p\neq q$, the solution has the form $P_j=a+b(q/p)^j$. We use the values $P_{-n}=0$ and $P_n=1$ to solve for a and b, finding

$$P_j=\frac{p^{2n}-p^{n-j}q^{n+j}}{p^{2n}-q^{2n}},\quad -n\le j\le n.$$

Then the probability that A wins the series is

$$P_0=\frac{p^n}{p^n+q^n},$$

which is still correct if $p=q$.

(b) Let E_j denote the expected number of games remaining after A leads B by j games. For $-n<j<n$, we have

$$E_j=pE_{j+1}+qE_{j-1}+rE_j+1,$$

which we rewrite as

$$E_{j+1}=(1+q/p)E_j-(q/p)E_{j-1}-1/p.$$

Again for $p\neq q$, this nonhomogeneous recurrence has solution of the form

$$E_j=a+b(q/p)^j+j/(q-p).$$

The conditions $E_{-n}=E_n=0$ lead to

$$E_0=a+b=\frac{p^nE_n+q^nE_{-n}+n(p^n-q^n)/(p-q)}{p^n+q^n}=\frac{n(p^n-q^n)}{(p-q)(p^n+q^n)}.$$

When $p=q$, this translates to $n^2/(2p)$.

The probability function in the special case $n=2$ may be derived as in the first solution.

Also solved by Michael H. Andreoli, Jean Bogaert (Belgium), Con Amore Problem Group (Denmark), Rob Pratt (graduate student), Michael Vowe (Switzerland), and the proposers. There was one incorrect solution. There were five incomplete solutions that left one or more expressions as unevaluated sums.

Powers with Rational Differences **October 1999**

1583. *Proposed by George T. Gilbert, Texas Christian University, Fort Worth, Texas.*

Classify all pairs of complex numbers a and b for which $a^2 - b^2$, $a^3 - b^3$, and $a^5 - b^5$ are rational numbers.

Solution by Michael Reid, Brown University, Providence, Rhode Island.

We will show that $a^2 - b^2$, $a^3 - b^3$, and $a^5 - b^5$ are all rational if and only if a and b are both rational or $a = b$.

If a and b are both rational, then $a^2 - b^2$, $a^3 - b^3$, and $a^5 - b^5$ are also rational, and the same conclusion holds trivially if $a = b$.

Conversely, suppose that $a^2 - b^2$, $a^3 - b^3$, and $a^5 - b^5$ are rational and $a \neq b$. If a or b is zero, the other is clearly rational. Now suppose that $ab \neq 0$ and $a \neq b$. Let $t = a/b$, which we are assuming is not 0 or 1; we claim that t is rational. Define

$$f(X,Y,Z) := X^7Y - 3X^6Z + X^4Y^3 + 7X^3Y^2Z - 3X^2YZ^2 + XY^5 - 9XZ^3 + 5Y^4Z.$$

Then

$$f(a^2 - b^2, a^3 - b^3, a^5 - b^5) = -2b^{17}t^2(t-1)^4 p(t),$$

where $p(T)$ is the symmetric polynomial

$$p(T) := 20T^8 + 80T^7 + 158T^6 + 200T^5 + 209T^4 + 200T^3 + 158T^2 + 80T + 20.$$

Consider first the possibility that $p(t) = 0$. Observe that $t^3 \neq 1$, so that

$$u := \frac{(a^2 - b^2)^6}{(a^3 - b^3)^4} = \frac{(t^2 - 1)^6}{(t^3 - 1)^4}$$

is rational. However, using $p(t) = 0$, we verify that u satisfies $27u^2 + 18u - 125 = 0$, so u is irrational. This contradiction shows that $p(t) \neq 0$, so $f(a^2 - b^2, a^3 - b^3, a^5 - b^5) \neq 0$. We also have

$$f(b^2 - a^2, b^3 - a^3, b^5 - a^5) = -2a^{17}\left(\frac{1}{t}\right)^2\left(\frac{1}{t} - 1\right)^4 p\left(\frac{1}{t}\right) = -2b^{17}t^3(t-1)^4 p(t),$$

so that $t = f(b^2 - a^2, b^3 - a^3, b^5 - a^5)/f(a^2 - b^2, a^3 - b^3, a^5 - b^5)$ is rational. Finally, noting that $(a^5 - b^5)(t^3 - 1) \neq 0$, we have

$$b = \frac{(a^3 - b^3)^2}{a^5 - b^5} \cdot \frac{t^5 - 1}{(t^3 - 1)^2},$$

so b and consequently a are rational.

Also solved by the proposer.

Answers

Solutions to the Quickies on page 322.

A903. Yes. Setting $y = x^{n-1}$, we have $x^n = x^{2n-1}$. It follows that

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

We may now observe that

$$x * y = y * x^n = x^n * y^n = y^n * (x^n)^n = y^n * x^{n^2} = y^n * x^n = x * y^n = y * x.$$

A904. For u real and n a positive integer,

$$\frac{1-u^{2^n}}{1-u} = (1+u)(1+u^2)(1+u^{2^2})\cdots(1+u^{2^{n-1}}).$$

Let $x > 2$, and put $u = 1/x$ followed by $u = 1/(x-1)$. Then

$$\begin{aligned}
\ln\frac{1-\dfrac{1}{x^{2^n}}}{1-\dfrac{1}{x}} &= \sum_{j=0}^{n-1}\ln\left(1+\frac{1}{x^{2^j}}\right) = \sum_{j=0}^{n-1}\frac{1}{x^{2^j}}\ln\left(1+\frac{1}{x^{2^j}}\right)^{x^{2^j}} \\
&< \sum_{j=0}^{n-1}\frac{1}{x^{2^j}}\ln e = \sum_{j=0}^{n-1}\frac{1}{x^{2^j}} \\
&< \sum_{j=0}^{n-1}\frac{1}{x^{2^j}}\ln\left(1+\frac{1}{(x-1)^{2^j}}\right)^{x^{2^j}} = \sum_{j=0}^{n-1}\ln\left(1+\frac{1}{(x-1)^{2^j}}\right) \\
&= \ln\frac{1-\dfrac{1}{(x-1)^{2^n}}}{1-\dfrac{1}{x-1}}.
\end{aligned}$$

Letting $n \to \infty$ gives the desired result.

Letter continued from page 334

According to *Science News*, March 4, 2000, pp. 152–153, the Fermat numbers $F_k = 2^{2^k} + 1$ are now known to be composite for all $k = 5, \ldots, 24$, as well as selected others! (Note: The *Science News* article did not realize that everyone considers $F_0 = 3$ to be a Fermat prime. My letter pointing this out appears in *Science News*, May 13, 2000, p. 307.)

Solomon Golomb, Communication Sciences Institute
University of Southern California, Los Angeles, CA 90089-2565

REVIEWS

PAUL J. CAMPBELL, *Editor*
Beloit College

Assistant Editor: Eric S. Rosenthal, West Orange, NJ. Articles and books are selected for this section to call attention to interesting mathematical exposition that occurs outside the mainstream of mathematics literature. Readers are invited to suggest items for review to the editors.

Clay Mathematics Institute, Millennium Prize Problems, `http://www.ams.org/claymath/`.

The Clay Mathematics Institute (CMI) of Cambridge, Mass., has designated a \$7 million prize fund for the solution of seven mathematical problems (\$1 million each), with no time limit. The problems are $\mathcal{P}$ vs. $\mathcal{NP}$, the Hodge conjecture, the Poincaré conjecture, the Riemann hypothesis, the mass gap in the Yang-Mills equation, a mathematical theory for the Navier-Stokes equations, and the Birch and Swinnerton-Dyer conjecture. The Web page cited above links to short descriptions of each problem, which themselves link to sophisticated mathematical descriptions, as well as to the contest rules and to other details. CMI, founded by Boston businessman Landon T. Clay, engages in other activities to foster the creation of mathematical knowledge, disseminate mathematical insight, inspire talented students, and recognize mathematical achievement.

Alperin, Roger C., A mathematical theory of origami constructions and numbers, *New York Journal of Mathematics* 6 (2000) 119–133.

Origami, the originally Japanese craft of paper-folding, can be interpreted as an apparatus for mathematical construction: A fold constructs a line, and the intersection of two folds constructs a point. Much as one can specify the points and corresponding numbers constructible by straightedge and compass (or by other geometrical tools), this article sets forth axioms for paper-folding and characterizes the resulting *origami-constructible* points and numbers. Alperin's axioms allow the constructions of angle bisections, tangents of a parabola, and tangents to two parabolas. These axioms give exactly the points and numbers obtained from intersections of conics (see Carlos R. Videla, On points constructible from conics, *Mathematical Intelligencer* 19 (1997) 53–57), viz., the smallest subfield of $\mathbb{C}$ closed under square roots, cube roots, and complex conjugation.

Buchanan, Mark, That's the way the money goes, *New Scientist* (19 August 2000) 22–26.

"Why do rich people have all the money?" Wealth in most societies is distributed according to a Pareto law: The proportion of people with wealth W is proportional to $1/W^L$, with $2 < L < 3$. In the U.S., 20% of the people have 80% of the wealth. Jean-Philippe Bouchaud and Marc Mezard (Paris-Sud University), in attempting to build a theory of economics "from the ground up," discovered that solutions to the equations that they had devised had already been found by condensed-matter physicists in their model for a directed polymer. What the solutions show is that under normal conditions, the distribution of wealth tends to follow a Pareto law. "Chop off the heads of the rich, and a new rich will soon take their place." Bouchaud and Mezard's model suggests that the route to more uniform distribution of wealth is by encouraging its movement, either by inhibiting its accumulation (through taxation) or by increasing the number of people with whom a person tends to trade (through free trade and competition).

Murray, Margaret A.M., *Women Becoming Mathematicians: Creating a Professional Identity in Post-World War II America*, MIT Press, 2000; xviii + 277 pp, $19.95. ISBN 0-262-13369-5.

About 200 women received Ph.D.'s in mathematics in the U.S. between 1940 and 1959. This well-written book, based on extensive interviews with 36 of them, seeks answers to the questions of how women in those times became mathematicians and how they fared in a field dominated and defined by men. Author Murray calls the prevailing model of career development for male mathematicians "the myth of the mathematical life course": early emergence of natural talent, single-minded pursuit of research, a spouse to tend to domestic concerns and fend off distractions, and reduced productivity with age. This path is common for men in many spheres of endeavor; in mathematics, the myth may loom larger that this is *the* way to success. But the women portrayed here did not have the luxury to "focus on mathematics to the exclusion of other social, personal, and intellectual concerns"—nor do most of their successors today.

Goodman, A.W., *A Victim of the Vietnam War, The Story of Virginia Hanly*, Pentland Press, 2000; viii + 130 pp, $18.50. ISBN 1-57197-972026.

This book recounts the story of a woman who definitely could not follow the "mathematical life course" described by Murray (see preceding review) nor did she find a viable alternative. Excited by mathematics as an undergraduate, Virginia Hanly—like many mathematicians and students of mathematics—became engrossed in opposing the Vietnam War, to the point of forceful nonviolent direct action. Author Goodman, a retired complex analyst who was her teacher, prints letters from her and other materials, as a kind of personal memorial to her. The story is tragic—she was an orphan, she was "probably the best student that I had in 46 years of teaching," and she committed suicide—but the gaps in her earlier and later life leave the reader unsatisfied. Her letters ended eight years before her death, after a sharp but polite exchange. Goodman suspects the FBI of hounding her to death; his Freedom of Information Act request about her FBI file is still pending. But since he cannot tell us anything from her last eight years, the motives that he attributes for her suicide can only be speculation based upon extrapolation. (Since he may have learned only in 1998 of her 1975 death, the long elapsed time would unfortunately have made it difficult to learn much more, though Web inquiries and hiring a detective—plus this book itself—might produce some leads.)

Berlekamp, Edwin, *The Dots-and-Boxes Game: Sophisticated Child's Play*, A K Peters, 2000; xii + 131 pp, $14.95. ISBN 1-56881-129-2.

"[P]erfect play at 3×3 Dots-and-Boxes is simpler than perfect play at 3×3 Tic-Tac-Toe. Yet the latter is known by many; the former, by very, very few." I remember being introduced to the game of Dots-and-Boxes by my English teacher in 9th grade, which is much later in life than author Berlekamp learned it. (For those who still haven't met the game: "Two players start from a rectangular array of dots and take turns to join two horizontally or vertically adjacent dots. If a player competes the fourth side of a square (box) he initials that box and must then draw another line. When all the boxes have been completed the game ends and whoever has initialed more boxes is declared the winner.") I did not realize that the game had any mathematical theory until I read Chapter 16 of *Winning Ways for Your Mathematical Plays* (1982) (by the author, John Conway, and Richard Guy; soon there will be a new edition from A K Peters). This exposition expands on that chapter, includes 100 positions to solve, and gets the reader started in combinatorial game theory.

NEWS AND LETTERS

Carl B. Allendoerfer Award – 2000

The Carl B. Allendoerfer Awards, established in 1976, are made to authors of expository articles published in *Mathematics Magazine*. Carl B. Allendoerfer, a distinguished mathematician at the University of Washington, served as President of the Mathematical Association of America, 1959–60. This year's award was presented at the July 2000 Mathfest, in Los Angeles. The citation follows.

Donald Teets and Karen Whitehead, "The Discovery of Ceres: How Gauss Became Famous," Mathematics Magazine 72 (April 1999). This engaging article tells the story of Carl Friedrich Gauss's computation of the orbit of the planetoid Ceres Ferdinandea shortly after its discovery by the Italian astronomer Joseph Piazzi in 1801. The problem Gauss solved was the following: From three geocentric observations of a planet, find two heliocentric vectors approximating the planet's position at two different times. Using these heliocentric vectors, the orbit of a planet can be completely described. Remarkably, Gauss solved this problem with only relatively elementary algebra and trigonometry. The article begins with a wonderfully written introduction and the fascinating historical background. It then goes on to present clearly the astronomical terminology and Gauss's solution to the problem. Through the article, the reader gains an appreciation of the significant role that astronomy plays in the history of mathematics. It also allows us a glimpse into the incredible mind and creative genius of Gauss.

Biographical Notes Having been born and raised in Boulder, Colorado, **Donald Teets** found the University of Colorado a natural choice for his undergraduate studies in mathematics. He received his bachelor's degree from C.U. in 1978, then went on to obtain a master's degree in mathematics from Colorado State University in 1982, and a Doctor of Arts in mathematics from Idaho State University in 1988. Since that time, Teets has taught mathematics at the South Dakota School of Mines and Technology, a small engineering and science college in Rapid City. In addition, he has served as Chair of the Department of Mathematics and Computer Science for the last three years. His current research interests lie somewhere in the intersection of the three great fields of mathematics, history, and astronomy, as reflected by the Gauss–Ceres article. When Donald Teets is not doing mathematics, he is likely to be found helping out at his wife's veterinary clinic, playing with his seven-year-old son, or pursuing his passion for outdoor sports, such as rock climbing, backpacking, skiing, and bicycling.

Following an eclectic start that included a B.A. in Germany, teaching English in Uganda, and working as a computer programmer, **Karen Whitehead** earned her Ph.D. in mathematics from the University of Minnesota in 1982. She has been at the South Dakota School of Mines and Technology since 1981 where, up until three years ago, she taught mathematics and an occasional computer science course. Karen Whitehead has served in a variety of positions from Assistant Professor to her current post as Vice President for Academic Affairs. Her professional interests range from improving mathematics preparation to applications of neural networks to the history of mathematics. Karen relaxes by gardening, hiking, biking, and reading. Her major avocation, however, is music. She has sung with student choral groups for 16 years, is a substitute organist, and serves on the board of a local chamber music society.

Letters to the Editor

Dear Editor:

The result presented in "How Many Magic Squares Are There?", by Libis, Phillips, and Spall (this MAGAZINE, Feb. 2000, p. 57), is incorrect.

The authors address the following question: How many 4×4 multiplicative magic squares can be constructed using the 16 divisors of 1995? To this end they construct a mapping f from the set of multiplicative squares, $M_{c,n}$, to the set of 880 additive magic squares, A_n, formable from the numbers $0, 1, \ldots, 15$.

The authors assert, correctly, that f is injective. But their further claim, "Checking that f is also surjective is left to the reader.", is mistaken, as an example will show.

Compare the additive magic square below with its image under the reverse mapping, $A_n \to M_{c,n}$:

$$\begin{bmatrix} 0 & 2 & 13 & 15 \\ 14 & 12 & 3 & 1 \\ 9 & 5 & 10 & 6 \\ 7 & 11 & 4 & 8 \end{bmatrix} \longmapsto \begin{bmatrix} 1 & 7 & 285 & 1995 \\ 105 & 15 & 133 & 19 \\ 57 & 95 & 21 & 35 \\ 665 & 399 & 5 & 3 \end{bmatrix}$$

The square at right, which uses the 16 divisors of 1995, is produced from the binary representation of each number in the square at left by replacing 1s with 3, 5, 7, or 19, depending on position:

$$\begin{bmatrix} 0000 & 0010 & 1101 & 1111 \\ 1110 & 1100 & 0011 & 0001 \\ 1001 & 0101 & 1010 & 0110 \\ 0111 & 1011 & 0100 & 1000 \end{bmatrix} \longmapsto \begin{bmatrix} 1 & 7 & 3\cdot 5\cdot 19 & 3\cdot 5\cdot 7\cdot 19 \\ 3\cdot 5\cdot 7 & 3\cdot 5 & 7\cdot 19 & 19 \\ 3\cdot 19 & 5\cdot 19 & 3\cdot 7 & 5\cdot 7 \\ 5\cdot 7\cdot 19 & 3\cdot 7\cdot 19 & 5 & 3 \end{bmatrix}$$

Yet the square at right is *not* multiplicatively magic; the two diagonal products differ. Hence f is not surjective, a bijection existing rather between a *subset* of A_n and $M_{c,n}$. What characterizes this subset?

Consider each binary string in the left-hand square above as a 4-vector. The vectorial sum of each row and column is then $[2, 2, 2, 2]$, while those of the diagonals are $[3, 1, 1, 0]$ and $[1, 3, 3, 4]$, respectively. A moment's thought reveals that the image square will be multiplicatively magic if and only if the binary square is vectorially magic. The number of multiplicative squares is thus equal to the number of (vectorially) additive magic squares that can be formed using this set of vectors.

To discover this number, observe that the 16 4-vectors correspond to the vertices of a unit hypercube centered on the point $[1/2, 1/2, 1/2, 1/2]$ in 4-space. Using methods similar to those in [1], we can show that the entries in any such magic square can be permuted to produce 384 distinct squares, one for each symmetry of the hypercube. Lack of space prohibits detailed explication, but in fact the same entries can be mapped onto the vertices of a hypercube in 10 further, fundamentally distinct ways, to yield a total of $11 \times 384 = 4224$ magic squares, rotations and reflections included. The total number of multiplicative squares using the divisors of 1995, rotations and reflection excluded, is thus $4224/8 = 528$, a figure independently confirmed using a brute force counting algorithm.

1. B. Rosser and R.J. Walker, On the transformation group for diabolic magic squares of order four, *Bull. Amer. Math. Soc.* XLIV, 1938, 416–20.

Lee Sallows, Johannaweg 12, Nijmegen 6523 MA, The Netherlands

Dear Editor:

Ay Caramba! Lee Sallows is right; while our f is certainly injective, it is not surjective. The problem with surjectivity is subtle: the diagonals of a multiplicative square corresponding to an additive magic square need not multiply to the proper number.

But we think the fact that f is only injective, and not surjective, actually makes our paper *more* interesting. It highlights the uniqueness of multiplicative magic squares; their image under our f forms a proper subclass of the class of additive magic squares. And this opens the door to more work on describing this class of magic squares.

Carl Libis, Richard Stockton College of NJ, Pomona, NJ 08240–0195
J.D. Phillips, Saint Mary's College of California, Moraga, CA 94575

Dear Editor:

In their article "A Postmodern View of Fractions and the Reciprocals of Fermat Primes" (this MAGAZINE, April 2000, 83–97), Rafe Jones and Jan Pearce conclude with two questions:

> *Question 1.* Does there exist, for each positive integer n, a natural number k such that $2^n(2k+1)+1$ is prime?

By Dirichlet's theorem on primes in arithmetic progressions, the answer is "Yes." In fact, for each n, there are infinitely many values of k such that $2^n(2k+1)+1 = 2^{n+1}k + (2^n+1)$ is prime. (Dirichlet's theorem asserts that $\{Ak+B\}$ takes on infinitely many prime values, provided that $(A, B) = 1$; clearly $(2^{n+1}, 2^n+1) = 1$.)

> *Question 2.* How many Fermat primes are there?

The partial answer the authors give begins:

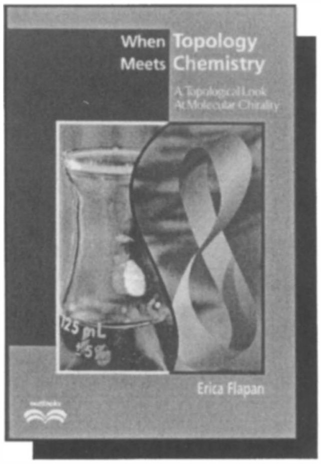
> No one has any idea; we know only that there are at least five. Pierre de Fermat thought that all numbers of the form $2^{2^k}+1$ were prime, but history has proven otherwise: All the numbers generated using $k = 5, \ldots, 11$ have turned out to be composite, as well as selected others, including the monstrous $2^{2^{23471}}+1\ldots$."

Continued on page 329

The Mathematical Association of America

When Topology Meets Chemistry

Erica Flapan



The MAA is proud to announce the publication of the first book in a series of interdisciplinary titles to be co-published with Cambridge University Press

Series: Outlooks

The applications of topological techniques for understanding molecular structures have become increasingly important over the past thirty years. In this topology text, the reader will learn about knot theory, 3-dimensional manifolds, and the topology of embedded graphs, while learning the role these play in understanding molecular structures. Most of the results that are described in the text are motivated by questions asked by chemists or molecular biologists, though the results themselves often go beyond answering the original question asked. There is no specific mathematical or chemical prerequisite; all the relevant background is provided. The text is enhanced by nearly 200 illustrations and more than 100 exercises.

Reading this fascinating book, undergraduate mathematics students can escape the world of pure abstract theory and enter that of real molecules, while chemists and biologists will find simple and clear but rigorous definitions of mathematical concepts they handle intuitively in their work.

Contents:

Stereochemical Topology; Detecting Chirality; Msbius Ladders and Related Molecular Graphs; Different Types of Chirality and Achirality; Intrinsic Topological Properties of Graphs; Symmetries of Embedded Graphs; Topology of DNA..

Catalog Code: TPC/JR 250 pp., Paperbound, 2000 ISBN 0-521-66482-9 List: $29.95 MAA Member: $24.95

Name________________ Credit Card No.________________

Address________________ Signature________________ Exp. Date___/___

City________________ Qty______ Price $________ Amount $________

State________ Zip ________ Shipping and Handling $________

Phone________________ Catalog Code: TPC/JR Total $________

Shipping and Handling: Postage and handling are charged as follows: USA orders (shipped via UPS): $2.95 for the first book, and $1.00 for each additional book. Canadian orders: $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. Overseas Orders: $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit card orders are accepted for all customers. All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Order Via:

Phone: 1 (800) 331.1622
Fax: (301) 206.9789
Mail: Mathematical Association of America
PO Box 91112
Washington, DC 20090-1112
Web: www.maa.org

CONTENTS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW
Washington, D.C. 20036