AUTHORS

TITLES

PROBLEMS

P, S, Q, *and* C *refers to Proposals, Solutions, Quickies, and Comments, respectively. Page numbers are given in parentheses. Thus,* P1338(56) *refers to proposal number* 1338, *which appears on page 56.*

Vol. 66 No. 5 December 1993

MATHEMATICS MAGAZINE

- Pythagoras—Space Traveller with a One-Way Ticket
- $6/\pi^2$
- Derangements and Bell Numbers

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

The aim of *Mathematics Magazine* is to provide lively and appealing mathematical exposition. This is not a research journal and, in general, the terse style appropriate for such a journal (lemma-theorem-proof-corollary) is not appropriate for an article 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. Articles on pedagogy alone, unaccompanied by interesting mathematics, are not suitable. Neither are articles consisting mainly of computer programs unless these are essential to the presentation of some good mathematics. Manuscripts on history are especially welcome, as are those showing relationships between various branches of mathematics and between mathematics and other disciplines.

The full statement of editorial policy appears in this *Magazine*, Vol. 64, pp. 71–72, and is available from the Editor. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, nor published by another journal or publisher.

Send new manuscripts to: Martha Siegel, Editor, Mathematics Magazine, Towson State University, Towson, MD 21204. Manuscripts should be typewritten and double spaced and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit the original and two 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. Illustrations should be carefully prepared on separate sheets in black ink, the original without lettering and two copies with lettering added. Do not use staples.

Cover: Pythagoras, from a fresco by Raphael. Taken from Elisha Scott Loomis, *The Pythogorean Proposition*, National Council of Teachers of Mathematics, Washington, DC, 1968, orginally from the collection of Dr. Jehuthiel Ginsburg, publisher of *Scripta Mathematica*.

AUTHORS

Hans R. Schneebeli studied mathematics at the Swiss Federal Institute of Technology (ETH), Zürich, and at the Weizmann Institute of Science, Rehovot, Israel. His Ph.D. thesis and early papers dealt with homological algebra and group theory. Since 1979 he teaches mainly at a Swiss Gymnasium. His principal effort is directed at renewing the curriculum, taking advantage of new technologies, and making mathematics (almost) as fascinating a subject to his students as for himself. He is interested in finding topics that are meaningful to both students and mathematics. Being involved with mathematical talent search, he encourages students to be active in junior research projects or in applying mathematics to help solving real problems. The present article grew out of the question 'How to lift theorems from plane geometry to three space and beyond?'

Vol. 66 No. 5 December 1993

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.

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

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

PRINTED IN THE UNITED STATES OF AMERICA

Statement of Ownership, Management and Circulation
(Required by 39 U.S.C. 3685)

1A. Title of Publication	1B. Publication No.	2. Date of Filing
MATHEMATICS MAGAZINE	0 0 2 5 5 7 0 X	October 10, 1993

3. Frequency of Issue	3A. No. of Issues Published Annually	3B. Annual Subscription Price
bi-monthly except July and August	Five	Library $77 Indiv. Member $18

4. Complete Mailing Address of Known Office of Publication *(Street, City, County, State and ZIP+4 Code) (Not printers)*
1529 Eighteenth St., N.W., Washington, DC 20036-1385

5. Complete Mailing Address of the Headquarters of General Business Offices of the Publisher *(Not printer)*
1529 Eighteenth St., N.W., Washington, DC 20036-1385

6. Full Names and Complete Mailing Address of Publisher, Editor, and Managing Editor *(This item MUST NOT be blank)*
Publisher *(Name and Complete Mailing Address)*
The Mathematical Association of America, 1529 Eighteenth St., N.W., Washington, DC, 20036-1385
Editor *(Name and Complete Mailing Address)*
Martha J. Siegel, Towson State Univ., Dept. of Mathematics, Baltimore, MD 21204
Managing Editor *(Name and Complete Mailing Address)*
Harry Waldman, MAA, 1529 Eighteenth St., N.W., Washington, DC 20036-1385

7. Owner *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.) (Item must be completed.)*

Full Name	Complete Mailing Address
The Mathematical Association of America, Inc.	1529 Eighteenth St., N.W. Washington, DC 20036-1385

8. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages or Other Securities *(If there are none, so state)*

Full Name	Complete Mailing Address
None	

9. For Completion by Nonprofit Organizations Authorized To Mail at Special Rates *(DMM Section 424.12 only)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*
(1) [X] Has Not Changed During Preceding 12 Months
(2) [ ] Has Changed During Preceding 12 Months
(If changed, publisher must submit explanation of change with this statement.)

10. Extent and Nature of Circulation *(See instructions on reverse side)*	Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
A. Total No. Copies *(Net Press Run)*	20,695	19,524
B. Paid and/or Requested Circulation 1. Sales through dealers and carriers, street vendors and counter sales	0	0
2. Mail Subscription *(Paid and/or requested)*	16,483	13,355
C. Total Paid and/or Requested Circulation *(Sum of 10B1 and 10B2)*	16,483	13,355
D. Free Distribution by Mail, Carrier or Other Means Samples, Complimentary, and Other Free Copies	1,263	2,270
E. Total Distribution *(Sum of C and D)*	17,746	15,625
F. Copies Not Distributed 1. Office use, left over, unaccounted, spoiled after printing	2,949	3,899
2. Return from News Agents	0	0
G. TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	20,695	19,524

11. I certify that the statements made by me above are correct and complete
Signature and Title of Editor, Publisher, Business Manager, or Owner
Kay Lamont
Membership Manager

PS Form 3526, January 1991 *(See instructions on reverse)*

ARTICLES

Pythagoras—Space Traveller with a One-Way Ticket

H. R. SCHNEEBELI
Aargauische Kantonsschule
Baden, Switzerland

1. Introduction

The Pythagorean Theorem of plane geometry says that for all right triangles with legs a, b and hypotenuse c the relation $a^2 + b^2 = c^2$ holds. Conversely, if three segments a, b, c satisfy the relation $a^2 + b^2 = c^2$, then they define a right triangle with hypotenuse c.

The purpose of this note is to investigate some generalizations of this famous theorem. We shall look at the three-dimensional case only, but the game could be played in higher dimensions as well. Our point of departure is a couple of logically equivalent characterizations of right triangles in plane geometry. As we generalize the two key properties to the context of space geometry, it turns out that two inequivalent notions result, those of *rectangular* and of *Pythagorean* tetrahedra.

A tetrahedron $ABCD$ is termed *rectangular at the vertex* D if all the edges meeting at D are mutually perpendicular.

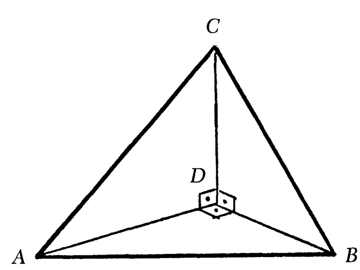


FIGURE 1
A rectangular tetrahedron.

Some special notation: To each vertex X of a tetrahedron $ABCD$, we associate the face opposite to X, i.e. the unique face of $ABCD$ not containing X. This face is denoted by X^*, and its area by x. If a tetrahedron $ABCD$ is rectangular at D then we might speculate that the relation

$$a^2 + b^2 + c^2 = d^2 \tag{1}$$

is a three-dimensional analogue of Pythagoras' Theorem.

Let us call a tetrahedron *ABCD Pythagorean with distinguished face* D^* if (1) is a true statement. The distinguished face will always be D^* unless explicitly stated otherwise. A word of caution: *'Pythagorean' in this context does not imply integral side length as is customary in number theory.*

Now two questions seem to be in order:

(Q1) Is every rectangular tetrahedron Pythagorean?

(Q2) Is every Pythagorean tetrahedron rectangular?

The answers to these questions are as follows:

(A1) *Every rectangular tetrahedron is Pythagorean.*

(A2) *Not every Pythagorean tetrahedron is rectangular.*

While (A1) is a well-known exercise (cf [3], Problem 32) and extends to higher dimensions as explained in [5], the construction of a Pythagorean but nonrectangular tetrahedron is one of the goals of Section 2. There we are led to generalize another characterization of right triangles that is traditionally attributed to Thales. We shall refer to the following statement as the '*Theorem of Thales*': *A triangle ABC is right angled at C if and only if the center of the circle circumscribed about ABC lies on the segment AB* (cf [4], Book III, Proposition 31). The circle with the segment AB as one of its diameters will be called the *Thales circle over AB.*

It is natural to ask whether the Theorem of Thales has some three-dimensional analogue. More specifically, let a triangle ABC be given in space. Then the following questions arise:

(Q3) What is the locus of all points D such that the tetrahedron $ABCD$ is rectangular at D?

(Q4) What is the locus of all points D such that the tetrahedron $ABCD$ is Pythagorean with distinguished face D^*?

The main results of Section 2 answer these questions as follows:

(A3) *For a given triangle ABC, the set of all points D such that the tetrahedron ABCD is rectangular at D consists of at most two points. The two-point set occurs if and only if all the angles of ABC are acute (i.e. strictly smaller than a right angle).*

(A4) *For a given triangle ABC the set of all points D such that the tetrahedron ABCD is Pythagorean with distinguished face* $ABC = D^*$ *is an ellipsoid centered at the centroid of* D^*. *The intersection of this ellipsoid with the plane* $\langle ABC \rangle$ *is the ellipse of minimal area containing* D^*.

In plane geometry the law of cosines, which applies to arbitrary triangles, reads as follows:

$$c^2 = a^2 + b^2 - 2ab\cos\gamma \tag{2}$$

where γ is the angle opposite side c. If γ is a right angle, this becomes the Pythagorean theorem, and we see why the right triangles are precisely the Pythagorean ones. This leads to an obvious question:

(Q5) Is there a three-dimensional analogue to the law of cosines that provides us with necessary and sufficient conditions for a tetrahedron $ABCD$ to be rectangular at the vertex D or Pythagorean with distinguished face D^*?

Such a three-dimensional version of the law of cosines will be dealt with in Section 3.

(A5) *Law of cosines*:

Let ABCD be any tetrahedron. Let α *denote the interior angle between the faces with common edge AD,* β, *and* γ *the angles between the faces with common edges BD and CD respectively. Then the relation*

$$d^2 = a^2 + b^2 + c^2 - 2(ab\cos\gamma + bc\cos\alpha + ca\cos\beta) \tag{3}$$

holds.

This result provides us with necessary and sufficient conditions for a tetrahedron $ABCD$ to be Pythagorean.

2. Thales Meets Pythagoras

Suppose a triangle ABC is given. What is the set of all points D such that the tetrahedron $ABCD$ is rectangular, or Pythagorean? We consider the rectangular case in (2.1), the Pythagorean in (2.2). The two cases are related by means of generalizations of the Theorems of Thales and of Pythagoras. If the two properties 'rectangular' and 'Pythagorean' were equivalent, then the two generalizations of Thales' theorem would be equivalent, too. Our discussion will show that this is not the case.

2.1. Existence of rectangular tetrahedra over a given triangle We now make an obvious extension of the Theorem of Thales from plane geometry to space geometry. First, we pass from Thales circles to Thales spheres. Let a segment AB be given in space, and let $\mathcal{S}$ be the sphere, one diameter of which is the given segment. Then the sphere $\mathcal{S}$ consists precisely of the points A, B, and of all other points C such that the two lines joining C to A or B are perpendicular.

We recall some terminology: The *orthocenter* is the common intersection of the three altitudes of a triangle or their extensions. Similarly, if a point common to all four altitudes (or their extensions) of a tetrahedron exists, it is termed the orthocenter. A necessary and sufficient condition for a tetrahedron $ABCD$ to admit an orthocenter is that all pairs of disjoint edges such as AB and CD are mutually perpendicular. This is easily understood by looking first at only two altitudes of a tetrahedron $ABCD$, say p_A through A and p_B through B. The line p_A is perpendicular to the plane $\langle BCD \rangle$ while p_B is perpendicular to $\langle ACD \rangle$. Now the two lines p_A and p_B cannot be parallel since $ABCD$ are in general position. If p_A and p_B intersect then their span is perpendicular to the intersection of the two planes $\langle BCD \rangle$ and $\langle ACD \rangle$, that is the line CD, and conversely. The necessary and sufficient criterion for the existence of an orthocenter now follows if this reasoning is applied to all pairings of altitudes in the tetrahedron $ABCD$.

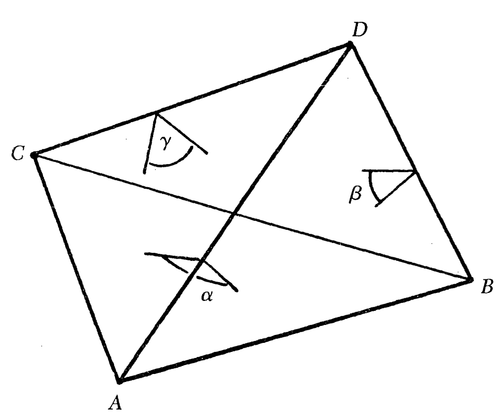


FIGURE 2
Labelling of angles between faces in a tetrahedron.

THEOREM 1 (THEOREM OF THALES, THREE-DIMENSIONAL VERSION, RECTANGULAR CASE).

(i) *Let ABC be a triangle. Then there are tetrahedra ABCD rectangular at D if and only if each angle of the triangle ABC is strictly smaller than a right angle.*

(ii) *If the tetrahedron ABCD is rectangular at D, then the point D is unique up to reflection in the plane $\langle ABC\rangle$, and the line joining the vertex D and its reflection intersects the plane $\langle ABC\rangle$ at the orthocenter of the triangle ABC.*

Proof.

(a) Suppose $ABCD$ is rectangular at D. Choose an orthonormal basis with origin at D and axes along DA, DB, DC respectively. We can now compute the following scalar product

$$\vec{AB}\cdot\vec{AC}=\left(\vec{DB}-\vec{DA}\right)\cdot\left(\vec{DC}-\vec{DA}\right)=\vec{DA}\cdot\vec{DA}>0,$$

the three vectors $\vec{DA}, \vec{DB}, \vec{DC}$ being orthogonal and different from $\vec{0}$. Similar computations yield $\vec{BA}\cdot\vec{BC}>0$ and $\vec{CA}\cdot\vec{CB}>0$. Therefore all three of α, β, γ are smaller than a right angle.

(b) Let ABC be any triangle with strictly acute angles. In order to find a point D such that $ABCD$ is rectangular, we determine the intersection of three Thales spheres, $\mathcal{S}_1$ over AB, $\mathcal{S}_2$ over BC, $\mathcal{S}_3$ over CA. (Here we are using the traditional version of the Theorem of Thales in a three-dimensional context.) Now $\mathcal{S}_1\cap\mathcal{S}_2$ is a circle containing B and spanning a plane normal to the side AC. Similarly, planes normal to AB, BC are determined by $\mathcal{S}_2\cap\mathcal{S}_3, \mathcal{S}_1\cap\mathcal{S}_3$ respectively. Clearly (see FIGURE 3), all three planes meet in a common line l perpendicular to the plane $\langle ABC\rangle$ and passing through the orthocenter of ABC. Since all the angles of the triangle ABC are acute, this orthocenter lies inside ABC, and $\mathcal{S}_1\cap\mathcal{S}_2\cap\mathcal{S}_3$ consists of two distinct points D_1, D_2 on the line l.

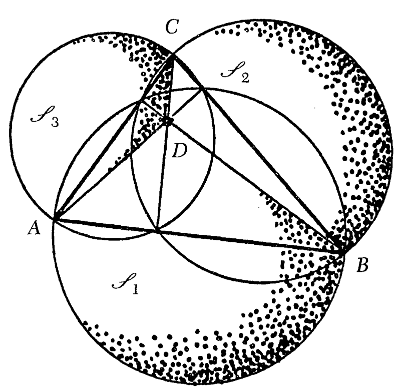


FIGURE 3
Normal projection of three Thales spheres onto the plane $\langle ABC\rangle$.

Remark. The four-dimensional version of Theorem 1 is rather interesting. If $ABCDE$ is a four-simplex that is rectangular at E, then all edges meeting at some vertex of $ABCD$ do so at acute angles, and moreover the tetrahedron $E^*=ABCD$ admits an orthocenter. The proof given for Theorem 1 generalizes to four-space, provided that E^* admits an orthocenter and all edges of E^* meet at acute angles.

2.2. Existence of Pythagorean tetrahedra with a prescribed face The Theorem of Thales may be restated: *A triangle ABC is Pythagorean with distinguished segment $C^*=AB$ if and only if C lies on the Thales circle over AB.*

We shall follow this pattern in our second generalization of the Theorem of Thales. Our proof will use the following terms: A triangle is *singular* if its vertices are collinear and *nonsingular* otherwise; similarly, a tetrahedron is singular if its vertices are coplanar and nonsingular otherwise.

THEOREM 2 (THEOREM OF THALES, THREE-DIMENSIONAL VERSION, PYTHAGOREAN CASE). *Let ABC be any nonsingular triangle, and $\mathscr{P}$ the plane $\langle ABC \rangle$. Then the set of all points D such that the (possibly singular) tetrahedron ABCD is Pythagorean with distinguished face D^* is an ellipsoid $\mathscr{E}$ centered at the centroid of ABC. The intersection $\mathscr{E}' = \mathscr{E} \cap \mathscr{P}$ is the ellipse of minimal area containing ABC.*

Proof. (a) We treat the problem algebraically. Choose an orthonormal coordinate system and let (x, y, z) be the coordinates of a point D such that the tetrahedron $ABCD$ is Pythagorean and $D^* = ABC$ the given face. Now equation (1) can be formulated as a quadratic equation in x, y, and z. (In principle, we could work out this equation explicitly and determine its normal form and the corresponding quadratic surface using standard algorithms from linear algebra. Instead of doing this, we use an approach that may give more geometric insight.)

(b) Next we show that the solution set contains at least three distinct points. If we let D coincide with A (yielding a singular tetrahedron) then equation (1) becomes $a^2 + 0 + 0 = d^2$. Letting D coincide with B and C yields two more such solutions, and three distinct solutions to equation (1) have been found.

(c) If PQR is a triangle of area m, then the set of all points Z in space such that the area of the triangle PQZ does not exceed m is a solid cylinder with axis PQ and radius equal to the distance of the point R from this axis. We use this fact to show that the solution set of equation (1) is bounded.

Indeed, if (1) holds then none of a, b, or c exceeds d. We conclude that any candidate for point D belongs to the intersection of three cylinders, each of which has one edge of ABC as axis, and radius determined by the remaining vertex. Because the axes of the three cylinders are not parallel, the solution set is bounded.

(d) We consult the list of affine normal forms for real quadratic forms on $\mathbf{R}^3$ (cf. [1], ch. 15) and find that among the corresponding quadrics only the empty set, singletons, and ellipsoids satisfy boundedness. Of these, only ellipsoids contain at least three different points. We conclude that the solution set, henceforth denoted by $\mathscr{E}$, is an ellipsoid.

(e) Finally we describe the ellipsoid $\mathscr{E}$. Clearly $\mathscr{E}$ is symmetric with respect to the plane $\mathscr{P}$, so two of the principal axes of $\mathscr{E}$ are axes of the ellipse $\mathscr{E}' = \mathscr{E} \cap \mathscr{P}$, while the third one is perpendicular to $\mathscr{P}$ and passes through the center of $\mathscr{E}'$. In order to see how the points A, B, C are related to this center, we first consider a special case.

Let $A_0B_0C_0$ be an equilateral triangle. Then the corresponding ellipsoid $\mathscr{E}_0$ and ellipse $\mathscr{E}_0'$ inherit the triangle's symmetry properties. In particular $\mathscr{E}_0'$ is invariant under rotations through $\pm 2\pi/3$ about its center. Hence $\mathscr{E}_0'$ is the circle circumscribed about $A_0B_0C_0$. Clearly $\mathscr{E}_0'$ has minimal area among all ellipses containing $A_0B_0C_0$ and its center is the centroid of the triangle. Moreover, the principal axes of $\mathscr{E}_0$ have lengths $s/\sqrt{3}, s/\sqrt{3}, s/\sqrt{6}$, where $s = |A_0B_0|$. This completes the description of $\mathscr{E}_0$.

For an arbitrary nonsingular triangle ABC, let $A_0B_0C_0$ be an equilateral triangle in the same plane $\mathscr{P} = \langle ABC \rangle$. There is a regular affine transformation Φ of $\mathscr{P}$ that carries A_0 to A, B_0 to B, and C_0 to C, and changes the area of each measurable set of $\mathscr{P}$ by multiplication by $\kappa = |\det \Phi|$ (see [2], section 3.9). Now if D_0 is any point in $\mathscr{P}$ and $D = \Phi(D_0)$, then $A_0B_0C_0D_0$ and $ABCD$ are singular tetrahedra and the area of each face of $ABCD$ is κ times the area of the corresponding face of $A_0B_0C_0D_0$. It follows immediately that $ABCD = \Phi(A_0B_0C_0D_0)$ is Pythagorean relative to D^* if and only if $A_0B_0C_0D_0$ is Pythagorean relative to D_0^*. Moreover, the ratio of the areas of $\mathcal{E}_0'$ and $A_0B_0C_0$ remains unchanged upon passing to $\Phi(\mathcal{E}_0') = \mathcal{E}'$ and $\Phi(A_0B_0C_0) = ABC$. Hence $\mathcal{E}' = \Phi(\mathcal{E}_0')$ is the unique ellipse of minimal area containing D^*. Furthermore, its center is the centroid of ABC. The proof is complete.

COROLLARY. *There exist Pythagorean tetrahedra that are not rectangular.*

Proof. Let ABC be a triangle with an obtuse angle. By Theorem 2 there exists a point D such that $ABCD$ is Pythagorean relative to D^* but according to Theorem 1, (i), $ABCD$ cannot be rectangular at D.

To construct explicit examples, simply avoid the situation of Theorem 1, (ii). Take any triangle ABC, and choose a line l perpendicular to the plane $\langle ABC \rangle$ such that l hits some point other than the orthocenter of ABC inside the ellipse $\mathcal{E}'$. Then l cuts out two points D of the ellipsoid $\mathcal{E}$ such that $ABCD$ is Pythagorean but not rectangular.

Example.

$A = (1,0,0) \qquad B = (-1,2,0) \qquad C = (-1,-2,0) \qquad D = (0,0,\pm\sqrt{1.25})$

In this case ABC is right angled at A and $(0,0,0)$ is not the orthocenter.

3. The Law of Cosines for Tetrahedra

We may look at the Pythagorean theorem as a special case of the law of cosines. Similarly, the Pythagorean theorem for rectangular tetrahedra is a special case of the following.

THEOREM 3 (THE LAW OF COSINES FOR TETRAHEDRA). *Let $ABCD$ be a nonsingular tetrahedron, and denote by α, β, γ the interior angles between the faces with common edge AD, BD, CD respectively. Then the relation*

$$d^2 = a^2 + b^2 + c^2 - 2(ab\cos\gamma + bc\cos\alpha + ca\cos\beta)$$

holds.

Proof. We imitate the standard argument that deduces the law of cosines from properties of the scalar product. For each face X^* of $ABCD$ let $\vec{x}$ be the outward normal vector on face X^* with length $|\vec{x}| = x$, the area of X^*. Each $\vec{x}$ may be written as a cross product involving the vectors $\overrightarrow{DA}$, $\overrightarrow{DB}$, $\overrightarrow{DC}$ that form the edges of the tetrahedron not in ABC.

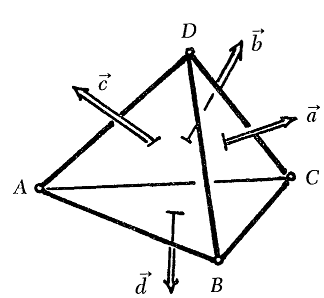


FIGURE 4
The tetrahedron $ABCD$ and its outward normals.

If we assume (without loss of generality) that $(\overrightarrow{DA}, \overrightarrow{DB}, \overrightarrow{DC})$ is a positively oriented, ordered basis of $\mathbf{R}^3$ then

$$2\vec{a} = \overrightarrow{DC} \times \overrightarrow{DB}, \qquad 2\vec{b} = \overrightarrow{DA} \times \overrightarrow{DC}, \qquad 2\vec{c} = \overrightarrow{DB} \times \overrightarrow{DA},$$

$$2\vec{d} = \overrightarrow{AB} \times \overrightarrow{AC} = \left(\overrightarrow{DB} - \overrightarrow{DA}\right) \times \left(\overrightarrow{DC} - \overrightarrow{DA}\right)$$

$$= \overrightarrow{DB} \times \overrightarrow{DC} - \overrightarrow{DB} \times \overrightarrow{DA} - \overrightarrow{DA} \times \overrightarrow{DC} = -2\vec{a} - 2\vec{b} - 2\vec{c},$$

whence

$$\vec{a} + \vec{b} + \vec{c} + \vec{d} = \vec{0}.$$

(More generally, this relation follows from the fact that the tetrahedron is a closed orientable polyhedral surface.)

Hence

$$\vec{d} \cdot \vec{d} = \left(\vec{a} + \vec{b} + \vec{c}\right) \cdot \left(\vec{a} + \vec{b} + \vec{c}\right),$$

and calculating the scalar products gives

$$d^2 = a^2 + b^2 + c^2 + 2(ab \cos \gamma' + bc \cos \alpha' + ca \cos \beta')$$

with $\alpha' = \pi - \alpha$, $\beta' = \pi - \beta$, $\gamma' = \pi - \gamma$. Since $\cos(\pi - \phi) = -\cos \phi$, the relation (3) follows.

COROLLARY. *A tetrahedron ABCD is*

(i) *Pythagorean if and only if* $ab \cos \gamma + bc \cos \alpha + ca \cos \beta = 0$;

(ii) *rectangular if and only if* $ab \cos \gamma = 0$, $bc \cos \alpha = 0$ *and* $ca \cos \beta = 0$.

Remarks. In plane geometry, the law of cosines connects four parameters of a triangle in such a way that three of them determine the fourth (possibly not uniquely). In space geometry, seven parameters are linked by the law of cosines and six of them fix the remaining one.

Given values of a, b, c, the additional requirement that the tetrahedron $ABCD$ be rectangular at D defines three more parameters via statement (ii) and thus leads to a solution unique up to reflections. In the Pythagorean context, the three data a, b, c together with statement (i) leaves two degrees of freedom for the remaining parameters. In general, a rectangular n-simplex is completely determined by the content of n of its boundary $(n-1)$-simplices and the choice of the distinguished face, while the same set of data for a Pythagorean n-simplex leaves $\frac{1}{2}n(n-1) - 1$ parameters undetermined.

Statement (ii) of the Corollary is of course just the dual form of our definition of *rectangular* tetrahedra: A tetrahedron is rectangular at D if and only if the *faces* A^*, B^*, C^* intersecting at D are mutually perpendicular.

Acknowledgement. The organization and exposition of this paper has greatly profited from the referees' comments as well as the Editor's advice. Their generous help is hereby gratefully acknowledged.

REFERENCES

1. Marcel Berger, *Geometry II*, Universitext, Springer Verlag, Berlin, 1987.
2. Judith N. Cederberg, *A Course in Modern Geometries*, Springer Verlag, Berlin, 1989.
3. Judita Cofman, *What to Solve?*, Clarendon Press, Oxford, 1990.
4. T. L. Heath, *The Thirteen Books of Euclid's Elements*, 2nd edition, Dover Publications, Mineola, NY, 1956.
5. Shwu-Yeng T. Lin and You Feng Lin, The n-dimensional Pythagorean Theorem, *Linear and Multilinear Algebra*, Vol. 26, 1990, pp. 9–13.

$6/\pi^2$

GARETH A. JONES
University of Southampton
Southampton SO9 5NH, UK

1. Introduction

There is a common tendency for teachers and writers of mathematics to delay mentioning results until enough machinery has been introduced to prove them rigorously. This has the advantage of allowing a logical and self-contained exposition, but it can sometimes prevent students from discovering results that they are capable of understanding and believing long before they are ready for complete proofs. Experience shows that heuristic proofs, well-chosen examples, and experimental evidence may provide insights that would be obscured by the technicalities needed for full rigour; this can, and should, come later when confidence and understanding have been developed at a more elementary level.

As an example, consider Theorems 332 and 333 of the classic book [**12**] on the theory of numbers by Hardy and Wright. These state that:

"The probability that two integers should be prime to one another is $6/\pi^2$", and

"The probability that a number should be quadratfrei [= square-free, i.e. a product of distinct primes] is $6/\pi^2$".

These very attractive results (based on theorems of Dirichlet and Gegenbauer respectively) raise several questions:

Why should π appear in a context that seems to have no connection with geometry?

Why should these two quite separate events have the same probability?

Are there other similar coincidences?

Hardy and Wright (along with most authors in this field) do not answer these questions; indeed, they do not even comment on the coincidence, leaving it to their proofs to show that the same probability occurs in two quite distinct ways. Their proofs are short, but not elementary, being part of a comprehensive development including Dirichlet series, the Möbius inversion formula, and asymptotic estimates. The aim of this note is to provide elementary and (one hopes) reasonably convincing heuristic proofs of these results, and thus to introduce the Riemann zeta-function and the Möbius function in a way that should be acceptable to a student who might be daunted by the more traditional expositions (which can be read later).

Although the probabilistic interpretations may be unfamiliar, the reader will find no "new" mathematics here, most of the results being due to Euler; indeed, this note can be read as a plea for more mathematics to be introduced in the style of Euler (insight and ingenuity, without too much justification), and only later reinforced in the most rigorous style of more modern times. This is, after all, how most of us conduct our research, "seeing" results before writing down formal proofs.

2. Random Integers

There are several technical problems involved in discussing randomly chosen integers $x \in \mathbb{N}$ (or indeed elements of any infinite set). For example, if p_n denotes the probability $\Pr(x = n)$ that n is chosen, then one expects that

$$\sum_{n=1}^{\infty} p_n = 1. \tag{2.1}$$

However, if one wants all integers n to have the same status, then p_n must be a constant, independent of n, so that Σp_n is either 0 or divergent.

We can avoid some of these difficulties by assigning probabilities only to certain subsets $S \subseteq \mathbb{N}$, rather than to individual integers. One way of doing this (used by Hardy and Wright) is to regard $\Pr(x \in S)$ as, in some sense, the proportion of $\mathbb{N}$ lying in S, and to approximate this by restricting the choice of x to the finite subset

$$\mathbb{N}_N = \{1, 2, \dots, N\},$$

for some integer $N \geq 1$. Now the proportion of $\mathbb{N}_N$ lying in S is

$$\frac{|S \cap \mathbb{N}_N|}{|\mathbb{N}_N|} = \frac{|S \cap \mathbb{N}_N|}{N},$$

and it is reasonable to expect better approximations from larger samples $\mathbb{N}_N$, so we define the probability that $x \in S$ to be

$$\Pr(x \in S) = \lim_{N \to \infty} \frac{|S \cap \mathbb{N}_N|}{N} \tag{2.2}$$

provided this limit exists. (It is not hard to find sets S for which this limit does not exist.)

This definition of probability has many of the properties we would normally expect: For instance, if S and T are disjoint subsets of $\mathbb{N}$ then one easily sees that

$$\Pr(x \in S \cup T) = \Pr(x \in S) + \Pr(x \in T)$$

provided both probabilities on the right-hand side exist. This rule extends, in the obvious way, to all finite disjoint unions, but not unfortunately to all infinite disjoint unions: In particular, (2.1) fails since (2.2) gives $p_n = 0$ for all n.

For certain sets S of interest to us, it is straightforward both to prove that $\Pr(x \in S)$ exists and to evaluate it; for example, if S is a congruence class mod n for some $n \geq 1$, then

$$\left\lfloor \frac{N}{n} \right\rfloor \leq |S \cap \mathbb{N}_N| \leq \left\lfloor \frac{N}{n} \right\rfloor + 1$$

(where $\left\lfloor \frac{N}{n} \right\rfloor$ denotes the integer part of $\frac{N}{n}$), so

$$\begin{aligned} \Pr(x \in S) &= \lim_{N \to \infty} \left\lfloor \frac{N}{n} \right\rfloor / N \\ &= \frac{1}{n}, \end{aligned}$$

as one would expect. Now the Chinese Remainder Theorem ([**12**], §8.1) asserts that if $(m, n) = 1$ then the solutions x of any pair of simultaneous congruences

$$x \equiv a \bmod m, \tag{2.3}$$

$$x \equiv b \bmod n \tag{2.4}$$

form a single congruence class $\bmod(mn)$; hence

$$\begin{aligned}\Pr(x \equiv a \bmod m \quad \text{and} \quad x \equiv b \bmod n) \\ = \frac{1}{mn} \\ = \Pr(x \equiv a \bmod m) \cdot \Pr(x \equiv b \bmod n),\end{aligned}$$

that is, the events (2.3) and (2.4) are statistically independent. (This fails if $(m, n) > 1$.)

There is an obvious extension of definition (2.2) to subsets $S \subseteq \mathbb{N}^k$ consisting of k-tuples $\underline{x} = (x_1, \dots, x_k)$ where each $x_i \in \mathbb{N}$; we define

$$\Pr(\underline{x} \in S) = \lim_{N \to \infty} \frac{|S \cap (\mathbb{N}_N)^k|}{N^k}$$

provided this limit exists. In particular, if $S = S_1 \times \cdots \times S_k$ where $S_i \subseteq \mathbb{N}$ and $\Pr(x_i \in S_i)$ exists for each i, then

$$\Pr(\underline{x} \in S) = \prod_{i=1}^{k} \Pr(x_i \in S_i).$$

Unfortunately, it is significantly harder to evaluate $\Pr(\underline{x} \in S)$, or even to prove that it exists, for certain other sets S we shall be interested in, such as the set of pairs of coprime integers, or the set of square-free integers. In these cases, we will assume that $\Pr(\underline{x} \in S)$ both exists and behaves reasonably, and then use these assumptions to calculate $\Pr(\underline{x} \in S)$. These assumptions can, with care and patience, be justified, but for the reasons given in §1 we will not do so here. It follows that our arguments are only heuristic, with no claim to rigour; for this, the serious reader is encouraged to read such excellent textbooks as [4], [12], or [14].

3. The Coincidence

Suppose that x and y are chosen randomly from $\mathbb{N}$ as in §2. Suppose also that they are chosen independently, that is,

$$\Pr(x \in S \quad \text{and} \quad y \in T) = \Pr(x \in S) \cdot \Pr(y \in T)$$

for all subsets $S, T \subseteq \mathbb{N}$ for which these probabilities are defined. We let

$$P = \Pr((x, y) = 1)$$

denote the probability that x and y are coprime. If $\mathrm{Sq}(x)$ denotes the largest square factor of x, then

$$Q = \Pr(\mathrm{Sq}(x) = 1)$$

is the probability that x is square-free.

We shall now present three "proofs" that

$$P = Q, \tag{3.1}$$

each illustrating a basic property of the Riemann zeta-function.

a) For each $n \in \mathbb{N}$, we have

$$(x, y) = n \quad \text{if and only if} \begin{cases} x \equiv 0 \bmod n, \\ y \equiv 0 \bmod n, \quad \text{and} \\ \left(\dfrac{x}{n}, \dfrac{y}{n}\right) = 1. \end{cases}$$

Now $x \equiv y \equiv 0 \bmod n$ with probability $1/n^2$, and if these congruences are satisfied then x/n and y/n are coprime with probability P, so

$$\Pr((x, y) = n) = \frac{P}{n^2}.$$

The sum of these probabilities, for all $n \in \mathbb{N}$, is equal to 1, so

$$P = 1\Big/\sum_{n \geq 1} n^{-2} = 1/\zeta(2), \tag{3.2}$$

where

$$\zeta(s) = \sum_{n \geq 1} n^{-s}$$

is the *Riemann zeta-function* (see **[10]** or **[17]**).

A similar argument can be used to calculate Q. We have

$$\operatorname{Sq}(x) = n^2 \quad \text{if and only if} \begin{cases} x \equiv 0 \bmod n^2, & \text{and} \\ x/n^2 \text{ is square-free.} \end{cases}$$

Now $x \equiv 0 \bmod n^2$ with probability $1/n^2$ in which case x/n^2 is square-free with conditional probability Q, so

$$\Pr(\operatorname{Sq}(x) = n^2) = \frac{Q}{n^2}$$

and hence

$$Q = 1\Big/\sum_{n \geq 1} n^{-2} = 1/\zeta(2). \tag{3.3}$$

Combining (3.2) and (3.3) we have

$$P = 1/\zeta(2) = Q. \tag{3.4}$$

(b) Cremona **[9]** has suggested a second proof of (3.1), based on the Euler product for $\zeta(s)$ (see §4). We have

$$(x, y) = 1 \quad \text{if, and only if,} \left\{ \begin{matrix} x \not\equiv 0 \bmod p \\ \text{or} \\ y \not\equiv 0 \bmod p \end{matrix} \right\} \text{for every prime } p.$$

For each p we have

$$\Pr(x \equiv 0 \bmod p \text{ and } y \equiv 0 \bmod p) = \Pr(x \equiv 0 \bmod p) \cdot \Pr(y \equiv 0 \bmod p) = p^{-2},$$

and hence

$$\Pr(x \not\equiv 0 \bmod p \text{ or } y \not\equiv 0 \bmod p) = 1 - p^{-2}.$$

By the comments about statistical independence in §2 we can multiply these probabilities for distinct primes p, so (taking care with the limits) we have

$$P = \prod_p (1 - p^{-2}) \tag{3.5}$$

where the product is over all primes $p \in \mathbb{N}$. Similarly, $\operatorname{Sq}(x) = 1$ if and only if

$x \not\equiv 0 \bmod p^2$ for every prime p, so

$$Q = \prod_p (1 - p^{-2}); \tag{3.6}$$

combining this with (3.5), we have

$$P = \prod_p (1 - p^{-2}) = Q. \tag{3.7}$$

c) A third proof of (3.1) is based on the Inclusion-Exclusion Principle (see **[12]**, Theorem 260). We have

$$(x, y) > 1 \quad \text{if, and only if,} \begin{Bmatrix} x \equiv 0 \bmod p \\ \text{and} \\ y \equiv 0 \bmod p \end{Bmatrix} \text{for some prime } p;$$

for each p this joint event has probability p^{-2}, so we obtain a contribution $S_1 = \Sigma_p p^{-2}$ to the probability $1 - P$ of the left-hand side. From this we must *subtract* a contribution

$$S_2 = \sum\sum_{p<q} (pq)^{-2}$$

to compensate for the double counting in S_1 of cases where (x, y) is divisible by distinct primes p and q. We must then *add* a contribution

$$S_3 = \sum\sum\sum_{p<q<r} (pqr)^{-2}$$

to allow for overcompensation in S_2, etc. Thus

$$1 - P = S_1 - S_2 + S_3 - \cdots \tag{3.8}$$

and so

$$P = 1 - S_1 + S_2 - S_3 + \cdots, \tag{3.9}$$

where

$$S_k = \sum \cdots \sum_{p_1 < \cdots < p_k} (p_1 \ldots p_k)^{-2} \tag{3.10}$$

for all $k \geq 1$, with each p_i ranging over the primes. (It is a useful exercise to establish that the coefficients in (3.8) really are all ± 1, as indicated.)

Similarly, $\mathrm{Sq}(x) > 1$ if, and only if, $x \equiv 0 \bmod p^2$ for some prime p; thus the same type of argument yields

$$Q = 1 - S_1 + S_2 - S_3 + \cdots, \tag{3.11}$$

giving (3.1). Indeed, if we define the *Möbius function* μ by

$$\mu(n) = \begin{cases} (-1)^k & \text{if } n \text{ is a product of } k \text{ distinct primes } (k \geq 0) \\ 0 & \text{if } p^2 \text{ divides } n \text{ for some prime } p, \end{cases}$$

then (3.9), (3.10) and (3.11) give

$$P = \sum_{n \geq 1} \frac{\mu(n)}{n^2} = Q. \tag{3.12}$$

These three "proofs" that $P=Q$ all depend on showing that both P and Q are equal to $1/\zeta(2)$, or to some equivalent expression; it would be interesting to have a direct proof, which did not involve evaluating both probabilities but instead proceeded by matching up the events to which they correspond.

Here we will make a slight digression to bring in two of the most important conjectures in modern number theory. If x is square-free then x is a product of k distinct primes, and one might expect k to be even or odd with equal conditional probabilities. This statement, which can be put more precisely in the form

$$\lim_{N\to\infty} \frac{1}{N} \sum_{x=1}^{N} \mu(x) = 0, \tag{3.13}$$

is indeed true, being a corollary of the Prime Number Theorem of Hadamard and de la Vallée Poussin [**4**, §4.9; **12**, §§18.6 and 22.17].

It is useful to know the rate of convergence in (3.13), and in 1897 Mertens conjectured that

$$\left| \sum_{x=1}^{N} \mu(x) \right| \le N^{1/2}$$

for all $N \in \mathbb{N}$. Despite impressive supporting evidence, this was eventually disproved by Odlyzko and te Riele in 1985 [**15**].

However, the slightly weaker *Riemann Hypothesis* is still open. The series $\zeta(s) = \sum n^{-s}$ converges absolutely for all $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 1$, and can be continued analytically to the whole of $\mathbb{C}\setminus\{1\}$ (there is a simple pole at $s=1$, where the harmonic series diverges). This extended function $\zeta(s)$ is known to have zeros at the points $s = -2k$ $(k \in \mathbb{N})$, and in 1859 Riemann conjectured that all other zeros lie on the line $\mathrm{Re}(s) = \frac{1}{2}$. If true, this would have many important consequences in number theory; in particular, it is equivalent to the assertion that if $\varepsilon > 0$ then

$$\left| \sum_{x=1}^{N} \mu(x) \right| \le N^{\frac{1}{2}+\varepsilon}$$

for all sufficiently large N [**4**, Ch. 13, Ex. 4; **17**, p. 315].

4. Generalizations

One can easily extend the arguments in §3 to show that, for each integer $s \ge 2$, the probability P_s that s randomly and independently chosen integers have highest common factor 1 is equal to the probability Q_s that a single randomly chosen integer is s-th power free (divisible by no s-th power greater than 1). Indeed, both probabilities are equal to $1/\zeta(s)$, $\prod(1-p^{-s})$ and $\sum\mu(n)/n^s$ by the arguments in (a), (b), and (c), so we have "proved" the identities

$$\frac{1}{\zeta(s)} = \prod_p (1-p^{-s}) = \sum_{n\ge 1} \frac{\mu(n)}{n^s} \tag{4.1}$$

for all integers $s \ge 2$. (The equation $\zeta(s) = \prod_p (1-p^{-s})^{-1}$, implied by (4.1), is called the *Euler product* representation of $\zeta(s)$.)

A further generalization is due to Benkoski [**6**], who has shown that the highest common factor of the s integers is r-th power free with probability $1/\zeta(rs)$; the result $P_s = Q_s$ is a special case of this.

Of course, (4.1) is valid (but without any probabilistic interpretation) for all $s \in \mathbb{C}$ for which the relevant series and products converge absolutely, that is, for $\mathrm{Re}(s) > 1$. If one postpones details of convergence, then one can justify the second equation in (4.1) by multiplying out the factors $1 - p^{-s}$, and noting that the general term in the expansion is

$$(-1)^k (p_1 p_2 \dots p_k)^{-s} = \frac{\mu(n)}{n^s},$$

where $n = p_1 p_2 \dots p_k$ is a product of distinct primes. Similarly, one can justify the first equation by multiplying out the inverse factors

$$(1 - p^{-s})^{-1} = 1 + p^{-s} + p^{-2s} + \cdots ;$$

the general term is now n^{-s}, with each $n \in \mathbb{N}$ appearing exactly once by the Fundamental Theorem of Arithmetic, so the product is equal to $\zeta(s)$.

This leads immediately to Euler's celebrated proof that there are infinitely many primes: If there were only finitely many, then $\prod_p (1 - p^{-s})$ would converge to a finite nonzero limit as $s \to 1+$, contradicting the fact that $\zeta(s) \to +\infty$.

5. Evaluating $\zeta(2)$

Having shown in §3 that

$$P = Q = 1/\zeta(2),$$

we must now demonstrate the well-known result

$$\zeta(2) = \frac{\pi^2}{6} \tag{5.1}$$

(which Hardy and Wright assert without proof, see [**12**], §17.2).

There are many attractive proofs of (5.1). Some, like [**5**], are quite elementary and direct, while others provide interesting applications of such topics as Fourier series [**8**, p. 99] or complex function theory [**1**, p. 190]; Knopp [**14**, pp. 237, 266, 324, 376] gives the reader a splendid choice of proofs. The demonstration we shall outline here is in keeping with the philosophy of §1, in that the calculations are simple though the technical details needed for a rigorous proof have been omitted; these can be found in §§3.8 and 6.4 of [**13**].

We start with the infinite product expansion

$$\sin z = z \prod_{n \neq 0} \left(1 - \frac{z}{n\pi}\right) = z \prod_{n \geq 1} \left(1 - \frac{z^2}{n^2 \pi^2}\right), \tag{5.2}$$

which one can justify by regarding $\sin z$ as a "polynomial of infinite degree" with zeros at $z = n\pi$ ($n \in \mathbb{Z}$); this gives a "factorization"

$$\sin z = cz \prod_{n \neq 0} \left(1 - \frac{z}{n\pi}\right),$$

where the constant c is given by

$$c = \lim_{z \to 0} \frac{\sin z}{z} = 1.$$

We also have the Taylor series expansion

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \cdots, \tag{5.3}$$

and by comparing coefficients of z^3 in (5.2) and (5.3) we get

$$-\sum_{n\geq 1} \frac{1}{n^2\pi^2} = -\frac{1}{3!},$$

which "proves" (5.1).

In view of the remarks in §4 (and for many other reasons) it would be useful to know the value of $\zeta(s)$ for *every* integer $s \geq 2$. Little is known about the case where s is odd: Indeed, only in 1978 did Apéry [**2**] prove that $\zeta(3)$ is irrational (see also [**7**] and [**16**]). However one can, with a little extra work [**3**; 11, §6.5; **14**, p. 237] use (5.2) to show that for all even integers $s = 2k \geq 2$,

$$\zeta(2k) = \frac{(-1)^{k-1}2^{2k-1}\pi^{2k}B_{2k}}{(2k)!},$$

where B_m is the m-th *Bernoulli number*, a rational number defined by the generating function

$$\sum_{m\geq 0} \frac{B_m t^m}{m!} = \frac{t}{e^t - 1},$$

or equivalently by the recurrence relation

$$\sum_{m=0}^{r-1} \binom{r}{m} B_m = \begin{cases} 1 & \text{if } r = 1, \\ 0 & \text{if } r > 1. \end{cases}$$

Thus

$$\zeta(2) = \frac{\pi^2}{6},$$

$$\zeta(4) = \frac{\pi^4}{90},$$

$$\zeta(6) = \frac{\pi^6}{945}, \text{ etc.},$$

giving

$$P_2 = Q_2 = \frac{6}{\pi^2} = 0.6079\ldots,$$

$$P_4 = Q_4 = \frac{90}{\pi^4} = 0.9239\ldots,$$

$$P_6 = Q_6 = \frac{945}{\pi^6} = 0.9829\ldots,$$

and so on. These values illustrate how

$$P_s \to 1 \quad \text{as} \quad s \to \infty,$$

a fact that is easily proved by using the comparison test to show that

$$1 \leq \zeta(s) \leq 1 + \frac{1}{2^s} + \frac{1}{2^s} + \frac{1}{4^s} + \frac{1}{4^s} + \frac{1}{4^s} + \frac{1}{4^s} + \frac{1}{8^s} + \cdots$$
$$= (1 - 2^{1-s})^{-1}$$

for all $s > 1$.

6. Conclusion

I have tried to use this example to show how, by postponing technical details, one can introduce students to a wide range of interesting and closely related topics; for instance, in addition to the ideas presented above, one could use computer-generated random numbers to obtain experimental evidence for (3.1), or one could interpret P_s geometrically as the probability that a randomly chosen lattice-point $\underline{x} \in \mathbb{Z}^s$ is visible from the origin. A good teacher can use this approach to develop students' confidence by presenting them with material within their competence and understanding, while ensuring that they are well-informed about the main problems and achievements of mathematics. In this way, one can capture their interest and motivate them to go on to study the subject in greater depth.

Acknowledgements. I am grateful to a number of my colleagues, in particular John Cremona, Adrian Oldknow and David Singerman, and to the referees, for their helpful and perceptive comments.

REFERENCES

1. L. V. Ahlfors, *Complex Analysis*, 3rd edition, McGraw-Hill, New York, 1979.
2. R. Apéry, Irrationalité de $\zeta(2)$ et $\zeta(3)$, *Astérisque* 61 (*Journées Arithmétique de Luminy*, 1978) (1979), 11–13.
3. T. M. Apostol, Another elementary proof of Euler's formula for $\zeta(2n)$, *Amer. Math. Monthly* 80 (1973), 425–431.
4. T. M. Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag New York, Inc., New York, 1976.
5. T. M. Apostol, A proof that Euler missed: evaluating $\zeta(2)$ the easy way, *Math. Intelligencer* 5 (1983), 59–60.
6. S. J. Benkoski, The probability that k positive integers are relatively r-prime, *J. Number Theory* 8 (1976), 218–223.
7. F. Beukers, A note on the irrationality of $\zeta(2)$ and $\zeta(3)$, *Bull. London Math. Soc.* 11 (1979), 268–272.
8. R. V. Churchill, *Fourier Series and Boundary Value Problems*, 2nd edition, McGraw-Hill Inc., New York, 1963.
9. J. E. Cremona, private communication.
10. H. M. Edwards, *Riemann's Zeta Function*, Academic Press, New York, 1974.
11. R. L. Graham, D. E. Knuth and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, Reading, MA, 1989.
12. G. H. Hardy and E. M. Wright, *Introduction to the Theory of Numbers*, 5th edition, Oxford University Press, Oxford, 1979.
13. G. A. Jones and D. Singerman, *Complex Functions, an Algebraic and Geometric Viewpoint*, Cambridge University Press, Cambridge, 1987.
14. K. Knopp, *Theory and Application of Infinite Series*, 2nd edition, Blackie, London, 1951.
15. A. M. Odlyzko and H. J. J. te Riele, Disproof of the Mertens conjecture, *J. für die Reine und Angew. Math.* 357 (1985), 138–160.
16. A. J. van der Poorten, A proof that Euler missed..., Apéry's proof of the irrationality of $\zeta(3)$, *Math. Intelligencer* 1 (1978/79), 195–203.
17. E. C. Titchmarsh, *Theory of the Riemann Zeta Function*, Clarendon Press, Oxford, 1951.

NOTES

Derangements and Bell Numbers

ROBERT J. CLARKE
MARTA SVED
University of Adelaide
Adelaide, South Australia

1. Introduction We will establish an interesting connection between derangements and Bell numbers. As preliminaries we summarize some well-known properties of these combinatorial numbers.

A permutation of a set of elements ranking from 1 to n is called *derangement* if none of the elements is left at its original place. A well-known example of this is the case of the absent-minded secretary who has to place n letters into n addressed envelopes and puts *each* letter into a wrong envelope. The problem of enumerating the derangements has a history of some 200 years. It was first posed by Pierre Rémond de Montmort (1678–1719) and was called "le problème de rencontres". The number of derangements of a set of n elements is denoted by $D(n)$. Thus $D(0) = 1$, $D(1) = 0$, $D(2) = 1$, $D(3) = 2$, $D(4) = 9$.

Bell numbers, denoted by $B(n)$, count the number of partitions of a set of n distinguishable objects into unordered non-empty subsets. Thus $B(1) = 1$, $B(2) = 2$, $B(3) = 5$, $B(4) = 15$ and by definition $B(0) = 1$. The Bell numbers are closely related to the well-known Stirling numbers of the second kind, $S(n,k)$. In fact, as $S(n,k)$ equals the number of partitions of a set of n distinguishable objects into k unordered nonempty subsets, we have

$$B(n) = \sum_{k=1}^{n} S(n,k), \quad n \geq 1.$$

Though these numbers are named after Eric Temple Bell (1883–1960), the problem of classifying and enumerating the partitions on n objects occupied mathematicians through much earlier periods. Euler made important contributions to the subject. We refer the reader to Biggs [1] and Stein [5] for further historical background, and to Stanley [4] for a modern treatment of Bell numbers, Stirling numbers, and derangements.

It is not difficult to find a closed formula for the number of derangements $D(n)$. It is obtained by the use of the inclusion-exclusion principle. Denote by $R = \{P_1, P_2, \dots, P_r\}$ a set of properties that each of the elements of some set N may or may not possess. Denote by N_i the number of elements in N having property P_i, N_{ij} the number having properties P_i and P_j, and generally by $N_{ij\dots k}$ the number of elements having all the properties $P_i, P_j, \dots, P_k$, (and possibly other properties belonging to R).

Denote by N_0 the number of elements of N that have none of the properties listed in R. Then

$$N_0 = |N| - \sum_i N_i + \sum_{ij} N_{ij} + \cdots + (-1)^r \sum N_{12\dots r} \tag{1}$$

The set to be considered now is the set of all the permutations of $\{1,2,\dots,n\}$. Then $|N| = n!$ Let property P_i be: The element i occupies place i, ($i \in \{1 \cdots n\}$). Then

$$\sum N_i = \binom{n}{1}(n-1)!, \qquad \sum N_{ij} = \binom{n}{2}(n-2)!$$

and generally

$$\sum N_{i_1 i_2 \cdots i_k} = \binom{n}{k}(n-k)!.$$

Thus by (1) we have

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

This result can be written in the form

$$D(n) = n!\left(1 - \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{(-1)^n}{n!}\right). \tag{2}$$

From here some interesting well-known properties of $D(n)$ follow immediately.

(i) $\dfrac{D(n)}{n!}$ converges quickly to e^{-1},

(ii) $D(n) = nD(n-1) + (-1)^n$, and

(iii) $D(n) = (n-1)(D(n-1) + D(n-2))$.

Although a nice combinatorial proof of (iii) can be given quite easily, (ii) is a different matter. We refer the reader to Remmel [2] for a rather difficult combinatorial proof of (ii). The two recursion formulae (ii) and (iii) allow fast evaluation of $D(n)$.

Recognizing that $\binom{n}{k}D(k)$ counts all permutations in which exactly k elements of the set $\{1,2,\dots,n\}$ are displaced, we obtain the relation

$$\sum_{k=0}^{n} \binom{n}{k} D(k) = n!. \tag{3}$$

Bell numbers are more elusive. They can be evaluated successively using an identity with a left-hand side similar to that of (3). The relation is

$$\sum_{k=0}^{n} \binom{n}{k} B(k) = B(n+1). \tag{4}$$

Here we note that $\binom{n}{k}B(n-k)$ counts all the partitions of the set $\{a_1, a_2, \dots, a_{n+1}\}$ where the last element a_{n+1} appears in some subset containing exactly $k+1$ elements. Such a partition is constructed by finding k elements out of the set $\{a_1, a_2, \dots, a_n\}$ sharing the subset with a_{n+1} and partitioning in some way the remaining set, hence the above result. Thus

$$B(n+1) = \sum_{k=0}^{n} \binom{n}{k} B(n-k) = \sum_{k=0}^{n} \binom{n}{n-k} B(k) = \sum_{k=0}^{n} \binom{n}{k} B(k),$$

as stated in (4).

2. An identity connecting the numbers $D(n)$ and $B(n)$ The following problem has been posed by Terry Tao [6].

Prove the identity

$$\sum_{m=0}^{n}\sum_{i=0}^{m}\frac{(-1)^i m^2}{i!(n-m)!}=n^2-2n+2. \tag{5}$$

Using equation (2), this identity can easily be arranged to

$$\frac{1}{n!}\sum_{m=0}^{n} m^2\binom{n}{m}D(m)=n^2-2n+2. \tag{6}$$

However, we note here that the identity is not valid for $n<2$.

We may look at (6) in probabilistic terms. Over the sample space of all the permutations of n elements define the random variable X as the number of displaced elements. Then the probability

$$P(X=m)=\binom{n}{m}\frac{D(m)}{n!}.$$

Thus equation (3) may be interpreted as

$$\sum_{m=0}^{n}\binom{n}{m}\frac{D(m)}{n!}=\sum_{m=0}^{n}P(X=m)=1.$$

Identity (6) states that the second moment of X about the origin is given by the quadratic n^2-2n+2.

It is also easy to show that the first moment of X, that is its expected value, is

$$\sum_{m=0}^{n} m\binom{n}{m}\frac{D(m)}{n!}=n-1.$$

Writing $P_0(n)=1$, $P_1(n)=n-1$ and $P_2(n)=n^2-2n+2$, (assuming the validity of the last one), it can be expected that $P_s(n)$, the s-th moment of X about the origin, is a polynomial in n of degree s.

In fact, assuming that

$$\sum_{m=0}^{n} m^{s-1}\binom{n}{m}D(m)=n!P_{s-1}(n)\quad\text{for } n\geq s-1, \tag{7}$$

where $P_{s-1}(n)$ is a polynomial of degree $s-1$, (established for degrees 0 and 1), we may write

$$\sum_{m=0}^{n} m^{s}\binom{n}{m}D(m)=\sum_{m=0}^{n}(n-(n-m))m^{s-1}\binom{n}{m}D(m).$$

Using

$$(n-m)\binom{n}{m}=n\binom{n-1}{m},$$

we get

$$\sum_{m=0}^{n} m^{s}\binom{n}{m}D(m)=n\cdot n!P_{s-1}(n)-n\sum_{m=0}^{n-1} m^{s-1}\binom{n-1}{m}D(m)$$

$$= n!(nP_{s-1}(n) - P_{s-1}(n-1)) = n!P_s(n),$$

where

$$P_s(n) = nP_{s-1}(n) - P_{s-1}(n-1). \tag{8}$$

The recursion (8) shows that the sth moment of X is indeed a polynomial of degree s in n.

Our aim in this note is to give a representation of the polynomials $P_s(n)$ in terms of Bell numbers:

$$P_s(n) = \sum_{j=0}^{s} (-1)^j \binom{s}{j} n^{s-j} B(j) \quad \text{for } n \geq s,$$

or equivalently, using (7),

$$\sum_{m=0}^{n} m^s \binom{n}{m} D(m) = n! \sum_{r=0}^{s} (-1)^j \binom{s}{j} n^{s-j} B(j) \quad \text{whenever } n \geq s. \tag{9}$$

We will give an *elementary* proof of relation (9). Before doing so, let us note that one can prove (9) algebraically, using generating functions. Namely, one can use (8) and, after a reasonable amount of work, construct generating functions for $P_s(n)$:

$$\sum_{s=0}^{\infty} P_s(n) x^s/s! = e^{nx} \exp(e^{-x} - 1).$$

Comparing these functions with the well-known generating function for the Bell numbers,

$$\sum_{n=0}^{\infty} B(n) x^n/n! = \exp(e^x - 1),$$

one obtains (9).

We note that C. C. Rousseau [3] has also generalised Terry Tao's identity, obtaining an equivalent result to (9) by a somewhat different algebraic route.

3. Combinatorial interpretation of identity (9) Let N be the set of permutations of $\{1, 2, \ldots, n\}$ as before. In addition we define a set of symbols

$$S = \{\sigma_1, \sigma_2, \ldots, \sigma_s\}, \quad \text{where } s \leq n$$

such that *each* symbol of S is associated with *exactly one* element a_i in some given permutation $a_1 a_2 \ldots a_n$ in N, subject to the following *restriction*.

A symbol of S may only be associated with an element a_i if a_i is a displaced element, that is $a_i \neq i$. (Note that more than one symbol of S may be associated with the same element a_i.)

We consider first the left-hand side of (9). Suppose that there are exactly m displaced elements in a certain permutation belonging to N. There are m^s ways in which each of the symbols can be associated to one of the m elements. The factor $\binom{n}{m} D(m)$ gives the number of permutations with m displaced elements. *Hence the sum from* $m = 0$ *to* n *counts all possible associations of the symbols* S *with the permutations belonging to* N *under the stated restriction.*

Next we use the inclusion-exclusion principle to show that the right-hand side of (9) gives the same count.

Disregarding first the restriction, there are $A = n^s n!$ ways in which the symbols belonging to S can be associated with the permutations belonging to N. However, by the convention adopted, an *error* occurs whenever at least one of the symbols is attached to an element a_i where $a_i = i$. Denote by A_j the sum of the numbers of associations, in which j *specified errors occur*, and accordingly by A_0 the number of "legitimate" associations with no errors. Then by (1)

$$A_0 = A - A_1 + A_2 + \cdots + (-1)^j A_j + \cdots + (-1)^s A_s. \qquad (10)$$

To determine A_j assume first that (at least) j symbols are attached to l elements that are not displaced, where

$$1 \le l \le j \quad (\text{we assume here that } l \le n).$$

There are $\binom{n}{l}$ ways of choosing these fixed elements. The number of ways in which j specified symbols can be divided between l ordered places is

$$l!S(j,l)$$

where $S(j, l)$ is the Stirling number of the second kind. The remaining $s - j$ symbols can be allocated freely to any place in the permutation considered, hence in n^{s-j} ways. Since there are $(n - l)!$ permutations where at least l elements are fixed, the number of illegal associations in which *at least* each of the specified j symbols is consigned to exactly l fixed elements is

$$n^{s-j} l! S(j,l) \binom{n}{l} (n-l)! = n! S(j,l) n^{s-j}. \qquad (11)$$

Summing for $l = 1$ to s and for all possible choices of j symbols out of the set S we obtain

$$A_j = \binom{s}{j} \sum_{l=1}^{s} n! n^{s-j} S(j,l) = n! \binom{s}{j} n^{s-j} B(j),$$

since

$$\sum_{l=1}^{s} S(j,l) = B(j).$$

(We note here that if $s > n$, then at least one of the l values in the range $\{1, 2, \ldots, s\}$ is greater than n, and so we cannot arrive at the left-hand side of (11) and our reasoning breaks down.)

Substituting into (1) we obtain the right hand side of (9). This completes the proof. It is easy to check that the identity is valid only for $n \ge s$.

REFERENCES

1. N. L. Biggs, The roots of combinatorics, *Historia Math.* 6 (1979), 109–136.
2. J. B. Remmel, A note on a recursion for the number of derangements, *European J. Combinatorics* 4 (1983), 371–374.
3. C. C. Rousseau, Binomial identity 31, *James Cook Mathematical Notes* 5 (1990), 5252–5253.
4. Richard P. Stanley, *Enumerative Combinatorics Volume I*, Wadsworth & Brooks/Coke, Monterey, CA, 1986.
5. Paul R. Stein, A brief history of enumeration, *Advances in Mathematics Supplementary Studies* 10 (1986), 169–206.
6. Terry Tao, Binomial identity 31, *James Cook Mathematical Notes* 5 (1990), 5227.

An Old Algorithm for the Sums of Integer Powers

DAVID M. BLOOM
Brooklyn College of CUNY
Brooklyn, NY 11210

For integers $k \geqslant 0$, let $P_k(x)$ be the polynomial such that

$$P_k(n) = 1^k + 2^k + \cdots + n^k \quad (\text{all integers } n \geqslant 0).$$

A recent article [1] in this MAGAZINE gives a new proof of a well-known method for finding P_k if $P_0, \ldots, P_{k-1}$ are known, and then goes on to show that in fact the work can be cut in half: P_k can be expressed in terms of $P_{k-2}, P_{k-4}, \ldots$. Thus, for instance, to compute P_{20} we need only 10 of the preceding polynomials, not all 20. The author obtains even greater efficiency when k is odd.

We exhibit here a simple algorithm to find P_k if only the *single* polynomial P_{k-1} is known. The method is mentioned very briefly in [3, pp. 229–230] and is essentially a one-step deduction from a standard formula for P_k; but, among other sources of the standard formula that we have looked at ([2], [4, p. 90], [5]), none takes that one extra step. This suggests that the algorithm is not well known; the purpose of this note is to make it known more widely.

The "standard formula" for P_k is

$$P_k(x) = \sum_{j=1}^{k+1} a_j^{(k)} x^j, \quad a_j^{(k)} = \binom{k}{j-1} B_{k+1-j}/j \quad (1 \leqslant j \leqslant k+1), \tag{1}$$

valid for all $k \geqslant 0$, where B_n is the value at 0 for the n-th derivative of the function $f(x) = xe^x/(e^x - 1)$ (with $f(0)$ defined to be 1). The B's are the famous "Bernoulli numbers".* Elementary proofs of (1) appear in [4, p. 90] and [5]. Now for the one-step deduction: Since

$$\binom{k}{j-1}/j = (k/j) \cdot \binom{k-1}{j-2}/(j-1) \quad \text{if } j \geqslant 2,$$

the second of the two equations (1) immediately implies that $a_j^{(k)} = (k/j) a_{j-1}^{(k-1)}$ if $j \geqslant 2$. That is, *to get the coefficient of x^j in $P_k(x)$ when $j \geqslant 2$, just multiply the coefficient of x^{j-1} in $P_{k-1}(x)$ by k/j.*

(Yes, it's that simple!)

Thus all coefficients of $P_k(x)$ except for $a_1^{(k)} = B_k$ (the coefficient of x^1) are determined once P_{k-1} is known; and $a_1^{(k)}$ is then obtained from the equation

$$a_1^{(k)} + a_2^{(k)} + \cdots + a_{k+1}^{(k)} = P_k(1) = 1.$$

For example, given that $P_3(x) = \frac{1}{4}x^2(x+1)^2 = \frac{1}{4}x^4 + \frac{1}{2}x^3 + \frac{1}{4}x^2$, we obtain $P_4(x)$ as follows:

$$\begin{array}{lllll}
P_3(x) = \frac{1}{4}x^4 & + \frac{1}{2}x^3 & + \frac{1}{4}x^2 & + 0x & \\
\big\downarrow \text{ multiply by } \frac{4}{5}x & \big\downarrow \text{ multiply by } \frac{4}{4}x & \big\downarrow \text{ multiply by } \frac{4}{3}x & \big\downarrow \text{ multiply by } \frac{4}{2}x & \\
P_4(x) = \frac{1}{5}x^5 & + \frac{1}{2}x^4 & + \frac{1}{3}x^3 & + 0x^2 & - \frac{1}{30}x,
\end{array}$$

*The traditional definition of the Bernoulli numbers omits the factor e^x from the numerator of $f(x)$. We include it so as not to have to change the sign of B_1 to make formula (1) correct.

where the coefficient $B_4 = -1/30$ is the number needed to make the sum of the coefficients of $P_4(x)$ equal 1. Going from $P_4(x)$ to $P_5(x)$ would be even easier, since $B_5 = 0$; in fact, it is well known that the odd-numbered B's are zero beginning with B_3. Thus, once P_{2n} is known, calculation of P_{2n+1} is virtually instantaneous!

Remark. Since $B_3 = B_5 = B_7 = \cdots = 0$, it follows from (1) that

$$a_j^{(k)} = 0 \quad (j < k,\ j \equiv k \pmod 2). \tag{2}$$

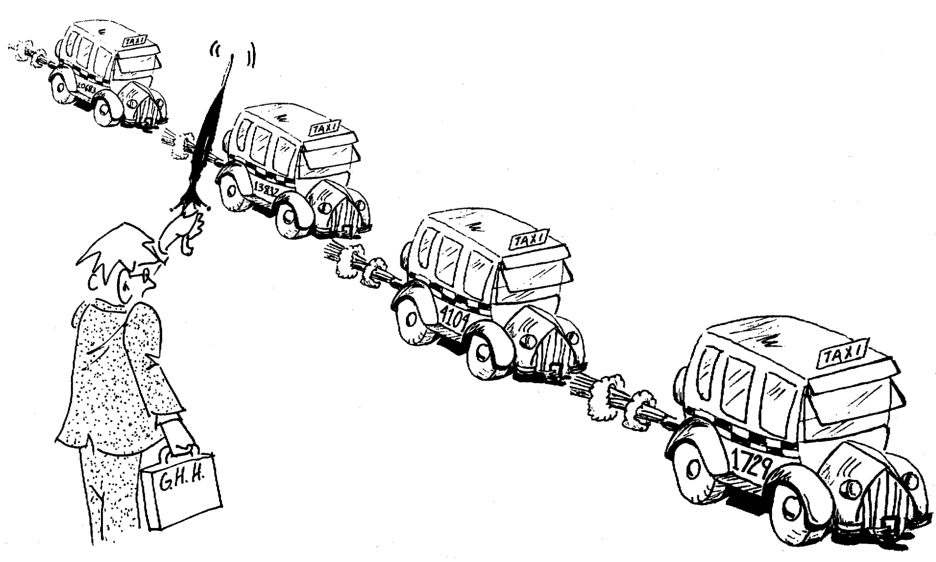
One can give a simple direct proof of (2) without any reference to Bernoulli numbers or to any explicit formula for $a_j^{(k)}$. Here is such a proof when k is even: Write $P = P_k$. Using $P(0) = 0$ and

$$P(x+1) - P(x) = (x+1)^k \tag{3}$$

(which must hold for all real x since it holds whenever $x \in \mathbf{Z}^+$), one shows by induction that $P(-n) = -P(n-1)$ for all $n \in \mathbf{Z}^+$, hence also for all real x in place of n. It follows that $\frac{1}{2}(P(x) + P(x-1))$ is an odd function of x; but the latter equals $P(x) - \frac{1}{2}x^k$ by (3), and the result follows. The proof for odd k is similar.

REFERENCES

1. D. Acu, Some algorithms for the sums of integer powers, this MAGAZINE 61 (1988) 189–191.
2. A. W. F. Edwards, Sums of powers of integers—a little of the history, *Math. Gazette* 66 (1982), 22–28.
3. H. Edwards, *Fermat's Last Theorem: A Genetic Introduction to Algebraic Number Theory*, Springer-Verlag New York, Inc., New York, 1977.
4. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 4th edition, Oxford, London, 1960.
5. J. Nunemacher and R. Young, On the sum of consecutive k-th powers, this MAGAZINE 60 (1987), 237–238.



—TIMOTHY STEPHANY
WILLIAMSON, NY 14589

The Prisoner's Dilemma

ROMAN WONG
Washington & Jefferson College
Washington, PA 15301

The following is an interesting problem in probability: *The prisoner's dilemma*: A prisoner was to be executed and he begged the king for mercy. The king decided to give the prisoner a chance. He gave the prisoner 50 white balls and 50 black balls, all identical in shape. The prisoner was supposed to distribute these balls into two identical bags in any way he liked and then pick one bag at random and draw one ball at random from that bag. His life would be spared if the ball drawn was a white ball. (There are several other problems under the same name. For instance, see **[1]**.)

The question now of course is to decide how the balls should be distributed so that the prisoner has the best (and the worst) chance to live. For a small number of balls, one can easily find the answer by computing the probabilities of all possible distributions. A solution to the above problem can be found in **[2]**. For a large number of balls, a calculus approach is more desirable.

Assume now that there are N white balls and N black balls, $N \geq 1$, to be distributed into two identical bags. Let w and b be the number of white and black balls in one of the bags, $0 \leq w,\ b \leq N$. Consider the domains

$$\mathfrak{D} = \{(w,b) \in \mathbb{Z} \times \mathbb{Z} \mid 0 \leq w,\ b \leq N,\ w+b \neq 0,\ w+b \neq 2N\}$$

and

$$\mathfrak{D}' = \{(w,b) \in \mathbb{R} \times \mathbb{R} \mid 0 \leq w,\ b \leq N,\ w+b \neq 0,\ w+b \neq 2N\}.$$

The probability that a ball drawn at random from one of the bags is a white ball is given by

$$P(w,b) = \frac{1}{2}\left(\frac{w}{w+b} + \frac{N-w}{2N-w-b}\right), \qquad (w,b) \in \mathfrak{D}. \tag{1}$$

We want to find the maximum and minimum of the function $P(w,b)$ in $\mathfrak{D}'$. The partial derivatives of P are rather messy. Nevertheless, it can be shown that

$$\frac{\partial P}{\partial w} = \frac{1}{2}$$
$$\left[\frac{(w+b)(2N-w-b)(N-2w) - (3Nw - 2w^2 - 2wb + Nb)(2N-2w-2b)}{(w+b)^2(2N-w-b)^2}\right]$$

and

$$\frac{\partial P}{\partial b} = \frac{1}{2}$$
$$\left[\frac{(w+b)(2N-w-b)(3N-4w-2b) - (3Nw - 2w^2 - 2wb + Nb)(2N-2w-2b)}{(w+b)^2(2N-w-b)^2}\right].$$

To find critical points of P, we set both partial derivatives to zero. This leads to

$$(w+b)(2N-w-b)(N-2w) = (w+b)(2N-w-b)(3N-4w-2b).$$

Thus,

$$(w+b)(2N-w-b)(2w+2b-2N)=0.$$

Since $w+b \neq 0$ and $w+b \neq 2N$, we must have $w+b=N$. Putting this back into $\partial P/\partial w = 0$, we find that $(w,b)=(N/2,N/2)$ is the only critical point of P in $\mathfrak{D}'$. From (1), we have $P(N/2,N/2)=1/2$. In fact, $P(w,b)|_{w+b=N}=1/2$. We shall see that this value is neither the maximum nor the minimum for P in $\mathfrak{D}'$.

We now consider the value of $P(w,b)$ on the boundary of $\mathfrak{D}'$. First we set $w=0$. Then

$$P(0,b)=\frac{N}{2}(2N-b)^{-1} \quad \text{so that} \quad \frac{\partial}{\partial b}P(0,b)=\left(\frac{N}{2}\right)(2N-b)^{-2},$$

which is always positive for any $(0,b)\in\mathfrak{D}$. Thus $P(w,b)$ is increasing along $w=0$ and so the maximum of $P(0,b)$ along the border $w=0$ is attained at $b=N$ with maximum value

$$P(0,N)=\frac{N}{2(2N-N)}=\frac{1}{2}.$$

By symmetry, $P(N,0)=1/2$ is the maximum of $P(N,b)$ along the border $w=N$.

We now consider the border $b=0$.

$$P(w,0)=\frac{1}{2}\left(1+\frac{N-w}{2N-w}\right)=\frac{1}{2}\left(\frac{3N-2w}{2N-w}\right)$$

and so

$$\frac{\partial P(w,0)}{\partial w}=\frac{1}{2}\left[\frac{(2N-w)(-2)+(3N-2w)}{(2N-w)^2}\right]=\frac{-N}{2(2N-w)^2}<0 \text{ in } \mathfrak{D}'.$$

Thus $P(w,0)$ is decreasing along the border $b=0$ and so the maximum of $P(w,0)$ in $\mathfrak{D}$ is attained at $w=1$ with maximum value

$$P(1,0)=\frac{1}{2}\left(\frac{3N-2}{2N-1}\right). \tag{2}$$

By symmetry, the maximum of P along the border $b=N$ is attained at $w=N-1$ with the same maximum value given by (2). Since $N\geq 1$, $3N-2\geq 2N-1>0$ so that

$$P(N-1,N)=P(1,0)=\frac{1}{2}\left(\frac{3N-2}{2N-1}\right)\geq\frac{1}{2}(1)=\frac{1}{2},$$

with equality only when $N=1$. Thus we have shown that the maximum of $P(w,b)$ is attained at $(1,0)$ and $(N-1,N)$ with maximum given by (2). Also, as a function of N,

$$\frac{dP(1,0)}{dN}=\frac{1}{2(2N-1)^2}>0,$$

so that the prisoner's chance of survival increases when more balls are used. The best chance for his survival is thus equal to

$$\lim_{N\mapsto\infty}\left(\frac{3N-2}{2(2N-1)}\right)=\frac{3}{4}.$$

The above discussions also indicate that the minimum value of $P(w, b)$ is attained on the border at (0, 1) and $(N, N-1)$ with minimum value equal to

$$\frac{N}{2(2N-1)} \le \frac{N}{2(2N-N)} = \frac{1}{2}.$$

Since $N/2(2N-1)$ is a decreasing function in N, the worst chance for the prisoner's survival is $1/4$.

The above argument can be extended to any number of bags used. Let k be any integer with $N \ge k \ge 2$. We want to find the maximum and minimum probabilities of drawing a white ball when N white balls and N black balls are distributed among the k bags. Let w_i and b_i be the number of white and black balls respectively to be distributed in bag i, $1 \le i \le k$. The probability that a white ball is drawn at random from the k bags is a function of $2k$ variables:

$$P = P(w_1, w_2, \dots, w_k; b_1, b_2, \dots, b_k) = \frac{1}{k} \sum_{i=1}^{k} \frac{w_i}{w_i + b_i},$$

where $0 \le w_i,\ b_i \le N$, $w_i + b_i \ne 0$ for each i, and

$$\sum_{i=1}^{k} w_i = N = \sum_{i=1}^{k} b_i.$$

With the result for $k = 2$ above and a simple induction on k, one can see that the maximum value of P is attained at points $(w_1, \dots, w_j, \dots w_k; b_1, \dots, b_j, \dots b_k)$ for each $1 \le j \le k$, where

$$w_j = N - k + 1,\ b_j = N \quad \text{and} \quad w_i = 1 \quad \text{and} \quad b_i = 0 \text{ for each } i \ne j.$$

It follows that the maximum probability is equal to

$$P_{\max} = P(1, 1, \dots, 1, N-k+1; 0, 0, \dots, 0, N) = \frac{1}{k}\left[k - 1 + \frac{N-k+1}{2N-k+1}\right]$$
$$= \frac{2Nk - k^2 + k - N}{2Nk - k^2 + k}.$$

Thus for a fixed k,

$$\frac{dP_{\max}}{dN} = \frac{k^2 - k}{(2Nk - k^2 + k)^2} > 0$$

so that $P_{\max}$ is an increasing function of N. Taking the limit of $P_{\max}$ as N approaches ∞, we see that the limit probability with a large number of balls is $(2k-1)/2k$. As more bags are used, this probability approaches 1 as a limit.

Similarly, the minimum P value is attained at $(0, \dots, 0, N; 1, \dots, 1, N-k+1)$ with minimum probability equal to

$$P_{\min} = P(0, \dots, 0, N; 1, \dots, 1, N-k+1) = \frac{N}{k(2N-k+1)} \le \frac{N}{k(2N-N)} = \frac{1}{2k}.$$

For a fixed k, this minimum probability approaches $1/2k$ when N is large, and approaches 0 when many bags are used. Thus, for example if $k = 5$ and $N = 10$, we see that we should fill each of the four bags with exactly one white ball and no black balls. The remaining bag should contain six white and 10 black balls. With $k = 5$, increasing N implies that $P_{\max} \to 0.90$.

Remark. The above discussion shows a good application of calculus in probability. When $k = 2$, the above result can also be established without using calculus. Denote $w + b$ by t, the number of balls in the first bag. Now (1) can be written as

$$P(w,b) = \frac{1}{2}\left(\frac{w}{t} + \frac{N-w}{2N-t}\right) = \frac{1}{2}\left(\frac{N}{2N-t}\right) + \frac{1}{2}w\left(\frac{1}{t} - \frac{1}{2N-t}\right).$$

The expression in the last set of parentheses is positive when $0 < t < N$ and is 0 if $t = N$. The maximum of $P(w,b)$ thus occurs when $w = t$ with value $P_t = P(t,b) = (1/2)(2 - N/(2N-t))$. The largest value for P_t is attained at $t = 1$ with value

$$\frac{1}{2}\left(2 - \frac{N}{2N-1}\right) = \frac{1}{2}\left(\frac{3N-2}{2N-1}\right).$$

The author is grateful to the referee for pointing out this method.

REFERENCES

1. Solomon Garfunkel, *For All Practical Purposes*, W. H. Freeman and Company, New York, 1988, p. 220–224.
2. Fred Schuh, *The Master Book of Mathematical Recreations*, Dover Publications, Mineola, NY, 1968, p. 180.

An Unconventional Orthonormal Basis Provides an Unexpected Counterexample

GIUSEPPE MARINO
PAOLAMARIA PIETRAMALA
Università della Calabria
Cosenza, Italy

Any good textbook about linear operators on real Hilbert spaces begins with the basic geometry of Hilbert spaces. In this context one introduces orthonormal bases and gives the classical examples of such bases in the spaces $L^2[a,b]$, with $[a,b]$ a compact interval of R: In this way the reader becomes familiar with the Haar basis and, via the Gram-Schmidt orthogonalization procedure, the Legendre, Hermite, and Laguerre polynomials.

We note that the usual examples of orthonormal bases in $L^2[a,b]$ are such that either *all* the functions in the basis are *continuous* functions or *all* the functions in the basis are *discontinuous* functions. So, the following problem arises:

QUESTION 1. *Does there exist an orthonormal basis in $L^2[a,b]$ containing only one discontinuous function?*

Here discontinuous means that such a function is not equal almost everywhere to another continuous function. In Proposition 1 we show that the answer to this question is "yes".

Remark. The above discussion shows a good application of calculus in probability. When $k=2$, the above result can also be established without using calculus. Denote $w+b$ by t, the number of balls in the first bag. Now (1) can be written as

$$P(w,b)=\frac{1}{2}\left(\frac{w}{t}+\frac{N-w}{2N-t}\right)=\frac{1}{2}\left(\frac{N}{2N-t}\right)+\frac{1}{2}w\left(\frac{1}{t}-\frac{1}{2N-t}\right).$$

The expression in the last set of parentheses is positive when $0<t<N$ and is 0 if $t=N$. The maximum of $P(w,b)$ thus occurs when $w=t$ with value $P_t=P(t,b)=(1/2)(2-N/(2N-t))$. The largest value for P_t is attained at $t=1$ with value

$$\frac{1}{2}\left(2-\frac{N}{2N-1}\right)=\frac{1}{2}\left(\frac{3N-2}{2N-1}\right).$$

The author is grateful to the referee for pointing out this method.

REFERENCES

1. Solomon Garfunkel, *For All Practical Purposes*, W. H. Freeman and Company, New York, 1988, p. 220–224.
2. Fred Schuh, *The Master Book of Mathematical Recreations*, Dover Publications, Mineola, NY, 1968, p. 180.

An Unconventional Orthonormal Basis Provides an Unexpected Counterexample

GIUSEPPE MARINO
PAOLAMARIA PIETRAMALA
Università della Calabria
Cosenza, Italy

Any good textbook about linear operators on real Hilbert spaces begins with the basic geometry of Hilbert spaces. In this context one introduces orthonormal bases and gives the classical examples of such bases in the spaces $L^2[a,b]$, with $[a,b]$ a compact interval of R: In this way the reader becomes familiar with the Haar basis and, via the Gram-Schmidt orthogonalization procedure, the Legendre, Hermite, and Laguerre polynomials.

We note that the usual examples of orthonormal bases in $L^2[a,b]$ are such that either *all* the functions in the basis are *continuous* functions or *all* the functions in the basis are *discontinuous* functions. So, the following problem arises:

QUESTION 1. *Does there exist an orthonormal basis in $L^2[a,b]$ containing only one discontinuous function?*

Here discontinuous means that such a function is not equal almost everywhere to another continuous function. In Proposition 1 we show that the answer to this question is "yes".

In the reading of our textbook one gets to the theory of self-adjoint operators. The requirement that an operator be self-adjoint is very restrictive. Nevertheless, a wide variety of problems in mathematics and physics give rise to self-adjoint operators. An important type of compact self-adjoint operator T on $L^2[a,b]$ is induced by a continuous function K on $[a,b]\times[a,b]$ via the formula

$$(Tf)(x)=\int_a^b K(x,y)f(y)\,dy.$$

Some time ago, in the setting of the Hilbertian formulation of quantum mechanics, one of the authors encountered the following problem:

QUESTION 2. *Let S be a real inner product space and let H be the real Hilbert space completion of S. Let $T\colon S\to S$ be a compact self-adjoint operator. Let $\overline{T}\colon H\to H$ be the continuous extension of T to H. If T is injective, does it follow that $\overline{T}$ must also be injective?*

At first, this question seems to have no relation to Question 1, but we will show in Proposition 2 that the answer to Question 2 is "no" by using the affirmative answer to Question 1.

PROPOSITION 1. *The answer to Question 1 is yes.*

Proof. We prove the assertion for $L^2[-1,1]$. The proof for $L^2[a,b]$ is analogous. Let $\phi_0\in L^2[-1,1]$ be the discontinuous function defined by

$$\phi_0(x)=\begin{cases}1 & x\geqslant 0\\ 0 & x<0\end{cases}\tag{1}$$

Let $\mathscr{A}=\phi_0^{\perp}$. Clearly, $\mathscr{A}=\{f\in L^2[-1,1]\colon \int_0^1 f(x)\,dx=0\}$. Let $\mathscr{B}=\mathscr{A}\cap\mathscr{C}[-1,1]$. Since $\mathscr{B}$ is a separable inner-product space, it has an orthonormal basis $\{\phi_n\}_{n\geqslant 1}$. So, it is enough to show that $\mathscr{B}$ is dense in $\mathscr{A}$ to conclude that $\{\phi_0\}\cup\{\phi_n\}_{n\geqslant 1}$ is an orthonormal basis of $L^2[-1,1]$. Thus we prove that $\mathscr{B}$ is dense in $\mathscr{A}$, i.e. that for any $f\in\mathscr{A}$ and for any $\varepsilon>0$, there exists $h\in\mathscr{B}$ such that $\|f-h\|_2<\varepsilon$ (here, as usual, $\|\cdot\|_2$ denotes the norm in $L^2[-1,1]$).

Let $\psi\in L^2[-1,1]$ be the continuous function defined by $\psi(x):=(\pi/2)\sin(\pi x)\phi_0(x)$. We note that $\int_0^1\psi(x)\,dx=1$. Given $f\in\mathscr{A}$ and $\varepsilon>0$, we know that there exists $g\in\mathscr{C}[-1,1]$ such that

$$\|f-g\|_2<\frac{\varepsilon}{1+2\|\psi\|_2}.$$

Let $\lambda=\int_0^1 g(x)\,dx$. Then

$$|\lambda|=\left|\int_0^1(f(x)-g(x))\,dx\right|\leqslant\int_{-1}^1|f(x)-g(x)|\,dx$$
$$\leqslant 2\|f-g\|_2<\frac{2\varepsilon}{1+2\|\psi\|_2}.$$

If h is the function defined by $h=g-\lambda\psi$, then h is in $\mathscr{B}$ and the following inequality holds:

$$\|f-h\|_2\leqslant\|f-g\|_2+|\lambda|\|\psi\|_2<\frac{\varepsilon}{1+2\|\psi\|_2}+\frac{2\varepsilon\|\psi\|_2}{1+2\|\psi\|_2}=\varepsilon.$$

PROPOSITION 2. *The answer to Question 2 is no.*

Proof. It is enough to give a counterexample. To accomplish this, let $S = \mathscr{C}[-1,1]$ and $H = L^2[-1,1]$. Moreover, let ϕ_0 be the function defined in (1) and let $\{\phi_n\}_{n\geqslant 1}$ be a sequence of orthonormal continuous functions defined in such a way that $\{\phi_0\} \cup \{\phi_n\}_{n\geqslant 1}$ is an orthonormal basis of $L^2[-1,1]$. (The existence of such a sequence $\{\phi_n\}_{n\geqslant 1}$ is assured by Proposition 1.) We note that from $\|\phi_n\|_2 = 1$ it follows that

$$M_n = \max\{|\phi_n(x)| : x \in [-1,1]\} \geqslant \frac{1}{\sqrt{2}}.$$

Hence we see that the series

$$\sum_1^\infty \frac{1}{n^2 M_n^2} \phi_n(x)\phi_n(y)$$

converges absolutely and uniformly on $[-1,1]\times[-1,1]$. This implies that it defines a continuous function on $[-1,1]\times[-1,1]$ that we denote by K:

$$K(x,y) = \sum_1^\infty \frac{1}{n^2 M_n^2} \phi_n(x)\phi_n(y).$$

Thus the function K induces the integral operator T defined by

$$(Tf)(x) = \int_{-1}^1 K(x,y)f(y)\,dy.$$

This integral operator is compact, self-adjoint and maps $\mathscr{C}[-1,1]$ into $\mathscr{C}[-1,1]$ (actually T maps the entire space $L^2[-1,1]$ into $\mathscr{C}[-1,1]$; see Chap. IV of [1]).

We show now that T: $\mathscr{C}[-1,1] \to \mathscr{C}[-1,1]$ is injective. Indeed, let $f \in \mathscr{C}[-1,1]$ be such that $Tf = 0$, i.e. for any $x \in [-1,1]$,

$$\begin{aligned} 0 = (Tf)(x) &= \int_{-1}^1 K(x,y)f(y)\,dy \\ &= \sum_1^\infty \frac{1}{n^2 M_n^2}\phi_n(x)\int_{-1}^1 \phi_n(y)f(y)\,dy \\ &= \sum_1^\infty \frac{1}{n^2 M_n^2} <f, \phi_n> \phi_n(x) \end{aligned} \tag{2}$$

where we have denoted by $\langle , \rangle$ the inner product on $L^2[-1,1]$.

From (2) it follows that

$$\left\| \sum_1^\infty \frac{1}{n^2 M_n^2} <f, \phi_n> \phi_n \right\|_2 = 0.$$

This implies that $\sum_1^\infty (1/n^4 M_n^4)\langle \phi_n, f\rangle^2 = 0$ and so $\langle \phi_n, f\rangle = 0$, for all $n \geqslant 1$. Since $\{\phi_0\} \cup \{\phi_n\}_{n\geqslant 1}$ is an orthonormal basis of $L^2[-1,1]$, one obtains that f is a multiple of ϕ_0, and the unique *continuous* function multiple of ϕ_0 is the 0 function. Summarizing, T is injective on $\mathscr{C}[-1,1]$, but since $T\phi_0 = 0$, we conclude that T is not injective on $L^2[-1,1]$.

Acknowledgement. The authors wish to thank L. Colzani for useful comments and suggestions.

REFERENCE

1. F. Gohberg and S. Goldberg, *Basic Operator Theory*, Birkhäuser, Boston, 1981.

Conversion of Sequences from Deductive to Inductive Form

DAVID F. SKRBINA
University of Michigan-Dearborn
Dearborn, MI 48128-1491

There are two general ways of expressing a sequence of numbers: (1) the *deductive* form, which specifies a function relating the elements of the sequence to the natural numbers, and (2) the *inductive* form, in which each term of the sequence is specified as a function of the preceding term.

It is apparent that many sequences can be expressed in *both* deductive and inductive forms. A question of interest is: Given the deductive form, is there an algorithm that will generate the inductive equivalent? Consider Diagram 1.

$$\begin{array}{ccccl}
n & \xrightarrow{f(n)} & a_n & & \\
\hline\hline
1 & & a_1 & \big\downarrow & g(a_1) \\
2 & & a_2 & & \\
\vdots & & \vdots & & \\
n & \xleftarrow{f^{-1}} & a_n & \big\downarrow & g(a_n) \\
n+1 & \xrightarrow{f} & a_{n+1} & & \\
\vdots & & \vdots & &
\end{array}$$

Diagram 1.

We can see that "f" maps n to a_n, and "g" maps a_n to a_{n+1}. However, we can also accomplish the mapping of "g" by taking $f^{-1}(a_n)$ back to "n", adding 1 to get $n+1$, then applying "f". Thus we have

$$a_{n+1} = f\big(f^{-1}(a_n) + 1\big), \tag{1}$$

or equivalently,

$$g(a_n) = f\big(f^{-1}(a_n) + 1\big).$$

So under the assumption that f^{-1} exists, we can generate "g" from "f":

Example. Consider the deductive sequence

$$S_1 = 1, 4, 9, 16, \ldots, \text{defined by } a_n = n^2.$$

The "intuitive" method of converting this to an inductive sequence (using both a_n and n) would result in

$$a_1 = 1, \quad \text{and} \quad a_{n+1} = a_n + (2n+1) \quad \text{for} \quad n \geq 1.$$

However, using (1) we get

$$a_{n+1} = \big((a_n)^{1/2} + 1\big)^2$$

$$\text{or} \quad a_1 = 1, \quad \text{and} \quad a_{n+1} = a_n + 2a_n^{1/2} + 1 \quad \text{for} \quad n \geq 1. \tag{2}$$

Example. For the sequence

$$S_2 = e^{-1}, e^{-2}, e^{-3}, \dots, \quad \text{defined by } a_n = e^{-n}$$

the algorithm produces

$$a_{n+1} = \exp[-(-\ln(a_n) + 1)]$$

$$\text{or} \quad a_1 = e^{-1}, \quad \text{and} \quad a_{n+1} = a_n \cdot e^{-1} \quad \text{for} \quad n \geq 1. \tag{3}$$

Example. As a third type of sequence, consider

$$S_3 = 1/2, 2/3, 3/4, \dots, \quad \text{defined by } a_n = n/(n+1).$$

Here, $f^{-1}(n) = n/(1-n)$, so $a_1 = 1/2$ and

$$\begin{aligned} a_{n+1} &= (a_n/(1-a_n) + 1)/[(a_n/(1-a_n) + 1) + 1] \\ &= 1/(2-a_n) \quad \text{for} \quad n \geq 1, \end{aligned} \tag{4}$$

which again is not an obvious solution.

In fact, we can solve all three of the above examples in the general sense. Using equation (1) we obtain the inductive equivalent of each of three classes of deductive sequences (see Table 1).

TABLE 1

Deductive	Inductive	
$a_n = n^p$	$a_{n+1} = [a_n^{1/p} + 1]^p$	(5)
$a_n = r^n$	$a_{n+1} = r^{[(\ln a_n)/(\ln r) + 1]}$	(6)
$a_n = \dfrac{An + B}{Cn + D}$	$a_{n+1} = \dfrac{A\Gamma + A + B}{C\Gamma + C + D}$, where $\Gamma = \dfrac{B - D(a_n)}{C(a_n) - A}$	(7)

It can readily be seen that the examples (2), (3), and (4) are special cases of the general solutions (5), (6), and (7), respectively. Thus for certain classes of deductive sequences (namely, ones in which f^{-1} exists), it is possible to explicitly generate the inductive equivalent.

Converting inductive sequences $a_{n+1} = g(a_n)$ to deductive form amounts to solving first order difference equations. Elementary examples can be found in any discrete mathematics textbook.

The Square Roots of 2 × 2 Matrices

DONALD SULLIVAN
University of New Brunswick
Fredericton, N.B., Canada

Introduction In a recent article MacKinnon [1] describes four methods that may be used to find square roots of 2×2 matrices. The first of these methods requires that the matrix for which the square roots are sought be diagonalizable and, subsequently, this method was used by Scott [2] to determine all the square roots of 2×2 matrices. A surprising conclusion is that scalar 2×2 matrices possess double-infinities of square roots whereas nonscalar 2×2 matrices have only a finite number of square roots.

The purpose of this article is to show how the Cayley-Hamilton theorem may be used to determine explicit formulae for all the square roots of 2×2 matrices. These formulae indicate exactly when a 2×2 matrix has square roots, and the number of such roots.

By definition, the square roots of a 2×2 matrix, A, are those 2×2 matrices, X, for which

$$X^2 = A. \tag{1}$$

However, for each square matrix X, the Cayley-Hamilton theorem states that

$$X^2 - (\operatorname{tr} X)X + (\det X)I = 0. \tag{2}$$

Thus, if a 2×2 matrix A has a square root X, then we may use (2) to eliminate X^2 from (1) to obtain

$$(\operatorname{tr} X)X = A + (\det X)I.$$

Further, since $(\det X)^2 = \det X^2 = \det A$, then $\det X = \varepsilon_1\sqrt{\det A}$, that is $\det\sqrt{A} = \varepsilon_1\sqrt{\det A}$, so that the above result simplifies to the identity:

$$(\operatorname{tr} X)X = A + \varepsilon_1\sqrt{\det A}\, I, \quad \varepsilon_1 = \pm 1. \tag{3}$$

Case 1: A is a scalar matrix. If A is a scalar matrix, $A = aI$, then (3) gives

$$(\operatorname{tr} X)X = (1+\varepsilon_1)aI, \quad \varepsilon_1 = \pm 1.$$

Hence, either $(\operatorname{tr} X)X = 0$ or $(\operatorname{tr} X)X = 2aI$. The first of these possibilities determines the general solution of (1) as

$$X = \begin{pmatrix} \alpha & \beta \\ \gamma & -\alpha \end{pmatrix}, \quad \alpha^2 + \beta\gamma = a \tag{4a}$$

and it covers the second possibility if $a = 0$. On the other hand, if $a \neq 0$ then the second possibility, $(\operatorname{tr} X)X = 2aI$, implies X is scalar and has only the pair of solutions

$$X = \pm\sqrt{a}\, I. \tag{4b}$$

For this case we conclude that if A is a zero matrix then it has a double infinity of square roots as given by (4a) with $a = 0$, whereas if A is a nonzero, scalar matrix then it has a double-infinity of square roots plus two scalar square roots as given by (4a) and (4b).

Case 2: A is not a scalar matrix. If A is not a scalar matrix then $\operatorname{tr} X \neq 0$ in (3). Consequently, every square root X has the form:

$$X = \tau^{-1}\left(A + \varepsilon_1 \sqrt{\det A}\, I\right), \quad \tau \neq 0$$

Substituting this expression for X into (1) and using the Cayley-Hamilton theorem for A we find

$$A^2 + \left(2\varepsilon_1\sqrt{\det A} - \tau^2\right)A + (\det A)I = 0$$

$$((\operatorname{tr} A)A - (\det A)I) + \left(2\varepsilon_1\sqrt{\det A} - \tau^2\right)A + (\det A)I = 0$$

$$\left(\operatorname{tr} A + 2\varepsilon_1\sqrt{\det A} - \tau^2\right)A = 0.$$

Since A is not a scalar matrix then A is not a zero matrix, so

$$\tau^2 = \operatorname{tr} A + 2\varepsilon_1\sqrt{\det A}, \quad (\tau \neq 0,\ \varepsilon_1 = \pm 1). \tag{5}$$

If $(\operatorname{tr} A)^2 \neq 4 \det A$ then both values of ε_1 may be used in (5) without reducing τ to zero. Consequently, it follows from (3) that we may write X, the square root of A, as

$$X = \varepsilon_2 \frac{A + \varepsilon_1\sqrt{\det A}\, I}{\sqrt{\operatorname{tr} A + 2\varepsilon_1\sqrt{\det A}}}. \tag{6a}$$

Here each $\varepsilon_i = \pm 1$, and if $\det A \neq 0$ the result determines exactly four square roots for A. However, if $\det A = 0$ then result (6a) determines two square roots for A as given by

$$X = \pm \frac{1}{\sqrt{\operatorname{tr} A}} A. \tag{6b}$$

Alternatively, if $(\operatorname{tr} A)^2 = 4 \det A \neq 0$, then one value of ε_1 in (5) reduces τ to zero whereas the other value yields the results, $2\varepsilon_1\sqrt{\det A} = \operatorname{tr} A$ and $\tau^2 = 2 \operatorname{tr} A$. In this case, A has exactly two square roots given by

$$X = \pm \frac{1}{\sqrt{2 \operatorname{tr} A}}\left(A + \frac{1}{2}(\operatorname{tr} A)I\right). \tag{6c}$$

Finally, if $(\operatorname{tr} A)^2 = 4 \det A = 0$ then both values of ε_1 reduce τ to zero in (5). Hence it follows by contradiction that A has no square roots.

For this case we conclude that a nonscalar matrix, A, has square roots if, and only if, at least one of the numbers, $\operatorname{tr} A$ and $\det A$, is nonzero. Then the matrix has four square roots given by (6a) if

$$(\operatorname{tr} A)^2 \neq 4 \det A, \ \det A \neq 0$$

and two square roots given by (6b) or (6c) if

$$(\operatorname{tr} A)^2 \neq 4 \det A, \ \det A = 0 \quad \text{or} \quad (\operatorname{tr} A)^2 = 4 \det A, \ \det A \neq 0.$$

It is worth noting from (6a) that

$$\operatorname{tr} X = \operatorname{tr} \sqrt{A} = \varepsilon_2\sqrt{\operatorname{tr} A + 2\varepsilon_1\sqrt{\det A}}\ .$$

Hence using the identity, $\det\sqrt{A} = \varepsilon_1\sqrt{\det A}$ as applied in (3), result (6a) may be rewritten as

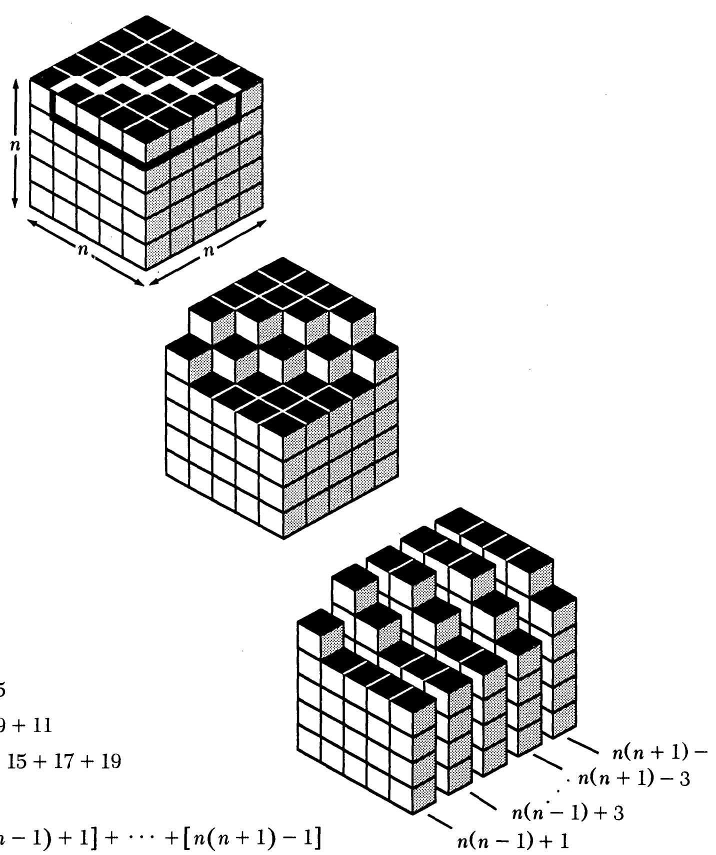
$$\sqrt{A} = \frac{1}{\operatorname{tr}\sqrt{A}}\left(A + \det\sqrt{A}\, I\right),$$

which is equivalent to the Cayley-Hamilton theorem for the matrix $\sqrt{A}$. This same deduction can be made, of course, for all other cases under which $\sqrt{A}$ exists.

REFERENCES

1. Nick MacKinnon, Four routes to matrix roots, *Math. Gaz.* 73 (1989), 135–136.
2. Nigel H. Scott, On square-rooting matrices, *Math. Gaz.* 74 (1990), 111–114.
3. Howard W. Eves, *Elementary Matrix Theory*, Dover Publications, Mineola, NY, 1980.
4. Roger A. Horn and Charles R. Johnson, *Topics in Matrix Analysis*, Camb. University Press, NY, 1990.

Proof without Words: Every Cube Is the Sum of Consecutive Odd Numbers



$1^3 = 1$

$2^3 = 3 + 5$

$3^3 = 7 + 9 + 11$

$4^3 = 13 + 15 + 17 + 19$

$\vdots$

$n^3 = [n(n-1)+1] + \cdots + [n(n+1)-1]$

—Roger B. Nelsen
Lewis and Clark College
Portland, OR 97219

Hence using the identity, $\det\sqrt{A} = \varepsilon_1\sqrt{\det A}$ as applied in (3), result (6a) may be rewritten as

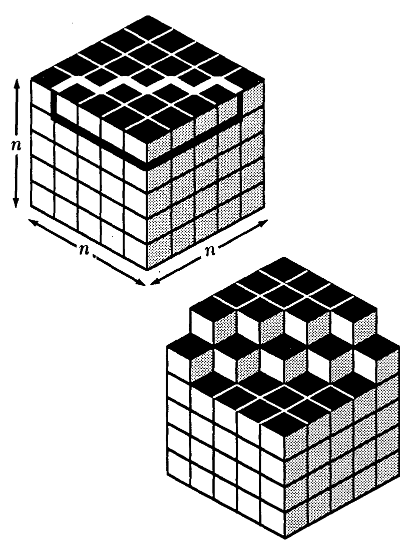
$$\sqrt{A} = \frac{1}{\operatorname{tr}\sqrt{A}}\left(A + \det\sqrt{A}\,I\right),$$

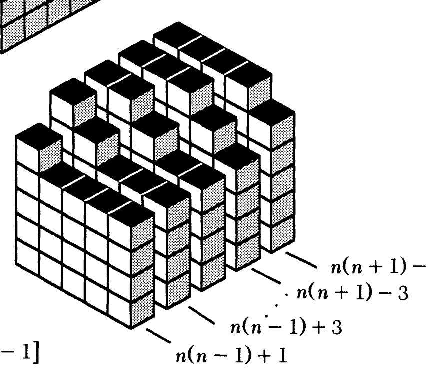
which is equivalent to the Cayley-Hamilton theorem for the matrix $\sqrt{A}$. This same deduction can be made, of course, for all other cases under which $\sqrt{A}$ exists.

REFERENCES

1. Nick MacKinnon, Four routes to matrix roots, *Math. Gaz.* 73 (1989), 135–136.
2. Nigel H. Scott, On square-rooting matrices, *Math. Gaz.* 74 (1990), 111–114.
3. Howard W. Eves, *Elementary Matrix Theory*, Dover Publications, Mineola, NY, 1980.
4. Roger A. Horn and Charles R. Johnson, *Topics in Matrix Analysis*, Camb. University Press, NY, 1990.

Proof without Words: Every Cube Is the Sum of Consecutive Odd Numbers





$1^3 = 1$

$2^3 = 3 + 5$

$3^3 = 7 + 9 + 11$

$4^3 = 13 + 15 + 17 + 19$

$\vdots$

$n^3 = [n(n-1)+1] + \cdots + [n(n+1)-1]$

—Roger B. Nelsen
Lewis and Clark College
Portland, OR 97219

A Graph of Primes

DONALD E. G. MALM
Oakland University
Rochester, MI 48309

Some time ago a colleague asked me a question about the graph formed by associating a vertex with each prime, and placing an edge between each pair of primes whose difference in absolute value is a nonnegative power of 2. His question was whether the graph formed in this way is connected. This kind of graph, which is called a *similarity* graph, is discussed in his text [**5**, p. 540]. A similarity graph is one in which vertices connected by an edge are not too different. In this case, primes that share an edge are similar in that their expressions in binary notation are close—their difference has only one nonzero binary digit.

We readily see that 2 shares an edge with 3, and so do 5, 7, 11, 19, etc. The prime 41 chains to 2: $41 \to 37 \to 5 \to 3 \to 2$. The prime 127 chains to 2, but not mononotically: $127 \to 131 \to 3 \to 2$. It's easy to check by hand that all primes less than a few hundred chain to 2. A computer or programmable calculator check verifies without too much effort that all primes less than 3000 also chain to 2, and with a little more effort one can check that all primes less than 40,000 chain to 2.

Despite the above evidence, the answer to my colleague's question is that the graph is not connected. Fred Cohen and J. L. Selfridge [**3**], extending a technique of Erdős, exhibit a 26-digit prime that is neither a sum nor a difference of a power of two and a prime, thus proving that the graph has an isolated node. The technique is a powerful but elementary use of congruences together with some help from computers. The purpose of this paper is to show how that technique works, in the process pushing a little further to obtain the result that there are infinitely many isolated nodes in the graph.

Actually, there are two natural graphs that one can consider—the graph consisting of all positive primes (the "small graph"), and the one that includes the negative primes as well (the "total graph"). Using the total graph, one can, for example, chain $97 \to -31 \to -23 \to -7 \to -3 \to -2 \to 2$, and one can chain $3181 \to -5011 \to -2963 \to -2707 \to -659 \to -643 \to -131 \to -3 \to 5 \to 3 \to 2$. In fact, if negative primes are not allowed, it is only possible to chain 3181 to 2 by using primes of 19 digits. This is because $3181 - 2^n$ is never a positive prime, while the smallest value of n for which $3181 + 2^n$ is prime is 60. (This is easily checked using a primality test and a computer.)

In 1950 P. Erdős [**4**] proved that there is an infinite arithmetic progression of odd integers, none of which can be written in the form $p + 2^n$, with p prime. Such an arithmetic progression will contain an infinite number of primes, and each of these primes shares no edge with a smaller prime. However, such a prime may very well chain to 2 anyway, just as 127 does. To deal with the graph of primes, one must also control $p - 2^n$.

The key concept used in the proof is that of a *covering system for a set S of integers*, which is a collection of k congruences of the form $y \equiv a_i \pmod{n_i}$, with $n_1 < n_2 < \cdots < n_k$, for which every integer y in S satisfies at least one of the congruences $y \equiv a_i \pmod{n_i}$. An example of a simple covering system for the set of all integers is $x \equiv 0 \pmod 2$, $0 \pmod 3$, $1 \pmod 4$, $1 \pmod 6$, $11 \pmod{12}$. If we drop the requirement that the n_i be distinct, we will call the result a *congruence set*. Guy [**6**, p. 140–141] discusses covering systems and unsolved questions about them.

The basic idea of the proof is this. If we can find a covering system $a_i \pmod{q_i}$ for the set of numbers 2^n, $n = 0, 1, 2, \ldots$ then by the Chinese Remainder Theorem there is an arithmetic progression for which every term x satisfies all the congruences $x \equiv -a_i \pmod{q_i}$. Suppose x is a member of this arithmetic progression that is bigger than all the q_i. Then $x + 2^n$ must be composite for all $n \geqslant 0$, since for any such n, there is an a_i with $2^n \equiv a_i \pmod{q_i}$, and therefore $x + 2^n \equiv 0 \pmod{q_i}$. To find isolated vertices in either graph, we must also handle $x - 2^n$ for $n \geqslant 0$. This can be done by using a second covering system $b_i \pmod{r_i}$ for the set 2^n, $n = 0, 1, 2, \ldots,$ where now we require $x \equiv b_i \pmod{r_i}$ rather than $x \equiv -b_i \pmod{r_i}$. The trouble is that these new congruences must not contradict the first ones, and we must also ensure that $x \pm 2^n$ is never equal to any of the moduli q_i or r_i, so that $x \pm 2^n \equiv 0$ implies that $x \pm 2^m$ actually is composite.

Covering systems for the set of powers of 2 can be constructed from certain congruence sets for the set of integers. The idea is illustrated by observing that if (say) $n \equiv 11 \pmod{12}$, then $n = 11 + 12m$ for some m, and $2^n = 2^{11} * 2^{12m}$. Since $2^{12} - 1$ has the prime factors 3, 5, 7, and 13, it follows that $2^{12} \equiv 1 \pmod{q}$ if q is any of the above primes, and $2^n \equiv 2^{11} \pmod{q}$. We see that we can get a covering system for the set of nonnegative powers of 2 from a congruence set for the set of all integers provided that the numbers $2^{n_i} - 1$ have sufficiently many different prime factors. In searching for such a covering system, the book [2] is invaluable, for it contains a table of the prime factors of $2^n - 1$ for positive $n < 250$. It is not necessary that the congruences for the set of all integers have distinct moduli, only that they lead to a bona fide covering system for the set of powers of 2. Table 1 lists two congruence sets for n and their corresponding covering sets for 2^n. These were found using a computer. (In this table, the number in parentheses following each residue is the corresponding modulus of the congruence.)

TABLE 1

$n \equiv$	$2^n \equiv$	$n \equiv$	$2^n \equiv$
0(2)	1(3)	1(2)	2(3)
0(3)	1(7)	0(4)	1(5)
1(8)	2(17)	0(5)	1(31)
1(9)	2(73)	0(7)	1(127)
11(12)	7(13)	2(10)	4(11)
7(18)	14(19)	2(14)	4(43)
5(24)	32(241)	1(15)	2(151)
31(36)	22(37)	2(16)	4(257)
13(36)	17(109)	14(20)	25(41)
		3(21)	8(337)
		6(28)	6(29)
		18(28)	97(113)
		8(30)	256(331)
		1(35)	2(71)
		58(60)	46(61)
		18(60)	586(1321)
		26(70)	163(281)
		206(21)	66(211)

Once the entries of the table have been found, it is simple to verify their correctness. The moduli for the first congruence set for n have l.c.m. 72, so one needs only to verify that each integer between 0 and 71 inclusive satisfies at least one of the congruences. The corresponding covering system for 2^n can be verified using the tables [2]. The second set of entries is handled in the same way. In this case the l.c.m. of the moduli for n is 1680.

Note that in each covering system for 2^n the moduli are distinct primes; also the only prime common to both systems is 3.

Now we can write down a system of congruences for which any solution will be an integer that is *not* of the form $q+2^n$ or $q-2^n$, with q prime. We indicate the modulus of each congruence by writing it in parentheses. The system is:

$$
\begin{array}{rrr}
x \equiv -1(3) & -1(7) & -2(17) \\
-2(73) & -7(13) & -14(19) \\
-32(241) & -22(37) & -17(109) \\
\hline
1(5) & 1(31) & 1(127) \\
4(11) & 4(43) & 2(151) \\
4(257) & 25(41) & 8(337) \\
6(29) & 97(113) & 256(331) \\
2(71) & 46(61) & 586(1321) \\
163(281) & 66(211) & \\
\hline
 & & 1000(2047)
\end{array}
\qquad (1)
$$

The 27 congruences of this system (1) come from three sources:

1) The first 9, from the first covering system, guarantee that if x satisfies each, then for each n, $x+2^n \equiv 0 \pmod p$ for at least one of the moduli p.

2) The next 17 are from the second covering system. Together with $x \equiv 2 \pmod 3$, which is the first congruence, they ensure that if x satisfies each of these 18 congruences, then for each n, $x-2^n \equiv 0 \pmod p$ for at least one of the moduli p. It follows that if x is a solution to the first 26 congruences, then every number of the form $x \pm 2^n$ must be composite unless it *is* one of the moduli.

3) The last congruence guarantees that the numbers of the form $x \pm 2^n$ cannot be equal to any of the moduli. For, since 2047 is one less than a power of two, the set of values of $2^n \pmod{2047}$ is $\{1,2,4,8,16,32,64,128,256,512,1024\}$. If x satisfies all 27 congruences, then $x \pm 2^n \pmod{2047}$ must be one of the numbers $\{1001, 1002, 1004, 1008, \ldots, 1512, 2024, 999, 998, 996, 992, \ldots, 488, -24\ (\equiv 2023)\}$. However, none of the prime moduli (mod 2047) are any of these values. It follows that if x is any solution of this system of 27 congruences, then $x \pm 2^n$ must be divisible by one of the prime moduli of the first 26, but cannot be equal to any of those moduli, and so must be composite.

The Chinese Remainder Theorem guarantees that there is an infinite arithmetic progression of solutions to system (1). (Note that the moduli are relatively prime, for $2047 = 23 * 89$.) Dirichlet's theorem [7, p. 31] guarantees that this arithmetic progression contains infinitely many primes, since the first term and the common difference must be relatively prime. These primes are isolated in both the total graph and the small graph.

The solution of system (1) is the arithmetic progression $a+bm$, where m is the parameter and

$$a = 16072\ 35727\ 03020\ 46589\ 74480\ 25537\ 91940\ 51479\ 85029\ 22751,$$
$$b = 23526\ 41407\ 50797\ 04441\ 84622\ 22680\ 98507\ 82474\ 11014\ 41905.$$

To exhibit a specific isolated prime as small as our covering systems allow, we omit the 27th congruence from (1), solve this smaller system, find the smallest positive prime q in the arithmetic progression of solutions, and verify directly that $q-2^n$ is never equal to any of the prime moduli in the first 26 congruences. This yields

$$q = 293\ 84382\ 54055\ 73891\ 53952\ 59805\ 37456\ 23284\ 74806\ 89009,$$

which is the smallest prime I have found that is isolated in the total graph, though the prime of Cohen and Selfridge mentioned above is smaller (their covering sets are different from ours).

One can find smaller primes that are isolated in the small graph by using only the first covering system to find an arithmetic progression of numbers x for which $x+2^n$ is never prime. Then find a positive prime q in this arithmetic progression for which $q-2^n$ is never a positive prime. This process yields $q = 5404\ 26473$, the smallest prime I found that is isolated in the graph of positive primes. This prime is not isolated in the total graph, for $q-2^{126}$ is prime.

While the original question is answered, very little else seems to be known about these two graphs. Natural questions are (we use the term degree in the usual graph-theoretic sense—the degree of a vertex is the number of edges it has):

1) Do any vertices have infinite degree?

2) In the total graph, are there any vertices of finite positive degree other than 2 and -2? In the small graph, one can find, as above, primes that don't chain upward and therefore have finite degree. I have found such vertices of degrees 0 through 6 inclusive. However, even in the small graph I do not know whether there are infinitely many vertices of finite positive degree nor whether there are vertices of arbitrarily large degree.

3) Does there exist an infinite connected component?

4) Is there a connected finite component of more than one vertex?

5) What is the smallest prime not in the component containing 2?

6) What is the smallest isolated node? Is it the same prime as in 5)?

Here, as in much of number theory, simple questions arise that seem difficult to answer.

I thank Jerrold Grossman for raising this problem, and for a careful reading of the manuscript. I also thank the referees for their helpful comments, and for bringing [3] to my attention. The numerical computations were done on Macintosh and PC computers, using my multiprecision number theory package (I distribute this package without charge via ftp, see [1] for details) and also routines distributed with the language UBASIC, most notably the Adelman-Pomerance-Rumely-Cohen-Lenstra primality test.

REFERENCES

1. Jon Barwise, Computers and Mathematics, *Notices Amer. Math. Soc.* 38 (1991), 104–110.
2. J. Brillhart, D. H. Lehmer, J. L. Selfridge, B. Tuckerman, and S. S. Wagstaff, Jr., *Factorizations of* b^n+1, $b = 2,3,5,6,7,10,11,12$ *up to High Powers*, Amer. Math. Soc., Providence, RI, 1983.
3. Fred Cohen and J. L. Selfridge, Not every number is the sum or difference of two prime powers, *Math. of Comp.* 29 (1975), 79–81.
4. P. Erdős, On integers of the form $2k+p$ and some related problems, *Summa Brasiliensis Mathematica* 2 (1950), 113–123.
5. J. Grossman, *Discrete Mathematics*, Macmillan Publishing Co., New York, 1990.
6. Richard K. Guy, *Unsolved Problems in Number Theory*, Springer-Verlag New York, Inc., New York, 1981.
7. Ivan Niven, Herbert S. Zuckerman, and Hugh L. Montgomery, *An Introduction to the Theory of Numbers*, 5th edition, John Wiley & Sons, Inc., New York, 1991.

Mathematical Ballroom Dancing

HOUSTON EULER

In [1] Yardley Beers refers to work done on a mathematical theory of ballroom dancing, but those results were never published, to my knowledge. Much progress has occurred in this field during the subsequent years, but again nothing has appeared in print, the results just circulating as folk dance theorems. It therefore seems appropriate to give at least a brief survey of the main approaches that have been pursued.

Integration and measure theory Since every piece of dance music is a composition of several measures, this approach can be used with great generality. Typically, a dance will be expressed as a function f such that $f(t)$ equals the place where the dancer steps at time t. In other words, f is a step function (or sometimes named, after one of the leading exponents of this method, Astaire step function), so its integral is easy to compute.

Plane geometry Although elementary geometry is more often associated with square dancing, it has also been successfully applied to Latin ballroom dances, such as the rhomba. With another Latin dance, properties of circles have yielded the famous Lemma on Merengue Pi.

Point-set topology Jitterbug dances lend themselves well to topological methods, especially those dances whose main step occurs on the second count of a measure. These second countable dances are exemplified by the Lindylof.

Vectors Techniques from elementary linear algebra and vector calculus are helping in dealing with a variety of dances. It suffices to mention the use of tango vectors and the polka-dot product.

REFERENCE

1. Y. Beers, in "American Institute of Useless Research" *A Random Walk in Science*, ed. by R. L. Weber and E. Mendoza, The Institute of Physics, 1973, pp. 23–24.

The Central Conic Sections Revisited

AYOUB B. AYOUB
The Pennsylvania State University
Abington, PA 19001

If a central conic section is given by the general quadratic equation

$$f(x,y) = ax^2 + 2nxy + by^2 + 2hx + 2ky + c = 0, \tag{1}$$

how do we find its center, the equations of its principal axes and the radii along these axes? Usually by translating and rotating the coordinate axes. Here, however, we will use a combination of techniques that will shed more light on this topic.

First, we will get the center by using the fact that it is the midpoint of any diameter passing through it. Assume that the center of the conic is (x_0, y_0). Then the equation of the diameter parallel to the x-axis is $y = y_0$, and if we substitute that in (1), we get $ax^2 + (2ny_0 + 2h)x + by_0^2 + 2ky_0 + c = 0$. If the roots of this equation are x_1 and x_2, then the x value of the center, x_0, is given by

$$x_0 = \frac{1}{2}(x_1 + x_2) = -\frac{1}{2}\left(\frac{\text{coefficient of } x}{\text{coefficient of } x^2}\right).$$

Thus,

$$x_0 = -\frac{1}{2}\left(\frac{2ny_0 + 2h}{a}\right), \quad \text{that is,}$$

$$ax_0 + ny_0 + h = 0. \tag{2}$$

Similarly if we use the diameter $x = x_0$, we get

$$nx_0 + by_0 + k = 0. \tag{3}$$

Equations (2) and (3) will determine the center (x_0, y_0) of the conic given by (1), and it is easy to see that these equations are equivalent to $\partial f/\partial x = 0$ and $\partial f/\partial y = 0$.

One may ask: What's so special about the diameters represented by (2) and (3)? You may notice that if (x_0, y_0) was a midpoint of any chord parallel to the x-axis, it also would have satisfied the equation $ax + ny + h = 0$. Hence this equation represents the diameter bisecting all chords parallel to the x-axis. The endpoints of the diameter, being the limiting cases of the chords, are the points where the tangents parallel to the x-axis touch the conic, and that explains why these points satisfy the condition $\partial f/\partial x = 0$. By the same token $nx + by + k = 0$ represents the diameter bisecting the chords parallel to the y-axis. FIGURE 1 depicts such two diameters in a case of an ellipse.

Solving equations (2) and (3) we get $x_0 = H/C$ and $y_0 = K/C$, where H, K, C are the cofactors of h, k, c, respectively, in the determinant

$$\Delta = \begin{vmatrix} a & n & h \\ n & b & k \\ h & k & c \end{vmatrix}.$$

Of course, if $C = 0$, this indicates that we are dealing with noncentral conic section, i.e. either a parabola or a pair of parallel lines.

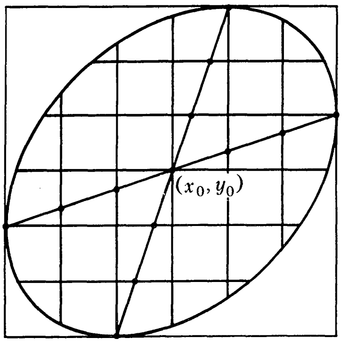


FIGURE 1

If we replace x and y of (1) by $x + x_0$ and $y + y_0$ to translate the origin to the center of the conic we get

$$ax^2 + 2nxy + by^2 + x_0(ax_0 + ny_0 + h) + y_0(nx_0 + by_0 + k) + hx_0 + ky_0 + c = 0.$$

Making use of (2), (3), $x_0 = H/C$, $y_0 = K/C$, and $hH + kK + cC = \Delta$ the equation of the conic, with its center now at the origin, becomes

$$ax^2 + 2nxy + by^2 + \frac{\Delta}{C} = 0$$

or

$$ax^2 + 2nxy + by^2 + \frac{\Delta}{ab - n^2} = 0.$$

If $\Delta = 0$ then the equation represents a pair of straight lines. Let us assume that $\Delta \neq 0$ and proceed to find the extreme radii of the conic in the nontrivial case when $n \neq 0$. To that end we use polar coordinates and in this case the previous equation takes the form

$$r^2(a\cos^2\theta + 2n\sin\theta\cos\theta + b\sin^2\theta) + \frac{\Delta}{ab - n^2} = 0.$$

Since $\Delta/(ab - n^2)$ is constant then r is maximum or minimum if $(a\cos^2\theta + 2n\sin\theta\cos\theta + b\sin^2\theta)$ is minimum or maximum, respectively, and this occurs when its derivative is equal to zero. But

$$-2a\cos\theta\sin\theta + 2n\cos^2\theta - 2n\sin^2\theta + 2b\sin\theta\cos\theta = 0$$

is equivalent to

$$n\tan^2\theta + (a - b)\tan\theta - n = 0. \tag{4}$$

If we let $\tan\theta = m$, then

$$nm^2 + (a - b)m - n = 0 \tag{5}$$

will produce the slopes of the principal axes of the conic section. Now using the substitution $m = (\lambda - a)/n$ (or $m = n/(\lambda - b)$) equation (5) takes the form

$$(\lambda - a)(\lambda - b) - n^2 = 0 \tag{6}$$

or

$$\begin{vmatrix} \lambda - a & n \\ n & \lambda - b \end{vmatrix} = 0.$$

The roots λ_1 and λ_2 of equation (6) are the eigenvalues of the matrix

$$A=\begin{bmatrix} a & n \\ n & b \end{bmatrix}.$$

Hence A is similar to the diagonal matrix

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

through some orthogonal matrix (isometric) transformation [1]. This implies that the equation $ax^2+2nxy+by^2+\Delta/(ab-n^2)=0$, which can be written as

$$[x \quad y]\begin{bmatrix} a & n \\ n & b \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}+\frac{\Delta}{ab-n^2}=0, \quad \text{would take the form}$$

$$[x \quad y]\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}+\frac{\Delta}{ab-n^2}=0.$$

Thus, $\lambda_1 x^2+\lambda_2 y^2+\Delta/(ab-n^2)=0$. But (6) implies that $\lambda_1\lambda_2=ab-n^2$. Hence,

$$\lambda_1 x^2+\lambda_2 y^2+\frac{\Delta}{\lambda_1\lambda_2}=0 \tag{7}$$

is the equation of the conic section referred to its principal axes. And this is exactly the equation we get if we rotate the coordinate axes through an angle θ of (4). From (7) we conclude that the squares of the radii along the principal axes are

$$-\frac{\Delta}{\lambda_1^2\lambda_2} \quad \text{and} \quad -\frac{\Delta}{\lambda_1\lambda_2^2}$$

and the equations of the corresponding principal axes are

$$y-y_0=m_1(x-x_0) \quad \text{and} \quad y-y_0=m_2(x-x_0),$$

where

$$m_1=\frac{\lambda_1-a}{n} \quad \text{and} \quad m_2=\frac{\lambda_2-a}{n},$$

and (x_0, y_0) is the center of the conic section.

Example. Consider the equation $17x^2-12xy+8y^2+12x-16y-12=0$. Here

$$\Delta=\begin{vmatrix} 17 & -6 & 6 \\ -6 & 8 & -8 \\ 6 & -8 & -12 \end{vmatrix}=-2000,$$

and

$$\begin{vmatrix} \lambda-17 & -6 \\ -6 & \lambda-8 \end{vmatrix}=0 \Rightarrow \lambda_1=5 \quad \text{and} \quad \lambda_2=20.$$

The equation of the ellipse referred to its principal axes is

$$5x^2+20y^2-\frac{2000}{5(20)}=0, \quad \text{that is,} \quad \frac{x^2}{4}+\frac{y^2}{1}=1.$$

Since the center satisfies $17x-6y+6=0$ and $-6x+8y-8=0$, the center is $(0,1)$.

The slopes of the principal axes are

$$m_1 = \frac{\lambda_1 - a}{n} = \frac{5-17}{-6} = 2 \quad \text{and} \quad m_2 = \frac{\lambda_2 - a}{n} = \frac{20-17}{-6} = -\frac{1}{2}.$$

The equations of the principal axes are

$$y - 1 = 2(x - 0) \quad \text{and} \quad y - 1 = -\tfrac{1}{2}(x - 0), \quad \text{or equivalently,}$$

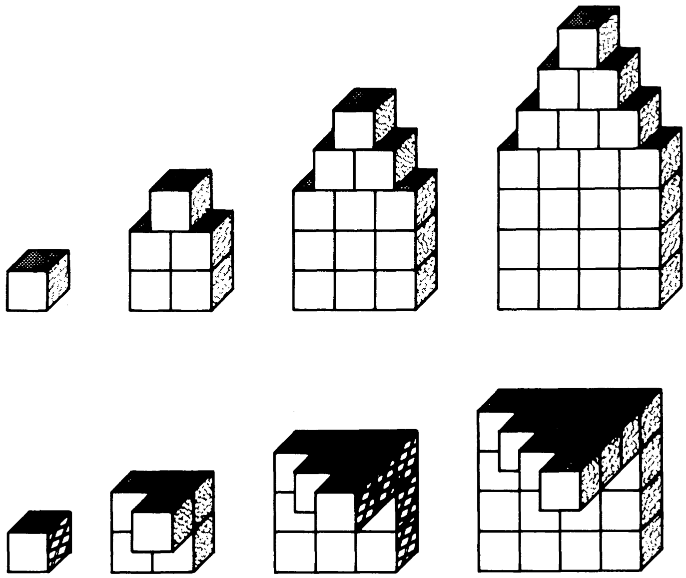
$$y = 2x + 1 \quad \text{and} \quad y = -\tfrac{1}{2}x + 1.$$

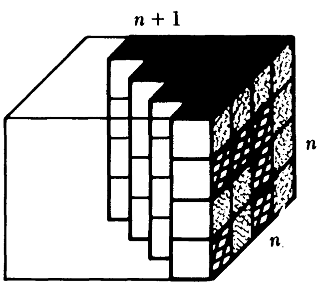
REFERENCE

1. H. Anton, *Elementary Linear Algebra*, 6th edition, John Wiley & Sons, Inc., New York, 1991, p. 418.

Proof without Words: Sum of Pentagonal Numbers

$$\frac{1\cdot 2}{2} + \frac{2\cdot 5}{2} + \frac{3\cdot 8}{2} + \cdots + \frac{n(3n-1)}{2} = \frac{n^2(n+1)}{2}$$





—William A. Miller
Central Michigan University
Mount Pleasant, MI 48859

The slopes of the principal axes are

$$m_1 = \frac{\lambda_1 - a}{n} = \frac{5-17}{-6} = 2 \quad \text{and} \quad m_2 = \frac{\lambda_2 - a}{n} = \frac{20-17}{-6} = -\frac{1}{2}.$$

The equations of the principal axes are

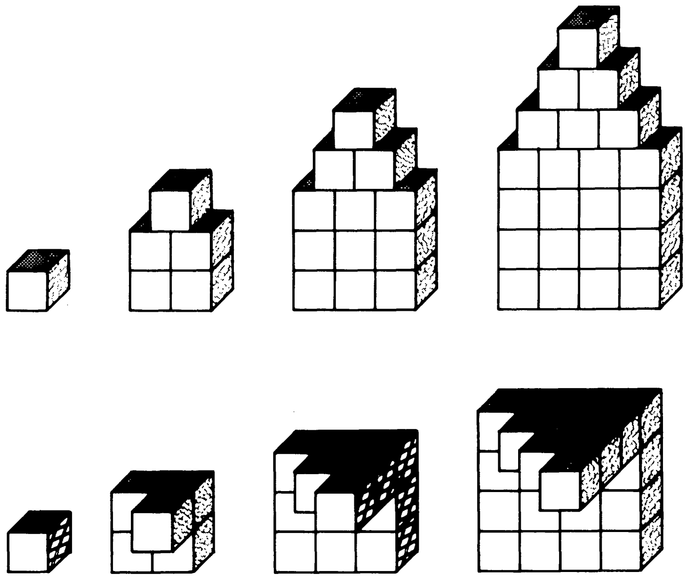
$$y - 1 = 2(x-0) \quad \text{and} \quad y - 1 = -\tfrac{1}{2}(x-0), \quad \text{or equivalently,}$$
$$y = 2x + 1 \quad \text{and} \quad y = -\tfrac{1}{2}x + 1.$$

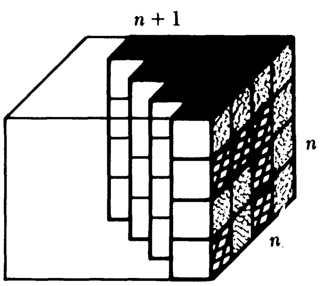
REFERENCE

1. H. Anton, *Elementary Linear Algebra*, 6th edition, John Wiley & Sons, Inc., New York, 1991, p. 418.

Proof without Words: Sum of Pentagonal Numbers

$$\frac{1\cdot 2}{2} + \frac{2\cdot 5}{2} + \frac{3\cdot 8}{2} + \cdots + \frac{n(3n-1)}{2} = \frac{n^2(n+1)}{2}$$





—William A. Miller
Central Michigan University
Mount Pleasant, MI 48859

Comparing Comparisons: Infinite Sums vs. Partial Sums

KENDALL RICHARDS
Southwestern University
Georgetown, TX 78626

The following problem originated during a classroom discussion:

Let C be the complex plane and let $D=\{z\in C\colon |z|<1\}$. Suppose that $f(z)=\sum_{n=0}^{\infty}a_n z^n$ and $g(z)=\sum_{n=0}^{\infty}b_n z^n$ are analytic in D. How does one compare two complex-valued functions? One common method is to use some type of "function measurement" or *functional*, F, that assigns real values $F(f)$ and $F(g)$ to f and g, respectively. In particular, we consider the following two conditions on f and g:

$$\text{(a)}\qquad \sum_{k=0}^{\infty}|b_k|^2r^{2k}\le\sum_{k=0}^{\infty}|a_k|^2r^{2k}\ \textit{for all}\ 0\le r<1.$$

$$\text{(b)}\qquad \sum_{k=0}^{n}|b_k|^2\le\sum_{k=0}^{n}|a_k|^2\ \textit{for all}\ n=0,1,2,\dots.$$

The following question arises: Which is the stronger condition, (a) or (b)? In other words, how do the two methods of comparison compare?

It is interesting to note that both conditions are known to hold in the special case that g is subordinate to f (i.e., $g(z)=f(w(z))$ for some w analytic in D satisfying $|w(z)|\le|z|$ for all $z\in D$). That is, we have the following:

LITTLEWOOD'S SUBORDINATION THEOREM. *If g is subordinate to f, then* (a) *holds.*

ROGOSINSKI'S THEOREM. *If g is subordinate to f, then* (b) *holds.*

Furthermore, an examination of the proof in Duren [1] of Rogosinski's Theorem reveals that Littlewood's Subordination Theorem is indirectly used. This observation might lead one to believe that, if a direct implication held, (a) would imply (b). Somewhat surprisingly, we have the following:

THEOREM. *Let $f(z)=\sum_{n=0}^{\infty}a_n z^n$ and $g(z)=\sum_{n=0}^{\infty}b_n z^n$ be analytic in D. If* (b) *holds then* (a) *holds. The converse is not true.*

Proof. Let f and g be analytic in D such that (b) is satisfied. Let $0\le r<1$ and let $A_n=\sum_{k=0}^{n}|a_k|^2$, $B_n=\sum_{k=0}^{n}|b_k|^2$. Since $r^{2k}\ge r^{2k+2}\ge\cdots\ge 0$ and $B_k\le A_k$ for all $k\ge 0$, we obtain (see [1] p. 193), using summation by parts, that

$$\sum_{k=0}^{n}|b_k|^2r^{2k}=\sum_{k=0}^{n-1}(r^{2k}-r^{2k+2})B_k+r^{2n}B_n$$
$$\le\sum_{k=0}^{n-1}(r^{2k}-r^{2k+2})A_k+r^{2n}A_n=\sum_{k=0}^{n}|a_k|^2r^{2k}.$$

Letting $n\to\infty$, we have that (a) holds.

To see that (a) does not imply (b), consider the following example: Let $f(z)=\sum_{k=0}^{\infty}z^{2k}$ and $g(z)=(1+\varepsilon)z+\sum_{k=1}^{\infty}z^{2k+1}$, where $\varepsilon>0$ is chosen so that $2\varepsilon+\varepsilon^2\le 1/2$. Define $I_2(r,f)=\sum_{k=0}^{\infty}|a_k|^2r^{2k}$. It follows that

$$I_2(r,f) = \sum_{k=0}^{\infty} r^{4k} = 1/(1-r^4) \text{ and}$$

$$\begin{aligned} I_2(r,g) &= (1+\varepsilon)^2 r^2 + \sum_{k=1}^{\infty} r^{4k+2} \\ &= (2\varepsilon + \varepsilon^2) r^2 + \sum_{k=0}^{\infty} r^{4k+2} \\ &\le 1/2 + r^2/(1-r^4). \end{aligned}$$

Hence, $I_2(r,f) - I_2(r,g) \ge (1-r^2)/(1-r^4) - 1/2 = 1/(1+r^2) - 1/2 \ge 1/2 - 1/2 = 0$. Thus (a) holds. But when $n = 1$, the left-hand side of (b) is equal to $(1+\varepsilon)^2$ while the right-hand side of (b) is equal to 1. Thus, (b) is not satisfied for this f and g. This proves the theorem.

Indeed, given any prescribed nonnegative integer N, we can show that inequality (a) can hold while the inequality in (b) holds for n up to but not including $2N+1$. Let f be as above and choose $\varepsilon > 0$ such that $(1+\varepsilon)^{4N+2} - 1 \le 1/2$. Define

$$g(z) = (1+\varepsilon)^{2N+1} z^{2N+1} + \sum_{k=0,\, k \ne N}^{\infty} z^{2k+1}.$$

$$\begin{aligned} \text{Now, } I_2(r,g) &= (1+\varepsilon)^{4N+2} r^{4N+2} + \sum_{k=0,\, k\ne N}^{\infty} r^{4k+2} \\ &= \left\{(1+\varepsilon)^{4N+2} - 1\right\} r^{4N+2} + \sum_{k=0}^{\infty} r^{4k+2} \\ &\le 1/2 + r^2/(1-r^4). \end{aligned}$$

Thus as above, $I_2(r,g) \le I_2(r,f)$ and so (a) holds. Also, the inequality in (b) is satisfied for all $n < 2N+1$. But for $n = 2N+1$, the left-hand side of (b) is equal to $N + (1+\varepsilon)^2$ while the right-hand side is equal to $N+1$. Therefore (b) is not satisfied.

Remark: One may also write $I_2(r,f) = 1/2\pi \int_0^{2\pi} |f(re^{i\theta})|^2 \, d\theta$, by Parseval's Identity [2].

REFERENCES

1. P. L. Duren, *Univalent Function Theory*, Springer-Verlag New York, Inc., New York, 1983.
2. W. Rudin, *Real and Complex Analysis*, 2nd edition, McGraw-Hill Book Co., New York, 1974.

Interchanging Parameters of the Hypergeometric Distribution

R. R. DAVIDSON
B. R. JOHNSON
University of Victoria
Victoria, B.C., Canada V8W 3P4

Recently in a finite mathematics class, errors made by some of our students led us to rediscover an interesting property of the hypergeometric distribution. Our students accidentally demonstrated that it is possible to interchange certain parameters of the hypergeometric distribution and still obtain the correct probability values. The purpose of this note is to give a probabilistic rationale for this phenomenon.

Consider an experiment in which n balls are chosen at random without replacement from an urn that contains N balls of which K are black and $N-K$ are white, and suppose the random variable of interest is the number, X, of black balls chosen. Instead of using the correct hypergeometric formula

$$P(X=x)=\binom{K}{x}\binom{N-K}{n-x}\Big/\binom{N}{n},$$

some students wrongly interchanged K with n, leading to the formula

$$P(X=x)=\binom{n}{x}\binom{N-n}{K-x}\Big/\binom{N}{K}.$$

The fact that this interchange does not change the probability values is verified immediately by expressing the hypergeometric formula in terms of factorials.

We offer the following probabilistic argument to demonstrate that the number of black balls drawn in both cases are identically distributed. We begin with an urn that contains N white balls. When Ms. Painter arrives, she picks K balls at random without replacement from the urn, then paints each of the drawn balls black with instant dry paint, and finally returns these K balls to the urn. When Mr. Carver arrives, he chooses n balls at random without replacement from the urn, then engraves each of the chosen balls with the letter C, and finally returns these n balls to the urn. Let the random variable X denote the number of black balls chosen (*i.e.* painted and engraved) when both Painter and Carver have finished their jobs. Since the tasks of Painter and Carver do not depend on which task is done first, the probability distribution of X is the same whether Painter does her job before or after Carver does his. If Painter goes first, then Carver chooses n balls from an urn that contains K black balls and $N-K$ white balls, so

$$P(X=x)=\binom{K}{x}\binom{N-K}{n-x}\Big/\binom{N}{n},$$

for x an integer satisfying $\max(0,n-(N-K))\leqslant x\leqslant\min(n,K)$.

On the other hand, if Carver goes first, then Painter draws K balls from an urn that contains n balls engraved with C and $N-n$ balls not engraved, so

$$P(X=x)=\binom{n}{x}\binom{N-n}{K-x}\Big/\binom{N}{K}$$

for x an integer satisfying $\max(0, K-(N-n)) \leqslant x \leqslant \min(n, K)$.

Thus, changing the order of who goes first provides an urn-drawing explanation of why the probability distribution of X remains the same when K is interchanged with n.

REFERENCES

1. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd edition, John Wiley & Sons, Inc., New York, 1968.
2. J. G. Kalbfleisch, *Probability and Statistical Inference*, Vol. 1, 2nd edition, Springer-Verlag New York, Inc., New York, 1985.
3. S. Ross, A *First Course in Probability*, 3rd ed., MacMillan Publishing Co., New York, 1988.

Proof Without Words: kth Power of a Natural Number n as the Sum of n Consecutive Odd Numbers ($k = 2, 3, \ldots$)

$$n^k = \sum_{r=1}^{n} (n^{k-1} - n + 2r - 1)$$

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

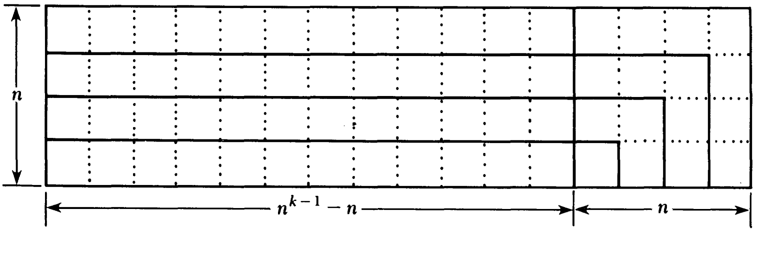
—N. GOPALAKRISHNAN NAIR
KERALA STATE, INDIA

for x an integer satisfying $\max(0, K-(N-n)) \leqslant x \leqslant \min(n, K)$.
Thus, changing the order of who goes first provides an urn-drawing explanation of why the probability distribution of X remains the same when K is interchanged with n.

REFERENCES

1. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd edition, John Wiley & Sons, Inc., New York, 1968.
2. J. G. Kalbfleisch, *Probability and Statistical Inference*, Vol. 1, 2nd edition, Springer-Verlag New York, Inc., New York, 1985.
3. S. Ross, *A First Course in Probability*, 3rd ed., MacMillan Publishing Co., New York, 1988.

Proof Without Words: kth Power of a Natural Number n as the Sum of n Consecutive Odd Numbers ($k = 2, 3, \ldots$)



$$n^k = \sum_{r=1}^{n} (n^{k-1} - n + 2r - 1)$$

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

—N. GOPALAKRISHNAN NAIR
KERALA STATE, INDIA

An Elementary Extension of Tietze's Theorem

MARTIN W. BARTELT
Christopher Newport University
Newport News, VA 23606

JOHN J. SWETITS
Old Dominion University
Norfolk, VA 23508

A famous, interesting, and old result usually encountered in an introductory topology or analysis course is the Tietze Extension Theorem ([3], p. 149):

THEOREM T. *Let X be a normal Hausdorff space and A a closed subset of X. If $f \in C(A)$, then f has a continuous extension $F \in C(X)$. Furthermore, if $|f(x)| \leq c$ for each $x \in A$, then F can be chosen so that $|F(x)| \leq c$ for each $x \in X$.*

Here $C(X)$ and $C(A)$ are the sets of continuous real-valued functions on X and A, respectively. F is an extension of f means $F(x) = f(x)$ if $x \in A$.

While Theorem T is included in almost any text on point-set topology, none of the many books we surveyed mentions anything more general than Theorem T, either as a corollary or as an exercise. Of course, some texts observe that by translation one can extend a function f satisfying $M_1 \leq f(x) \leq M_2$, $x \in A$, to a function F satisfying $M_1 \leq F(x) \leq M_2$, $x \in X$ when M_1 and M_2 are any two constants, not just $M_2 = c = -M_1$ as given in Theorem T.

It should be observed that the original Tietze Theorem was stated for metric spaces and later generalized by Urysohn to normal Hausdorff spaces. Also, some extensions of Theorem T different from what is presented here appear in the literature (cf. [2]).

This note considers the following question: Let A be a closed subset of X and assume $g, h \in C(X)$ such that $g(x) \leq h(x)$ for every $x \in X$. If f is a function in $C(A)$ lying between g and h, i.e. $g(x) \leq f(x) \leq h(x)$ for every x in A, can f be extended to a function F in $C(X)$ lying between the same two functions?

The following easy exercises lead directly from Theorem T to an affirmative answer to the question. However, we show later that there is a more general version that can be verified immediately from Theorem T.

Exercises

(1) If $f \in C(A)$ satisfies $0 \leq f(x) \leq c$, $x \in A$, then there is an extension F of f to X satisfying $0 \leq F(x) \leq c$.
(2) If $f \in C(A)$ and $g \in C(X)$ satisfy $f(x) \leq g(x)$, $x \in A$, then there is an extension F of f with $F(x) \leq g(x)$, $x \in X$.
(3) Given $f \in C(A)$, $g \in C(X)$ such that $0 \leq f(x) \leq g(x)$, $x \in A$, and $0 \leq g(x)$, $x \in X$, then there an extension F of f such that $0 \leq F(x) \leq g(x)$, $x \in X$.
(4) Given $f \in C(A)$, and $h, g \in C(X)$ with $h(x) \leq f(x) \leq g(x)$, $x \in A$ and $h(x) \leq g(x)$, $x \in X$, then there exists an extension F of f with $h(x) \leq F(x) \leq g(x)$, $x \in X$.

To verify (1) the student need only apply Theorem T to obtain a function $f_1(x)$ with $-c \leq f_1(x) \leq c$ and $f_1(x) = f(x)$ for every x in A, and then the function

$F(x) = \max\{f_1(x), 0\}$ will satisfy $0 \le F(x) \le c$ and $F(x) = f(x)$ for every x in A. Now (2) follows by applying (1) to the function $w(x) = g(x) - f(x)$, which satisfies $w(x) \ge 0$ for every x in A to obtain a function W defined on X and letting $F(x) = g(x) - W(x)$. Then (3) follows from (2) just as (1) followed from Theorem T. Finally, to verify (4), let $w(x) = f(x) - h(x)$ and apply (3) to obtain W in $C(X)$ and let $F(x) = W(x) + h(x)$.

For a more general version of Theorem T, we need the concept of equicontinuity (cf. [3]). A subset $S \subseteq C(X)$ is called *equicontinuous at* $x_0 \in X$ if given any $\varepsilon > 0$ there exists a neighborhood U of x_0 such that $|f(x) - f(x_0)| < \varepsilon$ for every $x \in U$ and every $f \in S$. The subset S is called *equicontinuous* if it is equicontinuous at each point of X.

THEOREM. *Let X be a normal Hausdorff space and A a closed subset of X. Let $f \in C(A)$ and let P and Q be nonempty, equicontinuous subsets of $C(X)$. Suppose that for each $p \in P$, $q \in Q$ and $x \in A$*

$$p(x) \le f(x) \le q(x). \tag{1}$$

Then f has an extension $F \in C(X)$ such that for each $x \in X$ and each $p \in P$ and $q \in Q$,

$$p(x) \le F(x) \le q(x) \tag{2}$$

if, and only if,

$$r(x) = \sup\{p(x) \colon p \in P\} \le \inf\{q(x) \colon q \in Q\} = s(x). \tag{3}$$

Proof. From (2) the necessity is obvious. To prove sufficiency, we note that the equicontinuity hypothesis implies that r and s are in $C(X)$, and

$$r(x) \le f(x) \le s(x), \quad \text{for each } x \in A.$$

To see, for example, that s is continuous, let $x_0 \in X$ and $\varepsilon > 0$ be given. Let U be a neighborhood of x_0 such that

$$q(x_0) - \varepsilon < q(x) < q(x_0) + \varepsilon$$

for each $x \in U$ and each $q \in Q$. Then taking the infimum over all $q \in Q$ we obtain

$$s(x_0) - \varepsilon \le s(x) \le s(x_0) + \varepsilon,$$

and s is continuous at x_0.

Define $w(x) = f(x) - r(x)$ for $x \in A$. By Theorem T, w has an extension $W \in C(X)$. Define $W_1 \in C(X)$ by

$$W_1(x) = \begin{cases} 0, & \text{if } W(x) < 0 \\ s(x) - r(x), & \text{if } s(x) - r(x) < W(x) \\ W(x), & \text{if } 0 \le W(x) \le s(x) - r(x). \end{cases}$$

Then it is easily verified that $F = W_1 + r$ is the desired extension of f.

The following example shows that the assumption of equicontinuity cannot be omitted.

Example. Let $X = [-1, 1]$ and $A = \{0\}$. For $n = 1, 2, \ldots$, let $p_n(x) = 0$, $-1 \le x \le 1$, and let

$$q_n(x) = \begin{cases} 0, & \text{if } -1 \le x \le -\dfrac{1}{n} \\ 0, & \text{if } \dfrac{1}{n} \le x \le 1 \\ 1, & \text{if } x = 0 \\ \text{linear}, & \text{if } -\dfrac{1}{n} \le x \le 0 \\ \text{linear}, & \text{if } 0 \le x \le \dfrac{1}{n} \end{cases}$$

Then, for any $x \in X$, $\sup_n p_n(x) \le \inf_n q_n(x)$ and

$$\inf_n q_n(x) = \begin{cases} 0, & x \ne 0 \\ 1, & x = 0. \end{cases}$$

Thus if $0 < f(0) \le 1$, f has no continuous extension of the desired type.

Theorem T and its generalization provide interesting examples of the interconnections of various areas of mathematics. They are useful in approximation theory where one of the central objects of study is the best approximation $B(f, X)$ from a subset $S \subseteq C(X)$ to a given function $f \in C(X)$ where X is the real closed interval $[a, b]$; i.e., $B(f, X)$ is a function in S such that

$$\| f - B(f, X) \|_X \le \| f - s \|_X \quad \text{for every } s \in S.$$

Here we use the uniform norm $\|f\|_X = \sup\{|f(x)|: x \in X\}$ with $\|f\|_A$ defined similarly. Let A be a closed subset of X. Let $B(f, A)$ denote the function in S that is the best approximate on A to a given function $f \in C(A)$. One seeks a function $F \in C(X)$ such that F is an extension of f, $B(f, A)$ is the best approximate to F on all of X, and $\|F - B(f, A)\|_X = \|f - B(f, A)\|_A$. Analytically, this is equivalent to requiring the existence of a function $F \in C(X)$ such that

$$B(f, A)(x) - \| f - B(f, A) \|_A \le F(x) \le B(f, A)(x) + \| f - B(f, A) \|_A,$$

for every $x \in X$. One can obtain this extension using the Theorem (or Exercise (4)) and the fact that

$$B(f, A)(x) - \| f - B(f, A) \|_A \le f(x) \le B(f, A)(x) + \| f - B(f, A) \|_A,$$

for every $x \in A$. For an introduction to approximation theory, the interested reader is invited to see [1].

REFERENCES

1. E. W. Cheney, *Introduction to Approximation Theory*, McGraw-Hill, New York, 1966.
2. J. Dugundji, An Extension of Tietze's Theorem, *Pacific J. Math* (1951), 353–357.
3. J. Dugundji, *Topology*, Allyn and Bacon, Boston, 1970.

The Fourth Side of a Triangle

B. F. SHERMAN
University of Adelaide
Australia 5000

One of the earliest geometrical facts we learn is that a triangle has three sides; proposing a fourth side is about on the same level as suggesting that two and two are not four. Nevertheless, the fourth 'side' found here, whilst not joining any of the vertices of the triangle, does arise in a natural way from the properties of the triangle. For this to be seen, the sides must be considered from the perspective of certain circles associated with the triangle—the circumcircle, the incircle, and the nine-point circle.

The *nine-point circle* is the circle through the three midpoints of the sides; it also passes through the feet of the altitudes, and the mid-points of the lines joining the orthocentre to the vertices. It is also known as the *Euler circle* or the *Feuerbach circle*.

From the point of view of the three circles, the sides satisfy the following conditions:
a) Each side has its endpoints on the circumcircle;
b) each side has its midpoint on the nine-point circle; and
c) each side touches the incircle.

In looking for line segments that satisfy these conditions, I was amazed to find that, for most triangles, there were not three such segments but four. Three of them were clearly the three sides as we know them; what else was there to call the last but the fourth side of the triangle?

The construction Guinand [2] showed that it was not possible, using compass and straightedge alone, to reconstruct a triangle given its circumcentre O, incentre I, and orthocentre H.

The line OH is known as the *Euler line*; also on it are G, the centroid, one third of the way from O to H, and N, the centre of the nine-point circle, halfway between O and H. If the incentre I is also on it, then the triangle is an isosceles triangle symmetrical about its Euler line.

Given the points O, I, and H (and hence also G and N), it is possible to reconstruct the circumcircle, incircle, and nine-point circle, so that Guinand's result tells us that, if we want to find the triangle from the three circles, we will have use to some instruments other than straightedge and compass.

In order for the three circles to be in those relationships to a triangle, they must be interrelated in certain ways:

(i) The incircle is inside both the circumcircle and the nine-point circle, and touches the nine-point circle.

(ii) Let the circumradius be R, the inradius r, and the distance between the incentre I and the circumcentre O be d; then

$$d^2 = R^2 - 2rR = (R-r)^2 - r^2.$$

(iii) The radius of the nine-point circle is $R/2$.

(See [1], [2], and [3].)

Given a point M on the nine-point circle, and inside the circumcircle, we can easily construct a line segment that satisfies a) and b), and has M as its midpoint (FIGURE 1).

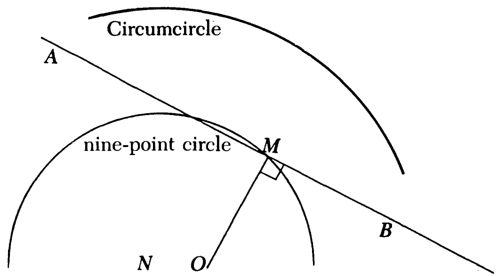


FIGURE 1

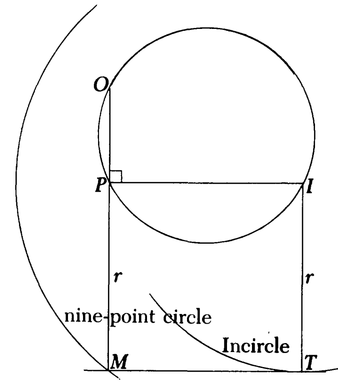


FIGURE 2

Join OM, and construct the line segment AB through M perpendicular to OM, meeting the circumcircle at A and B. If M is O, the circumcentre, then any line segment through it with its endpoints on the circumcircle is satisfactory.

We are looking for points on the nine-point circle for which this line is also tangent to the incircle. For the case when the circumcentre is on the nine-point circle (which happens if and only if the triangle is right-angled), the point O has either two tangents to the incircle through it, or one in the case when O is also the point where the nine-point circle touches the incircle (for which the triangle must also be isosceles).

There are three noteworthy lines perpendicular to a side of a triangle—the altitude, the perpendicular bisector OM, and the line from the incentre I to the tangent point T on the line. Using the last two of these, we form a rectangle three of whose vertices are the incentre I, the tangent point T, and the midpoint M. If we call the fourth vertex P, then (FIGURE 2)

(i) P is on the line OM;

(ii) $PM = r$, the inradius; and

(iii) $\angle OPI = 90°$.

As $\angle OPI$ is a right angle, it must lie on the circle that has OI as a diameter; so we construct this circle.

We thus are trying to find these points M on the nine-point circle such that P, the point other than O on both the circle with diameter OI and the line OM, is distant r from M. (Except that, if OM is tangent to the circle, then P is O.) It is at this point that we use something other than straightedge and compass.

One technique is to mark a point on a ruler, and then two more, each distant r either side of the first mark. The ruler is then placed so that the middle mark is on the circle with diameter OI, and the edge passes through O. It is then moved so that the middle mark traces out the circle, all the time keeping the ruler aligned so that the edge passes through O. Then any points where either of the other two marks pass across the nine-point circle are noted (e.g., J, K, L, and M in FIGURE 3); these will all produce lines which satisfy conditions a), b), and c).

The surprising thing is that, except for isosceles triangles and triangles in which one of the sides is parallel to OI, we get not three but four points from this procedure, and consequently four line segments that satisfy a), b), and c)—the three sides, plus one other. It is this other segment that I call the fourth side of the triangle.

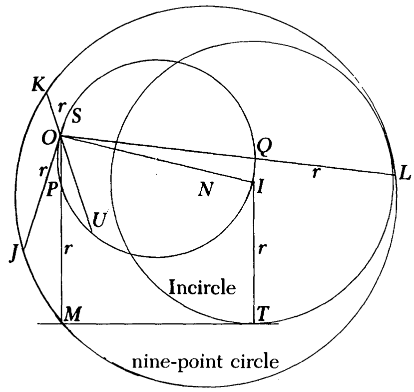


FIGURE 3

The proof In polar coordinates (using (ρ, θ) as coordinates, as r is already being used for the inradius), with origin O, and axis the line OI, the circle with diameter OI has the equation

$$\rho = d \cos \theta .$$

M will be distant r from this circle along OM, and hence on the curve

$$\rho = d \cos \theta \pm r ;$$

i.e.,

$$(\rho - d \cos \theta)^2 = r^2 .$$

This curve is a limaçon (Lawrence, [4]), with Cartesian equation

$$(x^2 + y^2 - dx)^2 = r^2(x^2 + y^2) .$$

M is also on the 9-point circle, which has the equation

$$(x - b)^2 + (x - c)^2 = R^2/4;$$

putting $k = R^2/4 - b^2 - c^2$, this becomes

$$x^2 + y^2 = 2bx + 2cy + k .$$

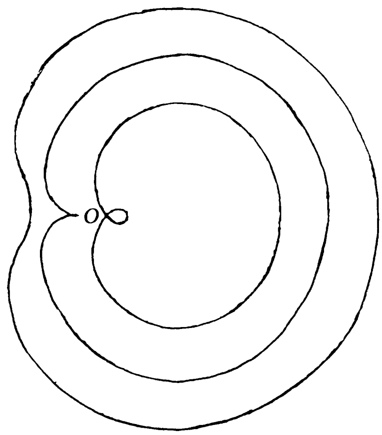
Hence M is also on

$$((2b - d)x + 2cy + k)^2 = r^2(2bx + 2cy + k),$$

which is clearly a parabola.

The circle and the parabola can meet in at most four points; and we already know that the three midpoints of the sides must satisfy the equations. So the limaçon and the nine-point circle have either three or four points of intersection.

The limaçon (FIGURE 4) either has a node at O if O is outside the incircle ($d > r$); a cusp at O if O is on the incircle ($d = r$); or an isolated point at O if O is inside the incircle ($d < r$). In this last case O is not on the curve represented by the original polar equation, having been introduced in the process of converting to Cartesian coordinates; and, in any case, if the nine-point circle passes through O, then the triangle is right-angled, and so O is on or outside the incircle.



Three nested limaçons—the outer, a single loop limaçon ($r < r$); the middle one, a cardioid ($d = r$); the inner one, a two loop limaçon ($d > r$).

FIGURE 4

The incircle has centre I and radius r; let the circle with centre I and radius $R - r$ be the *ring circle*. Then the annulus between these two circles contains both the limaçon and the nine-point circle (FIGURE 5), as the following discussion shows.

Take any point P on the limaçon, and join OP, extending it if necessary to meet the circle with diameter OI at Q. Then $\angle PQI$ is a right angle, PQ is of length r, and IQ is of length less than that of IO, which is d. Using the relation $d^2 = (R - r)^2 - r^2$, we see from the right-angled triangle $\triangle PQI$ that

(i) $PI \geqslant r$; and

(ii) $PI^2 = IQ^2 + PQ^2 \leqslant d^2 + r^2 = (R - r)^2$, so that $PI \leqslant R - r$.

Hence the limaçon lies inside the ring circle, touching it when $IQ = IO = d$, i.e., when $Q = O$, and P is either $(0, r)$ or $(0, -r)$; and outside the incircle, touching it when $PI = r = PQ$, i.e., when $Q = I$ and P is either $(d - r, 0)$ or $(d + r, 0)$.

Also, the incircle lies inside the nine-point circle, and touches it at one point. Consider the diameter of the nine-point circle that passes through this point; it also passes through the incentre I, and is of length R. The point at the other end of the diameter is then the furthest from I on the nine-point circle, and is distant $R - r$ from I—and so the nine-point circle is inside the ring circle, and touches it at one point. Note that the nine-point circle can pass through at most one of the four points $(0, r)$, $(0, -r)$, $(d - r, 0)$, and $(d + r, 0)$, except when the triangle is equilateral, in which case the incircle, the nine-point circle, the ring circle, and the limaçon all coincide.

If we put these four points in order around the limaçon—$(d - r, 0)$, $(0, r)$, $(d + r, 0)$, and $(0, -r)$—and consider the four segments of the limaçon cut off by them, then we see that each segment has one end-point on the incircle and the other on the ring circle; hence the nine-point circle must meet each segment once, with only two difficulties:

(i) It may meet two segments together at the node O—but this is just the right-angled triangle case;

(ii) it may meet two segments together at their common endpoint.

Except in these two cases, the nine-point circle meets the limaçon in four separate points, giving rise to four line segments each of which touches the incircle, has its

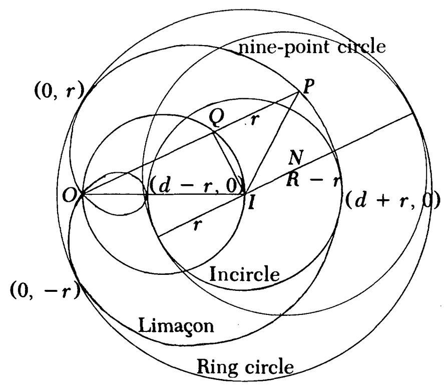


FIGURE 5

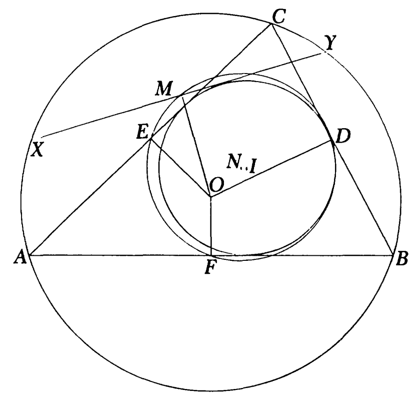


FIGURE 6

endpoints on the circumcircle, and its midpoint on the nine-point circle; three of these are the three regular sides of the triangle, and the other is the fourth side of the triangle.

Finally, we deal with the exceptional cases enumerated above; note that the second is split into two separate cases, depending on which of the four points are involved.

(i) In a right-angled triangle which is not isosceles, the three sides and the second tangent to the incircle from the circumcentre all satisfy the conditions a), b), and c), so that such a triangle also has a fourth side.

(iia) If the nine-point circle passes through either $(d-r, 0)$ or $(d+r, 0)$, then it touches the incircle (and hence also the limaçon) at that point, and so its centre N is on the line OI. But O, N, and H (the orthocentre) are all on the Euler line; and if the incentre I is on the Euler line, then the triangle is isosceles. The isosceles triangles with the nine-point circle through $(d-r, 0)$ are those with vertex angles greater than 60°, and those with it through $(d+r, 0)$ have vertex angle less than 60°. This double point is the midpoint of the base of the triangle, so that we could say that the fourth side coincides with the base. An equilateral triangle has O and I coincident, and incircle and nine-point circle one and the same, so that any line segment tangent to the incircle and with its endpoints on the circumcircle will also have its midpoint on the nine-point circle.

(iib) If, on the other hand, the nine-point circle passes through either $(0, -r)$ or $(0, r)$, then it touches the ring circle, and also the limaçon, at that point. For the side constructed at that point, the perpendicular distance from each of O and I to it is r, and so it is parallel to OI. Thus if a triangle has a side parallel to OI, then its fourth side coincides with that side.

FIGURE 6 shows a triangle ΔABC with its circumcircle (centre O), nine-point circle (centre N), incircle (centre I), midpoints D, E, and F, and its fourth side XY with midpoint M.

REFERENCES

1. H. S. M. Coxeter, and Greitzer, S. L., *Geometry Revisited*, New York, 1967.
2. A. P. Guinand, Triangles from centres out, *James Cook Mathematical Notes*, University of North Queensland 30 (1982), 3127–3132.
3. A. P. Guinand, Euler lines, tritangent centers, and their triangles, *Amer. Math. Monthly* 91 (1984), 290–300.
4. J. D. Lawrence, *A Catalog of Special Plane Curves*, Dover Publications, Mineola, NY, 1972.

PROBLEMS

LOREN C. LARSON, *editor*
St. Olaf College

GEORGE GILBERT, *associate editor*
Texas Christian University

Proposals

To be considered for publication, solutions should be received by May 1, 1994.

1433. *Proposed by Cristian Turcu, London, England.*

For $n \geq 1$, let a_n, b_n, c_n be nonzero integers, and suppose there are indices s and t such that $|a_s c_t - a_t c_s| = 1$. Let P_n denote the set of nonzero rational numbers x such that $\dfrac{a_n x^2 + c_n}{b_n x}$ is an integer. Prove that $\bigcap_{n=1}^{\infty} P_n \neq \emptyset$ if, and only if, b_n divides $a_n + c_n$ for each $n \geq 1$.

1434. *Proposed by Syd Bulman-Fleming and Edward T. H. Wang, Wilfrid Laurier University, Waterloo, Ontario, Canada.*

Let $\mathbf{Z}_n$ denote the ring of integers modulo $n, n > 1$. Consider the function $f\colon \mathbf{Z}_n \to \mathbf{Z}_n$ defined by $f(x) = x^n$ for all $x \in \mathbf{Z}_n$. Characterize those values of n for which f is a ring homomorphism.

1435. *Proposed by Florin S. Pîrvănescu, Slatina, Romania.*

Let $A_1, \ldots, A_n$ be point masses ($n \geq 3$) on the sphere $S(O, R)$ of radius R and center O, and let G be their centroid. Let M be an arbitrary point in the sphere having OG as a diameter, and let B_k be the other intersection of MA_k with the sphere $S(O, R)$. Show that

$$\sum_{k=1}^{n} MB_k \geq \sum_{k=1}^{n} MA_k .$$

 We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals should be accompanied by solutions, if at all possible, and by any other information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution. An asterisk (*) next to a problem number indicates that neither the proposer nor the editors supplied a solution.

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

Solutions and new proposals should be mailed in duplicate to Loren Larson, Department of Mathematics, St. Olaf College, 1520 St. Olaf Ave., Northfield, MN 55057-1098 or mailed electronically via fax: (507) 663-3549 or e-mail: larson@stolaf.edu.

1436. *Proposed by Randall K. Campbell-Wright, University of Tampa, Tampa, Florida.*

Suppose that $(z_n)_{n=1}^{\infty}$ is a bounded sequence of complex numbers and that $\lim_{n\to\infty}\alpha_n = 1$. If k is a positive integer such that $\lim_{n\to\infty}(z_n - \alpha_n z_{n+k}) = \beta$, prove that $\beta = 0$.

1437. *Proposed by Yuanan Diao, Kennesaw State College, Marietta, Georgia.*

Let $\mathbf{Q}[x]$ and $\mathbf{Z}[x]$ be the polynomial rings over the rational numbers and integers respectively. If $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbf{Z}[x]$ satisfies the Eisenstein Criterion, i.e., there is a prime integer p such that $p \nmid a_n, p | a_i, 0 \leq i \leq n-1$, and $p^2 \nmid a_0$, then $f(x)$, hence $af(bx+c)$, is irreducible in $\mathbf{Q}[x]$ for any rational numbers a, b, c, where $ab \neq 0$. Prove or disprove the converse: If $f(x) \in \mathbf{Q}[x]$ is irreducible, then there exist rational numbers $a, b, c, (ab \neq 0)$ such that $af(bx+c) \in \mathbf{Z}[x]$ and satisfies the Eisenstein Criterion.

Quickies

Answers to the Quickies are on page 344.

Q811. *Proposed by Benny N. Cheng, University of California, Santa Barbara, California.*

Show that

$$\int_0^{\alpha} \sqrt{1+\cos^2\theta}\, d\theta > \sqrt{\alpha^2 + \sin^2\alpha},$$

for $0 < \alpha \leq \pi/2$.

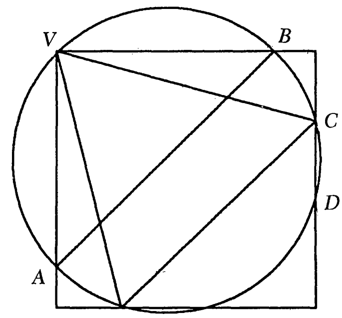
Q812. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

Determine the maximum value of

$$(x+y+z)\left(\sqrt{a^2-x^2} + \sqrt{b^2-y^2} + \sqrt{c^2-z^2}\right).$$

Q813. *Proposed by Edward Kitchen, Santa Monica, California.*

Inscribe an equilateral triangle in a square as shown in the following figure. Let V denote their common vertex, and let A, B, C, D be points on the circumference of the circumcircle of the triangle as indicated. Prove that $BC = CD$, and that AB bisects the circle, trisects the chord VC, and quadrisects the minor arc VC.



Solutions

Cancellative semigroup — December 1992

1408. *Proposed by Syd Bulman-Fleming and Edward T. H. Wang, Wilfrid Laurier University, Waterloo, Ontario, Canada.*

Let S be a commutative, cancellative semigroup and let $P(S)$ denote the semigroup of all nonempty subsets of S under the operation defined by $X \cdot Y = \{xy | x \in X, y \in Y\}$ for all $X, Y \in P(S)$. Characterize all those elements of $P(S)$ that are cancellative. That is, find all $A \in P(S)$ with the property that if $X, Y \in P(S)$ and $A \cdot X = A \cdot Y$, then $X = Y$.

Solution by Robert L. Doucette, McNeese State University, Lake Charles, Louisiana.

The cancellative elements of $P(S)$ are precisely those subsets of S that contain exactly one element.

Suppose that $a \in S$ and that $X, Y \in P(S)$. The equation $\{a\} \cdot X = \{a\} \cdot Y$ implies that for each $x \in X$, $ax = ay$ for some $y \in Y$. Since S is cancellative, $x \in Y$. Hence $X \subseteq Y$. Similarly, $Y \subseteq X$. It follows that $\{a\}$ is cancellative in $P(S)$.

Now suppose that A is cancellative in $P(S)$ and that A contains distinct elements $a, b \in S$. Let $X = A \cdot A$ and $Y = A \cdot A - \{ab\}$. Note that $a^2 \neq ab$ implies that $Y \neq \emptyset$. Since $Y \subseteq X$, it follows that $A \cdot Y \subseteq A \cdot X$. To establish the reverse inclusion let $z \in A \cdot X$. Then z can be written as the product cde where $c, d, e \in A$. If one of the products cd, ce, or de belongs to Y, then by commutativity $z \in A \cdot Y$. If none of cd, ce, or de belongs to Y, then it follows that $cd = ce = de = ab$. In this case the cancellation property leads to $c = d = e$. It follows that $c^2 = ab$, and, since $a \neq b$, that $c \neq a$. We may write $z = c^3 = cc^2 = c(ab) = a(cb)$. Since $c \neq a$ we have $cb \neq ab$. We again see that $z \in A \cdot Y$. Hence $z \in A \cdot X \Rightarrow z \in A \cdot Y$. Therefore $A \cdot X \subseteq A \cdot Y$. We have shown that $A \cdot X = A \cdot Y$. But clearly $X \neq Y$. This shows that if A is cancellative in $P(S)$, then it must contain fewer than two elements, and we are done.

Also solved by David Callan, Con Amore Problem Group (Denmark), Colonel Johnson, Jr., Marta Lewicka (student, Poland), John S. Sumner, Trinity University Problem Group, U.N.A.M. Problem Seminar (Mexico), Edward T. Wong, and the proposers.

Convex lattice pentagon — December 1992

1409. *Proposed by Gerald A. Heuer, Concordia College, Moorhead, Minnesota.*

Does there exist a convex pentagon, all of whose vertices are lattice points in the plane, with no lattice point in the interior? (Cf. 1990 Putnam Competition, A3.)

Solution by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands.

No, a convex lattice pentagon must have an interior lattice point.

To see this, note that every lattice point (x, y) belongs to one of the *four* classes K_{00}, K_{01}, K_{10} and K_{11}, where the index pair ij is determined by taking $i \equiv x$ and $j \equiv y$ modulo 2. A convex lattice pentagon has five vertices, so two of them, say P and Q, belong to the same class, which implies that their center, R, is also a lattice point.

(i) If P and Q are endpoints of a *diagonal*, then R is an interior lattice point.

(ii) If P and Q are endpoints of an *edge*, say edge AB of pentagon $ABCDE$, we continue by considering the convex lattice pentagon $ARCDE$ and so on. Because lattice points have no accumulation point we finally get an interior lattice point, because of (i).

Also solved by David Callan, Con Amore Problem Group (Denmark), Bill Correll, Jr. (student), Bill Doran and Francesco Regonati, Robert L. Doucette, Jiro Fukuta (Japan), Michael Golomb, Charles H.

Jepsen, Marta Lewicka (student, Poland), Reiner Martin (student), Edward D. Onstott, L. P. Pook (England), Allen J. Schwenk, U.N.A.M. Problem Seminar (two solutions, Mexico), and the proposer.

Floor function **December 1992**

1410. *Proposed by Seung-Jin Bang, Seoul, Republic of Korea.*

Prove that $\lfloor n^{1/3}+(n+1)^{1/3}\rfloor=\lfloor(8n+3)^{1/3}\rfloor$ for every positive integer n.

I. *Solution by the Con Amore Problem Group, Royal Danish School of Educational Studies, Copenhagen, Denmark.*

Since the function $y=x^{1/3}$ is concave for $x\geq 0$, for every positive integer n we have

$$\left(n+\frac{1}{2}\right)^{1/3}>\frac{n^{1/3}+(n+1)^{1/3}}{2}>\left(n^{1/3}\cdot(n+1)^{1/3}\right)^{1/2}=\left(n^2+n\right)^{1/6}$$

and consequently,

$$\begin{aligned}(8n+4)^{1/3}>n^{1/3}+(n+1)^{1/3}&>(64n^2+64n)^{1/6}\\ &>(64n^2+48n+9)^{1/6}=(8n+3)^{1/3}.\end{aligned}$$

Moreover, the existence of positive integers n and k such that

$$n^{1/3}+(n+1)^{1/3}\geq k>(8n+3)^{1/3}$$

would imply

$$8n+4>k^3>8n+3$$

which is impossible. We conclude that

$$\left\lfloor n^{1/3}+(n+1)^{1/3}\right\rfloor=\left\lfloor(8n+3)^{1/3}\right\rfloor$$

as claimed.

II. *Solution by David Callan, University of Wisconsin, Madison, Wisconsin.*

We claim that $8n+3$ and $(n^{1/3}+(n+1)^{1/3})^3$ have the same integral part. The desired result follows since $\lfloor a\rfloor=\lfloor b\rfloor$ readily implies $\lfloor a^{1/3}\rfloor=\lfloor b^{1/3}\rfloor$. Now $(n^{1/3}+(n+1)^{1/3})^3=2n+1+3n(1+1/n)^{1/3}+3n(1+1/n)^{2/3}$. For $n\geq 2$, the latter two summands each have a binomial expansion that is alternating, decreasing after the second term. Using the familiar "first omitted term" error estimate, we find $3n(1+1/n)^{1/3}=3n+1-\varepsilon_1$ with $0<\varepsilon_1<1/6$ and $3n(1+1/n)^{2/3}=3n+2-\varepsilon_2$ also with $0<\varepsilon_2<1/6$, and the claim follows.

Also solved by Sinefahopoulos Achilleas (student, Greece), J. C. Binz (Switzerland), W. E. Briggs, Sydney Bulman-Fleming and Edward T. H. Wang (Canada), Stan Byrd and Chris Mewata, Scott H. Brown (student), Bill Correll, Jr. (student), Qais H. Darwish (Oman), David Doster, Robert L. Doucette, Russell Jay Hendel, S. Paul and Julie Hess, Nathan Jaspen, Hans Kappus (Switzerland), Parviz Khajeh-Khalili, Hans Georg Killingbergrø and Ivar Skau (Norway), Satoshi Kondo, Václav Konečný, Kee-Wai Lau (Hong Kong), Peter W. Lindstrom, O. P. Lossers (The Netherlands), Kevin McDougal and Kandasamy Muthuvel, Allen J. Schwenk, Heinz-Jürgen Seiffert (Germany), Glenn A. Stoops, John S. Sumner and Kevin L. Dove, Trinity University Problem Group, University of Wyoming Problem Circle, WMC Problems Group, C. Wildhagen (The Netherlands), and the proposer.

A similar problem, $\lfloor\sqrt[3]{n}+\sqrt[3]{n+1}+\sqrt[3]{n+2}\rfloor=\lfloor\sqrt[3]{27n+26}\rfloor$, proposed by *Mihály Bencze (Braşov, Romania)*, appears as Problem P11 in the *Canadian Mathematical Society Notes* (solution and generalization, Vol. 17, No. 5, 1985, p.8).

Envelope of equichordal lines **December 1992**

1411. *Proposed by Miguel Amengual Covas, Cala Figuera (Santanyí), Mallorca, Spain.*

Let C_1 and C_2 be nonconcentric circles in the plane, and consider the set of lines that intersect C_1 and C_2 in equal chords. Show that these lines are all tangent to a single parabola.

I. *Solution by Jack V. Wales, Jr., The Thacher School, Ojai, California.*

We may assume without loss of generality that the circles are $x^2+y^2=R^2$ and $x^2+(y-a)^2=r^2$, $a\neq 0$, and $R\neq r$. (Note: The property cannot hold if the circles are congruent.) If the line $y=mx+b$ intersects $x^2+y^2=R^2$ at (x_1,y_1) and (x_2,y_2), then the distance $|x_2-x_2|$ is

$$\frac{2\sqrt{m^2R^2-b^2+R^2}}{m^2+1}. \tag{1}$$

For $x^2+(y-a)^2=r^2$, the corresponding distance is

$$\frac{2\sqrt{m^2r^2-(b-a)^2+r^2}}{m^2+1}. \tag{2}$$

The condition that the line intersects the two circles in equal chords dictates that the two distances be equal, so

$$m^2R^2-b^2+R^2=m^2r^2-(b-a)^2+r^2. \tag{3}$$

This reduces to

$$b=\frac{(m^2+1)(R^2-r^2)+a^2}{2a}.$$

We seek, therefore, the envelope of the set of lines

$$y=mx+\frac{(m^2+1)(R^2-r^2)+a^2}{2a} \tag{4}$$

as determined by the parameter m. (Note: All the lines that intersect the two circles in equal chords are in this set, but so are many other lines that do not intersect the circles at all. This is because equation (3) implicitly allows both its right and left sides to be negative, whereas the corresponding expressions in (1) and (2) must be positive.) Differentiating (4) with respect to the parameter m yields

$$m=\frac{-ax}{R^2-r^2},$$

and substituting this into (4) shows that the envelope is the parabola

$$y=-\frac{a}{2(R^2-r^2)}x^2+\frac{R^2-r^2}{2a}+\frac{a}{2}.$$

II. *Solution by Hans Georg Killingbergtrø, Horten, and Ivar Skau, Telemark College, Norway.*

Let F_1 be the center of C_1 and F_2 the center of C_2, and let F bisect F_1F_2. Let t be a line that intersects C_1 at A_1 and B_1, and C_2 at A_2 and B_2, and suppose that the length of A_1B_1 is equal to the length of A_2B_2. Let M_1 bisect A_1B_1, let M_2 bisect

A_2B_2, and let M bisect M_1M_2. Then FM is parallel to F_1M_1 and F_2M_2, and thus FM is perpendicular to t. It is easily seen that $MA_1 \cdot MB_1 = MA_2 \cdot MB_2$, so that the powers of M with respect to C_1 to C_2 are equal. All points with this property are located on a line f perpendicular to F_1F_2. If C_1 and C_2 are not congruent, F is off f. It is well known that a parabola p is completely determined by its focus F and "vertex" tangent f, and that a line t through *any* point M on f is tangent to p if and only if it is perpendicular to FM.

Also solved by Mangho Ahuja (two solutions), P. J. Anderson (Canada), Seung-Jin Bang, Kenneth L. Bernstein, Momcilo Bjelica (Yugoslavia), Con Amore Problem Group (Denmark), Robert L. Doucette, Ragnar Dybvik (Norway), Jiro Fukuta (Japan), R. Daniel Hurwitz, Hans Kappus (Switzerland), Murray S. Klamkin (Canada), Václav Konečný, Marta Lewicka (student, Poland), O. P. Lossers (The Netherlands), U.N.A.M. Problem Seminar (Mexico), University of Wyoming Problem Circle, and the proposer.

Klamkin proved an analogous three-dimensional result: The envelope of the family of planes cutting two non-concentric spheres in pairs of congruent circles is a paraboloid of revolution whose axis is the line joining the centers of the two spheres.

A limit problem — December 1992

1412. *Proposed by Sam Northshield, State University of New York, College at Plattsburgh, Plattsburgh, New York.*

Let $g\colon \mathbf{N} \to \mathbf{R}$ be a bounded function and $f\colon \mathbf{N} \to \mathbf{N}$ a function such that $\lim_{n\to\infty} f(n) = \infty$. Show that if $c > 1$ and $\lim_{n\to\infty}(cg(n) - g(f(n)))$ exists, then $\lim_{n\to\infty} g(n)$ exists.

I. *Solution by Kee-Wai Lau, Hong Kong.*

Since g is bounded, $\lim_{n\to\infty}\sup g(n)$ and $\lim_{n\to\infty}\inf g(n)$ exist.

Let $L = \lim_{n\to\infty}(cg(n) - g(f(n)))$. From

$$cg(n) = \big(cg(n) - g(f(n))\big) + g(f(n))$$

we have

$$c \lim_{n\to\infty} \sup g(n) \le L + \lim_{n\to\infty} \sup g(f(n)) \le L + \lim_{n\to\infty} \sup g(n),$$

or equivalently,

$$\lim_{n\to\infty} \sup g(n) \le \frac{L}{c-1}. \tag{1}$$

In a similar manner,

$$c \lim_{n\to\infty} \inf g(n) \ge L + \lim_{n\to\infty} \inf g(f(n)) \ge L + \lim_{n\to\infty} \inf g(n),$$

or equivalently,

$$\lim_{n\to\infty} \inf g(n) \ge \frac{L}{c-1}. \tag{2}$$

Combining the results of (1) and (2), we deduce that $\lim_{n\to\infty}\sup g(n) = \lim_{n\to\infty}\inf g(n)$, and therefore $\lim_{n\to\infty} g(n)$ exists.

II. *Solution by John S. Sumner, University of Tampa, Tampa, Florida.*

Let $|g(n)| \le M$ for all n and $\lim_{n\to\infty}(cg(n) - g(f(n))) = L$. Then

$$\lim_{n\to\infty} \big(g(n) - g(f(n))/c\big) = L/c.$$

Since $f(n) \to \infty$, we have $\lim_{n\to\infty}\big(g(f(n)) - g(f^2(n))/c\big) = L/c$. Combining the last two equations gives

$$\lim_{n\to\infty}\big(g(n) - g(f^2(n))/c^2\big) = L/c + L/c^2.$$

Inductively, we have

$$\lim_{n\to\infty}\big(g(n) - g(f^k(n))/c^k\big) = L/c + L/c^2 + \cdots + L/c^k \text{ for } k \geq 1.$$

Let $\varepsilon > 0$. Since $c > 1$, we may choose k such that $M/c^k < \varepsilon/3$ and

$$\left|\sum_{i=k+1}^{\infty} L/c^i\right| < \varepsilon/3.$$

Then there exists $N(\varepsilon, k)$ such that

$$\left|g(n) - g(f^k(n))/c^k - \sum_{i=1}^{k} L/c^i\right| < \varepsilon/3 \text{ for } n \geq N(\varepsilon, k).$$

Thus,

$$|g(n) - L/(c-1)| \leq \left|g(n) - g(f^k(n))/c^k - \sum_{i=1}^{k} L/c^i\right| + \left|g(f^k(n))/c^k\right| + \left|\sum_{i=k+1}^{\infty} L/c^i\right| < \varepsilon$$

for $n \geq N(\varepsilon, k)$. This shows that $\lim_{n\to\infty} g(n) = L/(c-1)$.

Also solved by P. J. Anderson (Canada), Michael H. Andreoli, Momcilo Bjelica (Yugoslavia), Con Amore Problem Group (Denmark), Bill Doran and Volkmar Welker, Robert L. Doucette, Herbert Gintis, Hans Georg Killingbergtrø and Ivar Skau (Norway), Gerry Ladas, Ronnie Levy, Marta Lewicka (student, Poland), Peter W. Lindstrom, O. P. Lossers (The Netherlands), Kandasamy Muthuvel, Andreas Müller (Germany), Stephen Noltie, Adam Riese, Heinz-Jürgen Seiffert (Germany), Shreveport Problem Group, Trinity University Problem Group, Tixiang Wang, Chris Wildhagen (The Netherlands), U.N.A.M. Problem Seminar (Mexico), University of Wyoming Problem Circle, and the proposer.

Killigbergtrø and *Skau* give examples to show that the implication fails if $|c| \leq 1$. *Ladas* shows that the conclusion is true when $c < -1$, and points out that a continuous version of the result is true (with $f(t) = t - c$). This result appears in the book *Oscillation Theory of Delay Differential Equations with Applications*, by Gerry Ladas and I. Györi, Clarendon Press, Oxford, 1991. This result is then used in the book to obtain oscillation theorems for neutral delay differential equations.

Answers

Solutions to the Quickies on page 339.

A811. The arclength of the graph of $\sin x$ from $(0,0)$ to $(\alpha, \sin\alpha)$ is greater than the length of the chord joining the two points.

Note: The general result for a C^1 function $f(x)$ with $f(0) = 0$ and $\alpha > 0$, is

$$\int_0^{\alpha} \sqrt{1 + (f'(x))^2}\, dx > \sqrt{\alpha^2 + (f(\alpha))^2}.$$

A812. The problem can be done in several ways using multivariate calculus. Here is a generalization by elementary means.

We determine the maximum value of

$$P = \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} \sqrt{a_i^2 - x_i^2}.$$

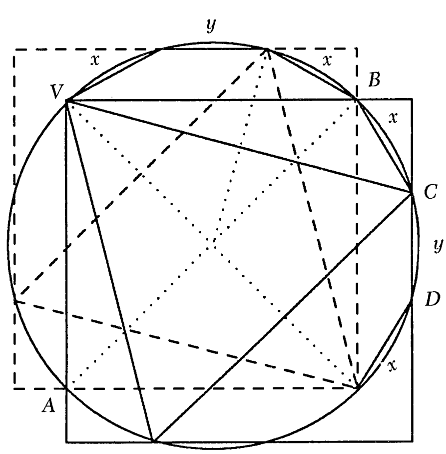
Let $x_i = a_i \sin\theta_i$, $-\pi/2 \le \theta_i \le \pi/2$, $a_i > 0$, so that

$$P = \sum_{i=1}^{n} a_i \sin\theta_i \cdot \sum_{i=1}^{n} a_i \cos\theta_i$$
$$= \frac{1}{2}\left(\sum_{i=1}^{n} a_i^2 \sin 2\theta_i + 2 \sum_{i=1}^{n} \sum_{j=1+1}^{n} a_i a_j \sin(\theta_i + \theta_j) \right).$$

Clearly the maximum is taken on when all $\theta_i = \pi/4$, so that

$$\max P = \frac{1}{2}\left(\sum_{i=1}^{n} a_i \right)^2.$$

A813. Consider the following figure, where the circumference of the circle is 1, and where x and y denote arcs as shown.



It is easy to see that $2x + y = 1/4$, and $3x + y = 1/3$, and together these imply that $x = y = 1/12$.

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 the mathematics literature. Readers are invited to suggest items for review to the editors.

Zorn, Eric, Math riots prove incalculable, *Chicago Tribune* (29 June 1993) Section 2, p. 1.

This newspaper column is an account of an imaginary aftermath of Wiles's proof of Fermat's last theorem, delightfully parodying articles about championship wins by sports teams from some large American cities: "Yes, admittedly, there was rioting and vandalism ... 'Math hooligans are the worst,' said a Chicago Police Department spokesman. 'But we learned from the Bieberbach riots. We were ready for them this time.' ... [T]here was talk of a dynasty, specifically that next year Wiles will crack the great unproven Riemann Hypothesis ('Rie-peat! Rie-peat!' the crowd cried) ...." This piece is funny enough to warrant reprinting in an MAA publication (maybe *Focus*?). There is a further mathematical connection: The author is the grandson of Max Zorn, of Lemma fame, who died earlier this year. (Thanks to Steve Conrad, Roslyn High School, NY, for the reference.)

Horgan, John, The death of proof, *Scientific American* (October 1993) front cover, 92–103.

"[W]as the proof of Fermat's last theorem the last gasp of a dying culture?" Proof has already disappeared from most high-school geometry courses in the U.S.; students find traditional proofs neither as convincing (they can't understand them) nor as relevant (they don't want to) as visual arguments. Calculus reform, under the pressure of client disciplines and the same pragmatic students, may be on its way to reforming proof out of calculus, too. No other science feels it has to prove its results to students, nor do students in those sciences feel insecure as a result. Maybe the era of mathematics-as-proof is drawing to a close anyway: Have all the "good" theorems—ones that are important but also can be proved in under a thousand pages—been exhausted? Will mathematicians be forced to make do with computer proofs, "video proofs," or even just conjectures that have not been falsified after enormous computations? Many mathematicians object to and resent the increasing role of computers in mathematics; after all, Euclid didn't use computers, and Wiles didn't either. Should the recently founded journal *Experimental Mathematics* be known instead as the *Journal of Unproved Theorems*? Will a generation of students brought up without proofs be able to produce a Wiles? This article surveys the deepening controversy, with examples and quotations from all sides.

Reid, Constance, *The Search for E.T. Bell, Also Known as John Taine*, MAA, 1993; x + 372 pp, $35. ISBN 0–88385–508–9

Many mathematicians know of E.T. Bell (1883–1960) from his *Men of Mathematics* (1937), *Mathematics, Queen and Servant of the Sciences* (1951), and *The Last Problem* (1961). They may not know that he was an acclaimed number theorist (200 papers) or that he was the author behind the name of John Taine, under which name he published pioneering science fiction novels (*The Time Stream Before the Dawn*, and *The Crystal Horde*). In preparing a memorial essay, Constance Reid—famous for her careful and readable biographies of mathematicians—discovered inconsistencies in accounts of his background. Possessed by the question, "Why did he falsify his past?", she recounts six years of investigation that led to surprising revelations about this unusual and many-sided man.

Gleick, James, Fermat's theorem, *New York Times Magazine* (3 October 1993) 52–53.

What is a science writer to make of Wiles's proof of Fermat's Last Theorem? "A half-dozen or so people on earth have some hope of understanding Wiles's proof from start to finish, and most of those people are now hard at work on it as referees. ... As the words wash over you, you can almost feel that the mathematicians are speaking English. ... It's slightly heartwarming to discover that your normal everyday mathematician also has trouble following this." And what of the 350 years that it took to prove the theorem? "Whole branches of mathematics have been born of failed proofs of Fermat's Last Theorem."

Forrest, Stephanie, Genetic algorithms: Principles of natural selection applied to computation, *Science* 261 (13 August 1993) 872–878.

More so than other popular sources, this article delves into more of the difficult details of how to represent the search space and why genetic algorithms work, including mathematical models of their behavior.

Pinder, Jeanne B., Fuzzy thinking has merits when it comes to elevators, *New York Times* (National Edition) (22 September 1993) C1, C7.

The first elevator employing fuzzy logic, manufactured by a U.S. company, is being installed in Osaka; it should be able to handle more people with less waiting than standard elevators. What doesn't come out in press reports about adapting machinery to fuzzy logic is that the improvements are accomplished by mathematicians, engineers, and computer scientists who *devise new algorithms*. Meanwhile, in an adjoining article on the same page, we find that in Japan—where trains and rice cookers have been "fuzzified"—elevator manufacturers are concentrating not on implementing fuzzy logic, nor on minimizing waiting time, but on minimizing transit time, with elevators that now travel almost 30 mph.

Kosko, Bart, and Satoru Isaka, Fuzzy logic, *Scientific American* (July 1993) 76–81.

This article explains how engineers go from a fuzzy-logic characterization of a situation to an actual product. The rules of a fuzzy system come from an expert system, sometimes based on a neural network. Current and planned applications are mentioned; I didn't know that the General Motors Saturn car uses fuzzy logic in its transmission.

von Bechtolsheim, Stephan, *TEX in Practice.* 4 vols. Vol. 1: *Basics.* xl + 386 pp. Vol. 2: *Paragraphs, Math, and Fonts.* xl + 368 pp. Vol. 3: *Tokens, Macros.* xlvi + 656 pp. Vol. 4: *Output Routines, Tables.* xxxvi + 422 pp. Springer-Verlag, 1993, $169. ISBN 0–387–97296–X

TEX is the standard for typesetting mathematics and will continue to be the standard for a long time to come. But it is one thing to use TEX (as I have for some years) and another to understand how TEX works (despite all the use, I have a very limited understanding). To increase your proficiency, there are several primers. To increase your understanding, you could read and work the exercises in Knuth's *The TEX book*; better yet, you could read the impressive TEX program itself, in Knuth's *TEX : The Program.* The first is somewhat dense, and the second mixes how TEX works with how TEX is implemented (once I understand how TEX works, I can begin to be curious about how the program makes it work). Now there is a leisurely treatment of details, roughly three times the length of *The TEXbook.* This four-volume work is for you, if you like to understand the tools that you use, if you learn better by example (there are lots), if you have reached the point of wanting to write your own macros, or if you are tired of not having text come out as you want because you you have a mistaken mental model of what TEX does. Thanks to these volumes, TEX will be a little less mysterious to its users.

NEWS AND LETTERS

ACKNOWLEDGEMENTS

In addition to our associate editors, the following have assisted the MAGAZINE *by refereeing papers during the past year. We appreciate the time and care they have given.*

Alexander, James C., *University of Maryland, College Park*
Alexanderson, Gerald L., *Santa Clara University*
Aminzadeh, Mostafa, *Towson State University*
Andrews, George E., *Pennsylvania State University, University Park*
Archibald, Thomas, *Acadia University*
Askey, Richard, *University of Wisconsin, Madison*
Bailey, Donald F., *Trinity University*
Bandy, C., *Southwest Texas State University*
Barbeau, Edward J., *University of Toronto*
Baxley, John V., *Wake Forest University*
Bivens, Irl C., *Davidson College*
Bloch, Norman, *SUNY College at Brockport*
Boas, Mary L., *Seattle, WA*
Booth, Ada, *Santa Clara University*
Borwein, Jonathan M., *University of Waterloo*
Comfort, Wistar, *Wesleyan University*
Chollet, John, *Towson State University*
Connelly, Robert, *Cornell University*
Cundy, H. Martyn, *Kendal, Cumbria, England*
Dancis, Jerry, *University of Maryland, College Park*
Dodge, Clayton, *University of Maine*
Drobot, Vlad, *San Jose State University*
Drucker, Daniel, *Wayne State University*
Dunham, William W., *Muhlenberg College*
Eddy, R.H., *Memorial University of Newfoundland*
Elenbogen, Bruce S., *University of Michigan-Duluth*
Faires, J. Douglas, *Youngstown University*
Fisher, J. Chris, *University of Regina*
Fleron, Julian, *SUNY at Albany*
Franzblau, Debbie, *Rutgers University*
Gardner, Martin, *Hendersonville, NC*
Ghahramani, Saeed, *Towson State University*
Gillman, Leonard, *University of Texas at Austin*
Gold, Bonnie, *Wabash College*
Goldstein, Larry J., *Goldstein Software, Brookeville, MD*
Goodson, Geoffrey, *Towson State University*
Gray, Mary, *American University*
Guy, Richard, *University of Calgary*
Hafner, Everett, *Williamsburg, MA*
Hall, Leon, *University of Missouri at Rolla*
Halmos, Paul R., *Santa Clara University*
Horak, Martin, *Towson State University*
Horn, Roger A., *The University of Utah*
Hutchinson, Joan, *Macalester College*
Ilgenfritz, Ernie, *Towson State University*
Jones, Sharon, *Towson State University*
Kaplan, Gail, *Annapolis, MD*
Kaplon, Howard, *Towson State University*
Kastner, Bernice, *Greenbelt, MD*
Keller, Edward L., *California State University, Hayward*
Kimberling, Clark H., *University of Evansville*
Klamkin, Murray S., *University of Alberta*
Kleiner, Israel, *York University*
Krieger, Henry, *Harvey Mudd College*
Kung, Sidney, *Jacksonville University*
Lagarias, Jeffrey C., *AT&T Bell Labs, Murray Hill, NJ*
Lax, Peter D., *New York University-Courant*
Leinbach, Carl, *Cornell University*
Lynch, Mark, *Millsaps College*
Markham, Elizabeth, *St. Joseph's College*
May, Coy, *Towson State University*
Merrill, Kathy, *Colorado College*
Moore, Thomas L., *Grinnell College*
Naiman, Daniel Q., *The Johns Hopkins University*
Nelsen, Roger B., *Lewis & Clark College*
Nunemacher, Jeffrey L., *Ohio Wesleyan University*
O'Neill, Barrett, *Pacific Palisades, CA*
Park, Efton, *Texas Christian University*
Pedersen, Jean, *Santa Clara University*
Pedoe, Daniel, *University of Minnesota*
Peeples, Joanne V., *El Paso Community College*
Raimi, Ralph, *University of Rochester*
Reiter, Harold, *University of North Carolina at Charlotte*
Ross, Sharon C., *DeKalb Community College*
Samuels, Stephen M., *Purdue University*
Schattschneider, Doris M., *Moravian College*
Scully, Erik, *Towson State University*
Segal, Sanford, *University of Rochester*
Smith, David A., *Duke University*
Sohrab, Houshang, *Towson State University*
Stern, Leonid, *Towson State University*
Stone, Alexander P., *University of New Mexico*
Strang, Gilbert, *Massachusetts Institute of Technology*
Stromquist, Walter, *Daniel H. Wagner Associates, Berwyn, PA*
Tanis, Elliot A., *Hope College*
Townsend, Douglas A., *Indiana-Purdue University at Fort Wayne*
Tucker, Alan, *SUNY at Stony Brook*
Tucker, Thomas W., *Colgate University*
Uhl, Jerry, *University of Illinois*
Wessner, John, *Towson State University*
Wilf, Herbert, *University of Pennsylvania, Philadelphia*
Williams, Kenneth S., *Carleton University*
Wolfskill, John, *Northern Illinois University*
Zimmerman, Jay, *Towson State University*

CONTENTS

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