INDEX

AUTHORS

Auer, Jan W., *Mathematical Preliminaries to Elementary Catastrophe Theory,* 13-20.

Bar-Hillel, Maya, see Falk, Ruma.

Beer, Gerald, *The Cobb-Douglas Production Function,* 44-48.

Berresford, Geoffrey C., *The Uniformity Assumption in the Birthday Problem,* 286-288.

Bolis, Theodore S., *Degenerate Critical Points,* 294-299.

Brams, Steven J., *Mathematics and Theology: Game-Theoretic Implications of God's Omniscience,* 277-282.

Brodie, Scott E., *Archimedes' Axioms for Arc-Length and Area,* 36-39.

Bruggeman, Tom, and Gush, Tom, *Nice Cubic Polynomials for Curve Sketching,* 233-234.

Collison, Mary Joan, *The Unique Factorization Theorem: From Euclid to Gauss,* 96-100.

Davis, Elwyn H., *Cauchy and Economic Cobwebs,* 171-173.

Falk, Ruma, and Bar-Hillel, Maya, *Magic Possibilities of the Weighted Average,* 106-107.

Fearnley-Sander, Desmond, *A Royal Road to Geometry,* 259-268.

Fisk, Robert, see Ruebush, John.

Fraenkel, Aviezri S., and Herda, Hans, *Never Rush to Be First in Playing Nimbi,* 21-26.

Frauenthal, James C., and Miller, Andrew B., *A Coin Tossing Game,* 239-243.

Friedlander, John B., and Wilker, John B., *A Budget of Boxes,* 282-286.

Gallian, Joseph A., and Rusin, David J., *Factoring Groups of Integers Modulo n,* 33-36.

Golomb, Solomon W., *On the Enumeration of Cryptograms,* 219-221.

Grünbaum, Branko, and Shephard, Geoffrey C., *Satins and Twills: An Introduction to the Geometry of Fabrics,* 139-161.

Gush, Tom, see Bruggeman, Tom.

Harary, Frank, and Kabell, Jerald A., *An Intuitive Approach to Interval Numbers of Graphs,* 39-44.

Herda, Hans, see Fraenkel, Aviezri S.

Kabell, Jerald A., see Harary, Frank.

Kasper, Toni, *Integration in Finite Terms: The Liouville Theory,* 195-201.

Kleiner, Alex, *A Weighted Voting Model,* 28-32.

Koehler, Don O., *The Cost Accounting Problem,* 3-12.

Levinger, Bernard W., *The Square Root of a 2x2 Matrix,* 222-224.

Lewis, Ted, and Willard, Stephen, *The Rotating Table,* 174-179.

Loyer, Milton W., *Equally Likely, Exhaustive and Independent Events,* 167-171.

Maurer, Stephen B., *The King Chicken Theorems,* 67-80.

Miller, Andrew B., see Frauenthal, James C.

Patashnik, Oren, *Qubic: 4x4x4 Tic-Tac-Toe,* 202-216.

Richards, Dana, *Generalized Magic Cubes,* 101-105.

Rickey, V. Frederick, and Tuchinsky, Philip M., *An Application of Geography to Mathematics: History of the Integral of the Secant,* 162-166.

Rogers, Kenneth, *Solving an Exponential Equation,* 26-28.

Ruberg, Stephen J., *The Tree Planting Problem on a Sphere,* 235-237.

Ruebush, John, and Fisk, Robert, *Cooking a Turkey,* 237-239.

Rusin, David J., see Gallian, Joseph A.

Sanderson, Donald E., *Advanced Plane Topology From an Elementary Standpoint,* 81-89.

Schattschneider, Doris, *Will it Tile? Try the Conway Criterion!,* 224-233.

Shephard, Geoffrey C., see Grünbaum, Branko.

Silvester, John R., *A Matrix Method for Solving Linear Congruences,* 90-92.

Straffin, Philip D., Jr., *Linear Algebra in Geography: Eigenvectors of*

Networks, 269-276.
Trigg, Charles W., *A Family of Ninth Order Magic Squares*, 100-101.
Tuchinsky, Philip M., see Rickey, V. Frederick.
Turner, Barbara, *Fourier's Seventeen Lines Problem*, 217-219.
———, *Sums of Powers of Integers via the Binomial Theorem*, 92-96.
Tymoczko, Thomas, *Computers, Proofs and Mathematicians: A Philosophical Investigation of the Four-Color Proof*, 131-138.
Ward, James E., III, *Vector Spaces of Magic Squares*, 108-111.
Wilker, John B., see Friedlander, John B.
Willard, Stephen, see Lewis, Ted.
Zage, Wayne M., *The Geometry of Binocular Visual Space*, 289-294.

TITLES

Advanced Plane Topology From an Elementary Standpoint, *Donald E. Sanderson*, 81-89.
Application of Geography to Mathematics: History of the Integral of the Secant, An, *V. Frederick Rickey and Philip M. Tuchinsky*, 162-166.
Archimedes' Axioms for Arc-Length and Area, *Scott E. Brodie*, 36-39.
Budget of Boxes, A, *John B. Friedlander and John B. Wilker*, 282-286.
Cauchy and Economic Cobwebs, *Elwyn H. Davis*, 171-173.
Cobb-Doublas Production Function, The, *Gerald Beer*, 44-48.
Coin Tossing Game, A, *James C. Frauenthal and Andrew B. Miller*, 239-243.
Computers, Proofs and Mathematicians: A Philosophical Investigation of the Four-Color Proof, *Thomas Tymoczko*, 131-138.
Cooking a Turkey, *John Ruebush and Robert Fisk*, 237-239.
Cost Accounting Problem, The, *Don O. Koehler*, 3-12.
Degenerate Critical Points, *Theodore S. Bolis*, 294-299.
Enumeration of Cryptograms, On the, *Solomon W. Golomb*, 219-221.
Equally Likely, Exhaustive and Independent Events, *Milton W. Loyer*, 167-171.
Factoring Groups of Integers Modulo n, *Joseph A. Gallian and David J. Rusin*, 33-36.
Family of Ninth Order Magic Squares, A, *Charles W. Trigg*, 100-101.
Fourier's Seventeen Lines Problem, *Barbara Turner*, 217-219.
Generalized Magic Cubes, *Dana Richards*, 101-105.
Geometry of Binocular Visual Space, The, *Wayne M. Zage*, 289-294.
Integration in Finite Terms: The Liouville Theory, *Toni Kasper*, 195-201.
Intuitive Approach to Interval Numbers of Graphs, An, *Frank Harary and Jerald A. Kabell*, 39-44.
King Chicken Theorems, The, *Stephen B. Maurer*, 67-80.
Linear Algebra in Geography: Eigenvectors of Networks, *Philip D. Straffin, Jr.*, 269-276.
Magic Possibilities of the Weighted Average, *Ruma Falk and Maya Bar-Hillel*, 106-107.
Mathematical Preliminaries to Elementary Catastrophe Theory, *Jan W. Auer*, 13-20.
Mathematics and Theology: Game-Theoretic Implications of God's Omniscience, *Steven J. Brams*, 277-282.
Matrix Method for Solving Linear Congruences, A, *John R. Silvester*, 90-92.
Never Rush to Be First in Playing Nimbi, *Aviezri S. Fraenkel and Hans Herda*, 21-26.
Nice Cubic Polynomials for Curve Sketching, *Tom Bruggeman and Tom Gush*, 233-234.
Qubic: 4x4x4 Tic-Tac-Toe, *Oren Patashnik*, 202-216.
Rotating Table, The, *Ted Lewis and Stephen Willard*, 174-179.
Royal Road to Geometry, A, *Desmond Fearnley-Sander*, 259-268.
Satins and Twills: An Introduction to the Geometry of Fabrics, *Branko Grünbaum and Geoffrey C. Shephard*, 139-161.
Solving an Exponential Equation, *Kenneth Rogers*, 26-28.
Square Root of a 2x2 Matrix, The, *Bernard W. Levinger*, 222-224.
Sums of Powers of Integers via the Binomial Theorem, *Barbara Turner*, 92-96.
Tree Planting Problem on a Sphere, The, *Stephen J. Ruberg*, 235-237.
Uniformity Assumption in the Birthday Problem, The, *Geoffrey C. Berresford*, 286-288.

Unique Factorization Theorem: From Euclid to Gauss, The, *Mary Joan Collison,* 96-100.
Vector Spaces of Magic Squares, *James E. Ward III,* 108-111.
Weighted Voting Model, A, *Alex Kleiner,* 28-32.
Will It Tile? Try the Conway Criterion!, *Doris Schattschneider,* 224-233.

PROBLEMS

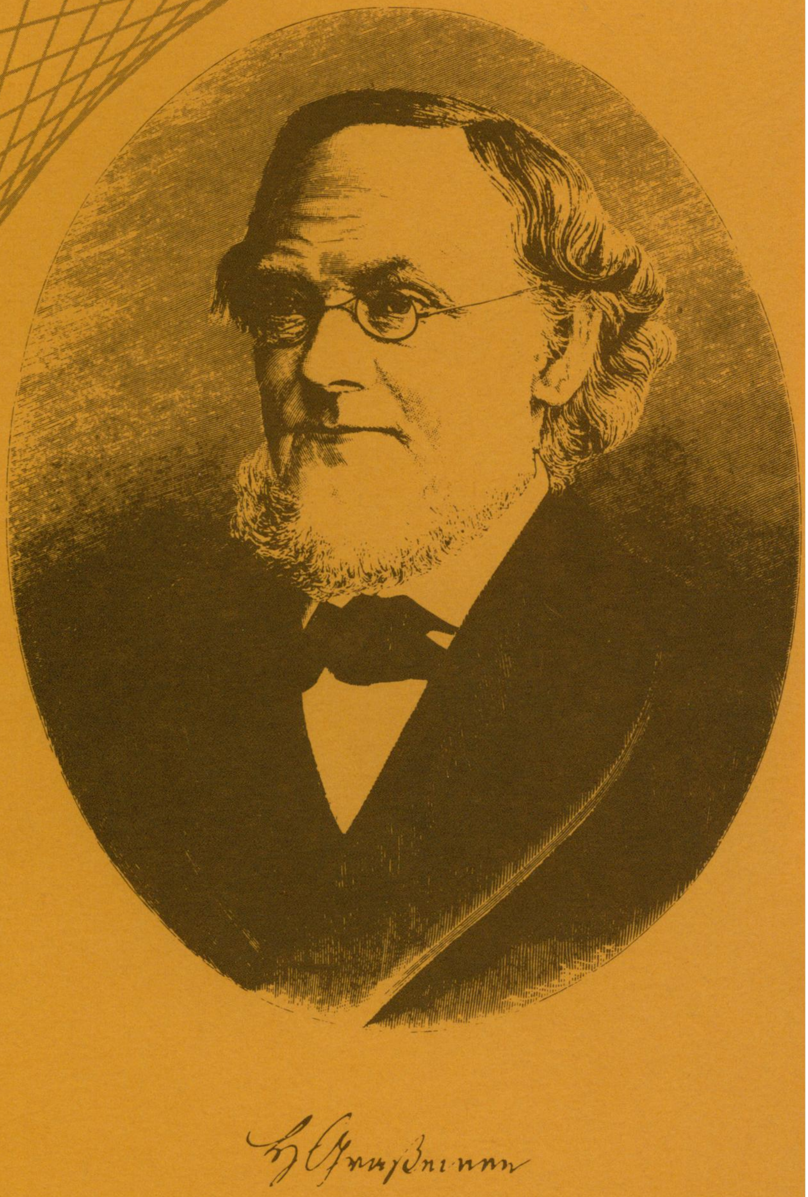
Proposals, Solutions and Quickies are indexed below by means of the code letters P, S, *and* Q, *respectively. Page numbers are given in parenthesis. Thus,* P1099(180) *refers to proposal number* 1099 *which appears on page* 180.

Ampe, John, S1053(51).
Anon., P1099(180).
Apostol, Tom M., P1090(49), P1091(49), S1064(183).
Beigel, Richard, S1083(303).
Bluger, Walter, S1057(113).
Bosch, William, S1069(245).
Bracken, Paul, S1077(250).
Brackett, A. Dickson, S1056(54).
Broline, Duane, S1074(248).
Clark, Robert, S1053(51), S1060(116).
Collison, D.M., S1057(114).
Creech, Roger L., P1092(49).
Ecker, Michael W., P1106(244).
Edgar, G.A., S1062(117).
Elliott, Donald, S1069(245).
Eves, Howard, S1054(52).
Finn, Michael, S1063(181).
Fisher, Gordon, S1081(302).
Forsey, Hal, P1103(244).
Fox, William F., S1052(50).
Franceschine, Nick, P1104(244).
Gardner, C.S., S1082(302).
Good, I.J., P1097(180).
Heuer, G.A., S1071(247), S1074(248), S1075(249), S1080(302).
Heuer, Karl, S1080(302).
Jensen, D.R., P1097(180).
Johnson, B.R., S1070(245).
Kappus, Hans, S1068(186), S1087(304).
Kestelman, H., S1058(114).
Klamkin, M.S., P1089(49), P1093(112), P1100(180), P1107(300), Q666(301).
Kleiner, David, S1066(184).
Lewan, Douglas, P1111(300).
Linis, Viktors, S1068(186).
Moore, Tom, P1094(112).
Moran, Dan, Q665(245).
Motteler, Zane C., S1051(50).
Nelsen, Roger B., S1074(248).
Nicely, Thomas R., P1109(300).
Noland, Hugh, S1079(301).
Ørno, Peter, P1096(180), P1098(180), P1105(244).
Propp, James, P1108(300).
Shapiro, Daniel B., P1101(244), P1102 (244).
Singmaster, David, S1059(115).
Stark, J.M., S1057(113), S1067(185).
Subbarao, M.V., P1100(180).
Todd, Philip, S1055(53).
Underwood, Douglas, S1086(304).
U. of Hartford Problem Solving Group, S1072(247), S1078(251), S1085(303).
Valdez, Eduardo, S1057(113).
Wang, Edward T.H., S1061(116).
Weening, Joseph S., S1065(184).
Wildenberg, Gerald, P1110(300).
Yeshurun, S., P1095(112).
Yeung, Imelda, S1076(249).
Yocom, Ken, S1077(250).
Young, Robert L., S1064(181).

NEWS

Acknowledgements, 315.
Allendoerfer, Ford, Pólya Awards, 255.
Conference on Undergraduate Mathematics, 60, 310.
ICME IV: The International Congress of Mathematical Education, 120.
Index, 316.
Mathematics Teachers Urge Reform of School Mathematics Programs, 190.
National Educational Computing Conference, 191.
1979 Putnam Exam Solutions, 122.
1979 William Lowell Putnam Mathematical Competition, 61.
1980 Chauvenet Prize, 59.
1980 Distinguished Service Award, 59.
1980 NSF-CBMS Regional Research Conferences, 192.
1980 U.S.A. Mathematical Olympiad, 191.
1980 U.S.A. Mathematical Olympiad Solutions, 311.
1981 Short Course on Numerical Linear Algebra, 310.
Olympiad News, 310.
Pi Mu Epsilon Student Conference, 191.
Pulitzer Prize, 190.
Reviews, 55, 118, 187, 252, 306.
Statistics Conference, 60.
Undergraduate Research Programs, 120.

MATHEMATICS MAGAZINE



Vol. 53, No. 5
November, 1980

BOXING IT ALL IN • ALGEBRA IN GEOMETRY AND GEOGRAPHY
THE PARADOX OF OMNISCIENCE • DEALING WITH DEGENERACY

MATHEMATICS MAGAZINE

Volume 53, Number 5, November 1980
(ISSN—0025-570X)

COVER: Hermann Günther Grassmann (1809–77) published a highly original but nearly unintelligible treatise on linear algebra and geometry in 1844. For a readable introduction to his ideas, see p. 259.

ARTICLES

259 A Royal Road to Geometry, *by Desmond Fearnley-Sander.*

269 Linear Algebra in Geography: Eigenvectors of Networks, *by Philip D. Straffin, Jr.*

NOTES

277 Mathematics and Theology: Game-Theoretic Implications of God's Omniscience, *by Steven J. Brams.*

282 A Budget of Boxes, *by John B. Friedlander and John B. Wilker.*

286 The Uniformity Assumption in the Birthday Problem, *by Geoffrey C. Berresford.*

289 The Geometry of Binocular Visual Space, *by Wayne M. Zage.*

294 Degenerate Critical Points, *by Theodore S. Bolis.*

PROBLEMS

300 Proposals Number 1107–1111.

301 Quickies Number Q666.

301 Solutions to Problems 1079–1083, 1085–1087.

305 Answers to Quickies.

REVIEWS

306 Reviews of recent books and expository articles.

NEWS AND LETTERS

310 Comments on recent issues; answers to problems from the 1980 USA Mathematical Olympiad; acknowledgement of reviewers.

INDEX

316 Authors, Titles, Problems, News

EDITORIAL POLICY

Mathematics Magazine is a journal of collegiate mathematics designed to enrich undergraduate study of the mathematical sciences. The *Magazine* should be an inviting, informal journal emphasizing good mathematical exposition of interest to undergraduate students. Manuscripts accepted for publication in the *Magazine* should be written in a clear and lively expository style. The *Magazine* is not a research journal, so papers written in the terse "theorem-proof-corollary-remark" style will ordinarily be unsuitable for publication. Articles printed in the *Magazine* should be of a quality and level that makes it realistic for teachers to use them to supplement their regular courses. The editors especially invite manuscripts that provide insight into applications and history of mathematics. We welcome other informal contributions, for example, brief notes, mathematical games, graphics and humor.

New manuscripts should be sent to: Doris Schattschneider, Editor-Elect, Mathematics Magazine, Department of Mathematics, Moravian College, Bethlehem, Pennsylvania 18018. Manuscripts should be prepared in a style consistent with the format of *Mathematics Magazine*. They should be typewritten and double spaced on $8\frac{1}{2}$ by 11 paper. Authors should submit the original and one copy and keep one copy as protection against possible loss. Illustrations should be carefully prepared on separate sheets of paper in black ink, the original without lettering and two copies with lettering added; the printers will insert printed letters on the illustration in the appropriate locations.

ABOUT OUR AUTHORS

Desmond Fearnley-Sander ("A Royal Road to Geometry") is a Senior Lecturer in Mathematics at the University of Tasmania. With J.P.O. Silberstein of the University of Western Australia he is translating Hermann Grassmann's *Ausdehnungslehre* of 1862. It was his interest in Grassmann's work which gave rise to the present paper, and to an historical appraisal ("Hermann Grassmann and the Creation of Linear Algebra") which appeared in the *American Mathematical Monthly*, December, 1979. He is at present working on the application of Grassmann's ideas to calculus on manifolds, and hopes that he "may help to bring closer the day when Grassmann's vision of a universal geometric algebra is fully realized."

Philip D. Straffin, Jr. ("Linear Algebra in Geography: Eigenvectors of Networks") is chairman of the mathematics department at Beloit College. He holds a B.A. from Harvard, an M.A. from Cambridge, and a Ph.D. in algebraic topology from Berkeley. He recently wrote a UMAP monograph on *Topics in the Theory of Voting*. The present article is a version of a talk given to the Wisconsin MAA workshop on Linear Algebra and its Applications in 1978. It grew out of a general interest in finding serious applications of mathematics in other disciplines which can be exciting to students in mathematics courses.

BUSINESS INFORMATION. Mathematics Magazine is published by the Mathematical Association of America at Washington, D.C., five times a year in January, March, May, September, and November. Ordinary subscriptions are $18 per year. Members of the Mathematical Association of America or of Mu Alpha Theta may subscribe at special reduced rates. Colleges and university mathematics departments may purchase bulk subscriptions (5 or more copies to a single address) for distribution to undergraduate students.

Subscription correspondence and notice of change of address should be sent to A. B. Willcox, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Back issues may be purchased, when in print, from P. and H. Bliss Co., Middletown, Connecticut 06457.

Advertising correspondence should be addressed to Raoul Hailpern, Mathematical Association of America, SUNY at Buffalo, Buffalo, New York 14214.

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

A Royal Road to Geometry

An introduction to Grassmann's affine geometry, an algebraic formalism which offers advantages of both the synthetic and analytic approaches.

DESMOND FEARNLEY-SANDER
University of Tasmania
Hobart, Tasmania, Australia 7001

Geometry is out of fashion. One can think of many reasons why this has come about. We have realized, a century after it was demonstrated, that Euclid is inadequate; and we have come to believe that a rigorous development of synthetic geometry must necessarily be long and tedious. We find an apparent superiority in analytic geometry, not realizing that a complete rigorous development here would be equally long and tedious and that much of what is called analytic geometry is not geometry at all, though it uses geometric language. Perhaps most significantly we have felt revulsion from a subject and a teaching method that became stultified over the centuries. And all the while, paradoxically, geometric thinking has come to pervade mathematics more and more deeply.

This paper deals with the part of geometry which is called affine geometry. From a traditional constructive point of view, it is that part in which we are allowed the use of an infinite ruler (or rather, finite rulers of arbitrary length) marked with real numbers, so that we may compare distances along any line, though we are not permitted to compare distances along distinct lines. However, as we shall see, it is in fact possible to compare distances along parallel lines. About a quarter of the theorems of Euclid belong to affine geometry. Affine geometry is in a sense the foundation on which Euclidean geometry rests, since it is only necessary to bring in a suitably defined general distance function (allowing the lengths of non-parallel line segments to be compared) to obtain the whole grand structure.

According to Proclus [**9**], Euclid was asked by King Ptolemy I if there is not a shorter road to geometry than through the Elements. Euclid replied that "there is no royal road to geometry." We are going to develop some elementary affine geometry in three dimensions from the point of view of Hermann Grassmann [**4**]. We suggest that here indeed is a royal road, combining, as it does, the advantages of the traditional synthetic and analytic approaches and minimizing their disadvantages. The exposition is necessarily terse, but may easily be adapted to the gradual intuitive development which is needed for a beginner.

We start with an informal presentation of the axiomatics approach based on Grassmann [**4**]. The body of the paper is devoted to an interpretation of the axioms in terms of the usual intuition of "point," "line," "plane," "vector," and "between." We include a number of examples of the utility of this particular algebraic formulation of affine geometry and close with a sketch of alternative approaches.

Algebraic Formalism

We denote **real numbers** by lower case letters $a, b, c, \ldots$ and **points** by capitals $A, B, C, \ldots$ (except that V, W, X, and Z will not stand for points). Of course points are different from numbers. We permit operations of addition, subtraction, and multiplication of points and numbers obtaining new symbols such as $(3ABC+2AB)(AD-C)$.The ordinary rules of algebra (associativity of addition and of multiplication, commutativity of addition and distributivity; also multiplication of a number and a point is commutative) are applicable except that instead of multiplication of points being commutative, we assume that for all points A and B, $BA=-AB$, and, in particular, $A^2=0$. Thus for example

$$\begin{aligned}(3ABC+2AB)(AD-C) &= 3ABC(AD-C)+2AB(AD-C)\\ &= 3ABCAD-3AVC^2+2ABAD-2ABC\\ &= 3A^2BCD-2A^2BD-2ABC\\ &= -2ABC.\end{aligned}$$

We also assume explicitly that there exist four points, A, B, C, and D, such that $ABCD \neq 0$. Treated as a formal matter the business of simplifying expressions involving sums and products of points and numbers is easily grasped.

Geometric Interpretation

All geometric interpretations of entities and operations follow step-by-step from a single fundamental one: if A and B are points and a and b positive real numbers with $a+b=1$, then $aA+bB$ is taken to be the point P which divides the line segment from A to B in the ratio b to a (as in FIGURE 1). (Formally, we are assuming that the set of all points is closed under affine combinations.) An interpretation of $P=aA+bB$ with $a>0$ and $b<0$ and $a+b=1$ is now forced by the fact that equivalently, $B=(-a/b)A+(1/b)P$ where $a/b>0$, $1/b>0$ and $(-a/b)+1/b=1$.

If P is a point, then $1\cdot P=P$, a point, while $0\cdot P=0$, a number; any other multiple of P is neither a point nor a number (for if $a\neq 1$ and aP was Q, a point, we would have a point

$$\frac{a}{a-1}P-\frac{1}{a-1}Q=0,$$

while if $a\neq 0$ and aP was a number b, we would have $P=b/a$). We do not need to interpret aP for $a\neq 0$ or 1.

If A and B are points, and $a+b\neq 0$, then

$$aA+bB=(a+b)\left(\frac{a}{a+b}A+\frac{b}{a+b}B\right)=(a+b)P,$$

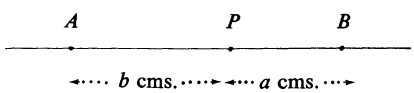


FIGURE 1.

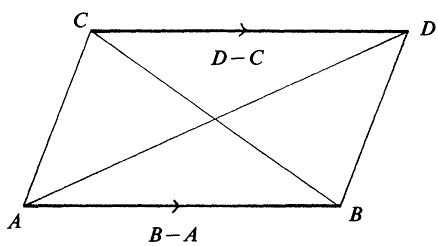


FIGURE 2.

a multiple of a point, P. For points A and B we call $B-A$ the **vector** from A to B (see FIGURE 2) since

$$B-A=D-C \Leftrightarrow \tfrac{1}{2}(B+C)=\tfrac{1}{2}(A+D),$$

we interpret equality of the vector from A to B with the vector from C to D as meaning that A, B, D and C are the vertices of a parallelogram, or, equivalently, that the directed line segments from A to B and from C to D have the same length and direction. A vector is not a point (since if $A-B$ was a point P we would have $\frac{1}{2}A=\frac{1}{2}B+\frac{1}{2}P$, a point, contrary to what was shown above); and the only vector which is a number is 0 (since $A-B=a\neq 0$ entails that $0=A^2-AB-BA+B^2=(A-B)^2=a^2$). We use the symbols V, W, X, Y and Z, to stand for vectors.

Any product of a number and a vector is a vector, since for $X=B-A$ we have

$$aX=[(1-a)A+aB]-A=P-A,$$

where $P=(1-a)A+aB$ divides the line segment from A to B in the ratio a to $1-a$; accordingly we interpret aX as having the same direction as X and a times its length.

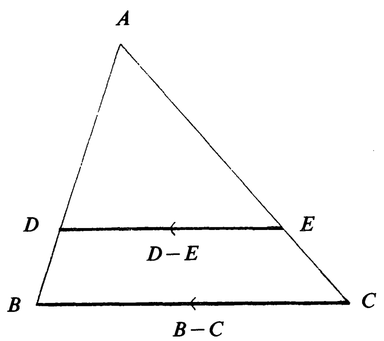


FIGURE 3.

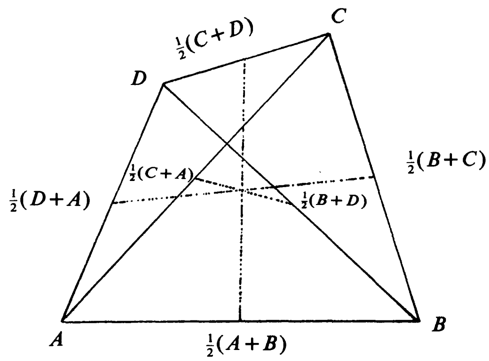


FIGURE 4.

EXAMPLE. Suppose that two edges of a triangle (A,B,C) are divided at D and E in the same ratio, b to a (see FIGURE 3):

$$D=aA+bB, \quad E=aA+bC.$$

Then $D-E=b(B-C)$; thus the line segment from D to E is parallel to the segment from B to C and has b times its length.

EXAMPLE. Many simple algebraic identities have elegant geometric interpretations. For example the identity

$$\tfrac{1}{2}\left[\tfrac{1}{2}(B+C)+\tfrac{1}{2}(D+A)\right]=\tfrac{1}{2}\left[\tfrac{1}{2}(A+B)+\tfrac{1}{2}(C+D)\right]=\tfrac{1}{2}\left[\tfrac{1}{2}(B+D)+\tfrac{1}{2}(C+A)\right]$$

entails that the three line segments joining mid-points of opposite edges of a tetrahedron intersect at their mid-points (see FIGURE 4). This is interesting even in the degenerate case of a quadrilateral. We leave the reader to interpret

$$\tfrac{1}{2}(B+C)-\tfrac{1}{2}(A+B)=\tfrac{1}{2}(C-A)=\tfrac{1}{2}(C+D)-\tfrac{1}{2}(D+A),$$

and

$$\tfrac{1}{4}A+\tfrac{3}{4}\left[\tfrac{1}{3}(B+C+D)\right]=\tfrac{1}{4}(A+B+C+D).$$

Addition of a point and a vector has an obvious interpretation:

$$B=X+A \Leftrightarrow X=B-A.$$

For a given point A, every vector $X=C-D$ may be written $X=B-A$ (by choosing $B=C-D+A$) and every point B may be written $B=A+X$ (by choosing $X=B-A$). The sum of two vectors is a vector, since for $X=B-A$ and $Y=C-B$ we have $X+Y=C-A$; this also indicates how to interpret addition of vectors. (Thus the set of all vectors is a vector space; but let us say emphatically that the set of points is not a vector space.)

It is easy to show that a linear combination of points $a_1A_1+a_2A_2+\cdots+a_nA_n$ is a point if $a_1+a_2+\cdots+a_n=1$, a vector if $a_1+a_2+\cdots+a_n=0$, and $a_1+a_2+\cdots+a_n$ times a point, otherwise.

Lines and Planes

The **line** through the distinct points A and B is the set of all points P of the form $P=aA+bB$ where (necessarily) $a+b=1$; equivalently it is the set of all P of the form $P=A+bX$ where $X=B-A$. One easily sees that points A, B and C are collinear if and only if $aA+bB+cC=0$ for suitable a, b and c, not all zero. (Formally, the existence of distinct points is assured, since otherwise all products $ABCD$ would vanish.)

Lines $\mathfrak{L}$ and $\mathfrak{L}'$ are **parallel** if there are distinct points A and B in $\mathfrak{L}$ and distinct points C and D in $\mathfrak{L}'$ such that $B-A=D-C$. Given a point C and a line $\mathfrak{L}$ through distinct points A and B, there exists a line $\mathfrak{L}'$ through C and parallel to $\mathfrak{L}$, namely the line through C and $D=C+(B-A)$. Readers may prove that this line is unique (Playfair's Axiom) and that distinct parallel lines are disjoint.

The **plane** through the noncollinear points A, B and C consists of all points P of the form $P=aA+bB+cC$; equivalently it is the set of all P of the form $P=A+bX+cY$, where $X=B-A$ and $Y=C-A$. (Formally, existence of three noncollinear points is assured, since otherwise for any four points A, B, C and D we would have either $A=B$ or $C=aA+bB$ and $D=cA+dB$, and, in either case, $ABCD=0$.)

Coordinates

We may now, if we wish, start to talk about coordinate systems. The triple (A,X,Y) is a **cartesian coordinate system** in the plane through A, $B=A+X$ and $C=A+Y$; the **origin** is the point A, and the **axes and scales** are determined by the vectors X and Y (as in FIGURE 5). When we say that P has **coordinates** (x,y) in this coordinate system we mean that $P=A+xX+yY$. Let us find the equation of the line through the points P_1 and P_2 with coordinates (x_1,y_1) and (x_2,y_2). This line is the set

$$\begin{aligned}\{P=aP_1+bP_2:a+b=1\}\\ &=\{A+(ax_1+bx_2)X+(ay_1+by_2)Y:a+b=1\}\\ &=\{A+xX+yY:x=ax_1+bx_2,y=ay_1+by_2\}\\ &=\{A+xX+yY:x=x_1+b(x_2-x_1),y=y_1+b(y_2-y_1)\}\\ &=\left\{A+xX+yY:\frac{y-y_1}{y_2-y_1}=\frac{x-x_1}{x_2-x_1}\right\},\end{aligned}$$

where, in the final equation, the line must not be parallel to an axis.

Coordinates are a device for handling particular real world problems. In geometry they are not needed, and we shall not use them.

Products

We now consider products of points; to begin with, we show that a necessary and sufficient condition for four points to be coplanar, or for three points to be collinear, or for two points to be coincident, is that their product be zero.

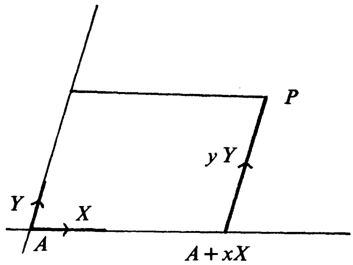
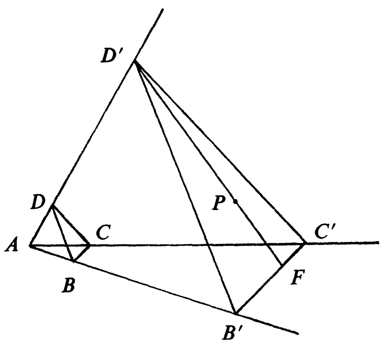


FIGURE 5.

If A, B, C and D are coplanar, one easily sees that $ABCD=0$. (Hence there exist four points which are not coplanar.) Conversely, suppose that A, B, C and D are not coplanar; we shall show that $ABCD\neq 0$. Indeed if P, Q, R and S are any other points, then (see FIGURE 6) we may write each as a linear combination of A, B, C and D (this being where the three-dimensionality of space comes in), and we then see that $PQRS=k\ ABCD$ for some k; thus $ABCD=0$ entails $PQRS=0$, and we have a contradiction of our hypothesis that there is at least one product of 4 points which is nonzero.

PROBLEM: Every product of 5 or more points is 0.

If three points A, B and C are collinear, one easily sees that $ABC=0$. Conversely, suppose that A, B and C are noncollinear; we shall now show that $ABC\neq 0$. Indeed if $ABC=0$, then for a point D not in the plane of A, B and C we would have $ABCD=0$, contrary to what was just proved (such a point D clearly exists). Similarly two points A and B are coincident if and only if $AB=0$.



$P=kD'+lF=kD'+mB'+nC'=aA+bB+cC+dD.$

FIGURE 6.

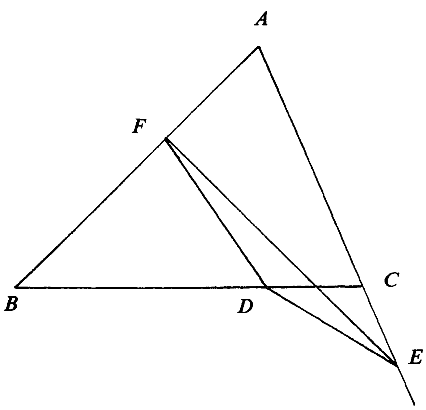


FIGURE 7.

EXAMPLE. (Menelaus' Theorem.) Let (A,B,C) be a proper triangle (i.e., $ABC\neq 0$), and let

$$D=aB+a'C,\quad E=bC+b'A,\quad F=cA+c'B$$

(as in FIGURE 7). Then

$$DEF=abcBCA+a'b'c'CAB=(abc+a'b'c')ABC,$$

and so the necessary and sufficient condition for D, E and F to be collinear is that $abc+a'b'c'=0$.

We note that in our approach Pasch's Axiom is a theorem. One of the gaps in Euclid's axiomatization of geometry was any formal consideration of betweenness. M. Pasch (in 1882) realized that the following should be an axiom: if a line, not passing through the vertices, cuts one side of a triangle internally, then it cuts one of the others internally. We may prove this by defining $F=cA+c'B$ to be **between** A and B if $c>0$ and $c'>0$ and deducing from Menelaus' theorem (using the above notation), that if the transversal cuts the edges of the triangle at points D, E and F, not vertices, with F between A and B, then either a and a' are both greater than 0 or b and b' are both greater than 0.

Now consider two distinct points A and B. If P and Q lie on the line through A and B, we may write $P=A+cX$ and $Q=P+dX$, where $X=B-A$, and we then have $PQ=(A+cX)[A+(c+d)X]=dAX=dAB$. Note that d is the ratio of the vector $Q-P$ to the vector $B-A$; and since the result is independent of c, we in fact have $P'Q'=dAB$ for any two points P' and Q' on the line through A and B such that $Q'-P'=d(B-A)$. Conversely, if PQ is a nonzero multiple of AB, the points A, B, P and Q are collinear; indeed, if $PQ=dAB$, with $d\neq 0$, then $ABP=0$ and $ABQ=0$. In particular, we have shown that $PQ=AB\neq 0$ if and only if the points A, B, P and Q are collinear and $Q-P=B-A$.

Similarly, consider three noncollinear points A, B and C. If P, Q and R lie in the plane through A, B and C, we may write

$$P=a_1A+b_1B+c_1C$$
$$Q=a_2A+b_2B+c_2C$$
$$R=a_3A+b_3B+c_3C,$$

and we see that $PQR=dABC$; indeed, though we do not need the fact, d is the determinant

$$d=\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

Conversely, if PQR is a nonzero multiple of ABC, then the points P, Q and R lie in the plane through A, B and C; indeed from $PQR=dABC$, with $d\neq 0$, we deduce that $0=ABCP=ABCQ=ABCR$. How is one to interpret the constant d which occurs here? To conform with traditional terminology, let us write $PQR: ABC=d$, the ratio of PQR to ABC. Also write $\mu(A,B,C)$ for the oriented area (in some units) of the triangle with vertices A, B and C; the oriented area is simply the (positive) area if (viewed from a fixed point outside the plane) motion from A to B to C to A is anticlockwise, and the negative of this area if (viewed from the same point) this motion is clockwise. Let us examine some special cases:

(i) If $A'=A+a(C-B)$ (see FIGURE 8), then

$$A'BC: ABC=ABC: ABC=1=\mu(A',B,C):\mu(A,B,C).$$

(ii) If $C'=B+a(C-B)$ (see FIGURE 9), then

$$ABC': ABC=aABC: ABC=a=\mu(A,B,C'):\mu(A,B,C);$$

in particular, $a<0$ if and only if C' is on the opposite side of B to C.

Thus, assuming two traditional theorems as facts about area, we see that in cases (i) and (ii), the ratio of the two triple products of points equals the ratio of the associated oriented areas.

Now suppose that we replace a vertex B of the triangle (A,B,C) by an arbitrary point B' in the plane, say $B'=aA+bB+cC=A+b(B-A)+c(C-A)$, where $a+b+c=1$. Then writing $B_1=A+b(B-A)$ we have

$$\begin{aligned} AB'C: ABC &= (AB'C: AB_1C)(AB_1C: ABC) \\ &= [\mu(A,B',C):\mu(A,B_1,C)][\mu(A,B_1,C):\mu(A,B,C)] \\ &= \mu(A,B',C):\mu(A,B,C). \end{aligned}$$

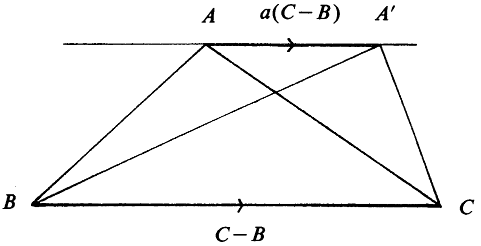


FIGURE 8.

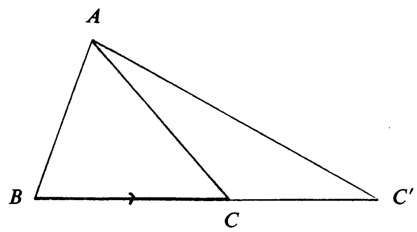


FIGURE 9.

Finally, if A', B' and C' are any 3 points in the plane of A, B and C we have

$$\begin{aligned} A'B'C':ABC &= (A'B'C':A'B'C)(A'B'C:A'BC)(A'BC:ABC)\\ &= [\mu(A',B',C'):\mu(A',B',C)][\mu(A',B',C):\mu(A',B,C)][\mu(A',B,C):\mu(A,B,C)]\\ &= \mu(A',B',C'):\mu(A,B,C), \end{aligned}$$

showing that, quite generally, the ratio $A'B'C':ABC$ is the ratio of the oriented areas of the triangles. (In a formal treatment we would define the ratio of areas this way, and then derive theorems about area.) In particular, we have shown that $A'B'C'=ABC\neq 0$ if and only if the triangles are nondegenerate, coplanar and have the same orientation and area.

EXAMPLE. In traditional language, "a diagonal of a parallelogram bisects it." The proof is immediate (see FIGURE 10): If A, B, C and D are the vertices of a parallelogram, we have $D=C+(A-B)$ and so $ADC=A(C+A-B)C=-ABC$; the negative sign occurs because the triangles are oppositely oriented.

EXAMPLE. Traditional geometric arguments often work unchanged in our algebraic approach. For example, consider the elegant proof of Ceva's Theorem given in [5]. Let (A,B,C) be a triangle, let $D=aB+a'C$, $E=bC+b'A$, $F=cA+c'B$, and suppose that lines through A and D, B and E, and C and F meet at a point P not on an edge of the triangle (see FIGURE 11). Then

$$\frac{BD}{DC}=\frac{ABD-PBD}{ADC-PDC}=\frac{ABP}{APC}$$

and so, using similar identities obtained by symmetry,

$$\frac{BD}{DC}\cdot\frac{CE}{EA}\cdot\frac{AF}{FB}=1.$$

While the steps here are geometrically intutitive, it is easy to prove them in our scheme: for the first equality we are multiplying numerator and denominator by $A-P(\neq 0)$, while from the fact

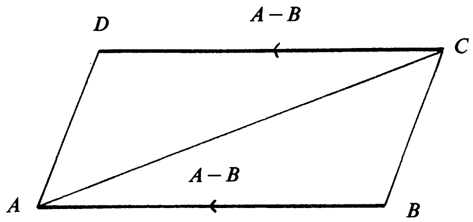


FIGURE 10.

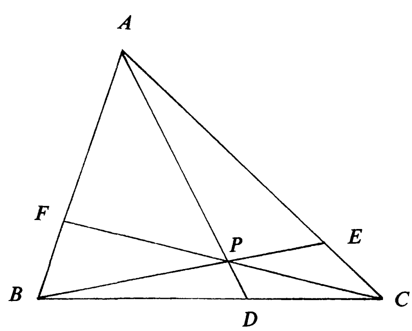


FIGURE 11.

that $D=mA+nP$ we have immediately that $ABD-PBD=nABP-mPBA=ABP$. (Some readers may prefer the following essentially equivalent argument. From $D=aB+a'C$ we have $0=ADP=aABP+a'ACP$ and hence $aABP=a'APC$; using the similar formulae $bBCP=b'BPA$ and $cCAP=c'CPB$, obtained by symmetry, we conclude that $abcABP=a'bcAPC=a'bc'CPB=a'b'c'BPA$ and hence that $abc=a'b'c'$.)

If two vectors, $X=B-A$ and $Y=C-A$, are **parallel** (meaning that one is a multiple of the other), then clearly $XY=0$. Conversely, suppose that $XY=0$; then $ABC=AXY=0$, so that A, B and C are collinear and hence X and Y are parallel.

EXAMPLE. (A Theorem of Pappus.) Let A, B, C, and A', B', C', be collinear triples of points and suppose that $A-B'$ is parallel to $B-A'$ and $B-C'$ is parallel to $C-B'$ (as in FIGURE 12); then $C-A'$ is parallel to $A-C'$. Indeed the collinearity hypotheses give $(A-B)(A-C)=0$ and $(A'-B')(A'-C')=0$, and hence, after some algebraic manipulation, $(A-B')(B-A')+(B-C')(C-B')+(C-A')(A-C')=0$. By the parallelism hypotheses, we have $(A-B')(B-A')=0$ and $(B-C')(C-B')=0$, so we conclude that $(C-A')(A-C')=0$.

We note (see FIGURE 13) that for vectors X, Y, X' and Y' one may show that $XY=X'Y' \Leftrightarrow AXY=AX'Y' \Leftrightarrow ABC=AB'C'$, where A is an arbitrary point and $B=A+X$, $C=A+Y$, $B'=A+X'$ and $C'=A+Y'$. Thus the product of X and Y equals the product of X' and Y' if and only if the associated triangles are coplanar and have equal orientation and area, or, equiva-

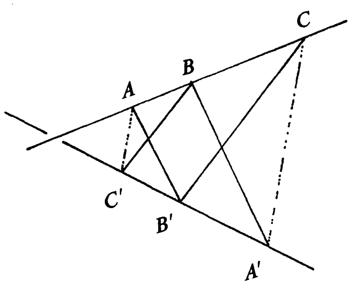


FIGURE 12.

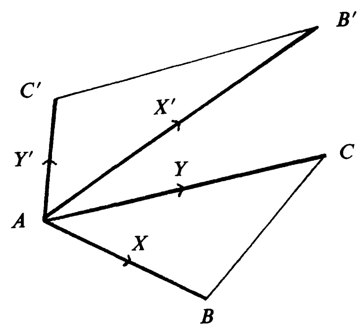


FIGURE 13.

lently, if and only if the vector products $X \times Y$ and $X' \times Y'$, in the traditional sense, are equal. We may therefore use our product XY as an alternative to the usual cross product. Of course XY is not a vector, but in the standard applications this is exactly as it should be.

Arguments similar to those given above show that the ratio $A'B'C'D' : ABCD$ is to be interpreted as a ratio of volumes. Our assumption that there is a nonzero 4-product $ABCD$ was necessary to ensure nontriviality of the notion of volume.

EXAMPLE. Volumes are invariant under translations (see FIGURE 14), since for any vector X $(A+X)(B+X)(C+X)(C+X)=ABCD$. We leave it as an exercise to show that the centroids of the edges of a tetrahedron form a new tetrahedron which has $1/27$ of its volume.

Other Approaches to Affine Geometry

Let us formally list our hypotheses. Our model for 3-dimensional affine geometry is a ring Λ which contains $\mathbb{R}$, the set of all real numbers, and a subset $\mathcal{P}$, the set of points, and which has the following properties:

(1) the number 1 is a unit element in Λ;
(2) $\mathcal{P}$ is an affine subspace of Λ and is spanned by 4 of its elements;

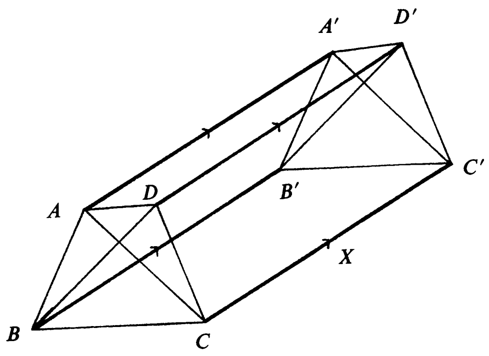


FIGURE 14.

(3) $\mathbb{R} \cup \mathcal{P}$ generates Λ;
(4) $aA = Aa$ for every $a \in \mathbb{R}$ and $A \in \mathcal{P}$;
(5) $A^2 = 0$ for every $A \in \mathcal{P}$;
(6) $0 \notin \mathcal{P}$;
(7) $\mathcal{P}^4 \neq 0$.

Here (2) means that if $A \in \mathcal{P}$, $B \in \mathcal{P}$ and $a + b = 1$ then $aA + bB \in \mathcal{P}$, and that there exist A, B, C and D in $\mathcal{P}$ such that each $P \in \mathcal{P}$ may be written in the form $P = aA + bB + cC + dD$.

In [**2**] we prove the existence and uniqueness of such a structure and we call it a 3-dimensional Grassmann geometry. It is the exterior algebra of Ω, the 4-dimensional linear space spanned by $\mathcal{P}$; the set $\mathcal{V}$ of all vectors is a 3-dimensional linear subspace of Ω, and if A is any element of $\mathcal{P}$ then $\mathcal{P} = A + \mathcal{V}$. It is obvious how to generalize the definition to n dimensions.

Many different applications of Grassmann's ideas are to be found in Forder [**3**], a remarkable book, but one that is not easy to read, largely because the terminology and presentation of the foundations do not improve much on Grassmann's own.

It remains to relate the present development to other ways of doing affine geometry. As well as the classical synthetic treatment, there are four other approaches that I am aware of.

(a) Any coordinate system (A, X, Y, Z) (where A is a point and X, Y and Z are linearly independent vectors) induces an isomorphism of Λ, as an algebra, with the exterior algebra of $\mathbb{R}^4$. Hence it is possible to develop the whole theory in coordinate terms. However this is extremely complicated and changes simple proofs into mind-bending computations with determinants. This kind of thing was attempted in the nineteenth century and is no doubt partly responsible for the neglect of Grassmann's ideas on geometry.

(b) The set $\mathcal{P}$ of all points in Λ may be turned into a linear space by choosing an origin A and defining operations $\oplus$ and $\odot$ by

$$B \oplus C = B + C - A$$
$$a \odot B = aB + (1-a)A.$$

This gives the standard linear algebra picture of affine geometry in which points and vectors, though geometrically different, are represented as the same.

(c) Each vector X induces a function $A \mapsto A + X$ on $\mathcal{P}$, called a **translation**. Conversely, it is possible to define an affine space as a pair consisting of a set $\mathcal{P}$ and a linear space of transformations of $\mathcal{P}$ with suitable properties. A typical such development may be found in [**1**].

(d) One way of interpreting the multiples of points in a Grassmann geometry is as mass-points; by doing this it is possible to model the theory of centres of mass. The converse idea of basing geometry on mass-points was developed at length by Möbius [**8**]. An interesting modern version is given by Hausner [**6**]. His approach is to define a mass-point as an ordered pair (m, P), where m is a positive number, to impose a suitable axiomatic structure on the set of mass-points, and to prove that the set of mass-points may be embedded in a linear space.

The most important respect in which these approaches differ from Grassmann's is that they do not bring in the multiplication of points. Although an exterior algebra is considered, it is the exterior algebra of the space $\mathcal{V}$ of vectors. The elegant Grassmann geometry structure in which numbers, points, vectors, affine length, area and volume are so simply represented is present only in part, and those proofs which need exterior algebra are complicated by the fact that they must be done in terms of vectors. Moreover in standard treatments the exterior algebra is developed by constructing it, using alternating multilinear forms. Like the constructive development of the real numbers from the natural numbers, and for the same reasons, this is conceptually difficult. It is much simpler to begin with a full axiomatization of the real numbers, or, for that matter, of affine geometry.

Teaching Geometry

In this paper I have purposely avoided an appeal to linear algebra. I want to draw attention to the fact that the algebraic approach to geometry which has been sketched here might well precede a course in linear algebra. I think it likely that to the uninitiated a Grassmann geometry structure may well be easier to work with than a linear space structure; this may seem paradoxical, since a Grassmann geometry is, among other things, a linear space, but here, as elsewhere in mathematics, more structure makes it easier to prove interesting theorems.

It has become standard to treat geometry, if at all, as a part of linear algebra; despite good intentions, this has tended to eliminate interesting geometrical theorems from the curriculum. Attention is focused on predictable facts about lines and planes and on the classification of conic sections, and proofs involve the sophisticated notion of a linear map. How much simpler it is to see that the plane through A, B and C is the set of all points P satisfying $ABCP=0$, than (what is essentially equivalent!) to view it as a translate of the kernel of a linear functional.

As a purely mechanical matter it is not hard to master the business of putting together points and numbers, adding and multiplying according to ordinary algebraic rules together with the geometrically natural condition that $BA=-AB$. The difficulties, if any, lie in the interpretation of this structure as a model of physical space. In a first treatment one would probably not show that everything follows from the rule for interpreting $P=aA+bB$, where $a+b=1$, $a \geqslant 0$ and $b \geqslant 0$, but rather include other interpretative rules as well (for example, that the equation $A'B'C'D'=kABCD$ is to be interpreted as meaning that the tetrahedron with vertices A', B', C' and D' has $|k|$ times the volume of the tetrahedron with vertices A, B, C and D, and has the same or opposite orientation according as k is positive or negative). After all, in viewing our algebraic structure as a model of physical space (which is how Grassmann viewed it) it is of no consequence that some rules of interpretation may be derived from others. What matters is that no theorem should conflict with physical fact—and, of course, on an everyday scale (not microscopic and not astronomical) and with everyday measuring instruments there is no such conflict.

References

[1] C. T. J. Dodson and T. Poston, Tensor Geometry, Pitman, San Francisco, 1977.

[2] Desmond Fearnley-Sander, Affine geometry and exterior algebra, Houston J. Math., 6 (1980) 53–58.

[3] H. Forder, The Calculus of Extension, Cambridge, London, 1941.

[4] H. G. Grassmann, Die Ausdehnungslehre, Berlin, 1862.

[5] H. S. Hall and F. H. Stevens, A School Geometry, Macmillan, London, 1903.

[6] Melvin Hausner, The centre of mass and affine geometry, Amer. Math. Monthly, 69 (1962) 724–737.

[7] ______, A Vector Space Approach to Geometry, Prentice-Hall, Englewood Cliffs, NJ, 1965.

[8] A. F. Möbius, Der barycentrische Calcul, Leipzig, 1827.

[9] G. R. Morrow (trans.), Proclus—a Commentary on the First Book of Euclid's Elements, Princeton Univ. Press, 1970.

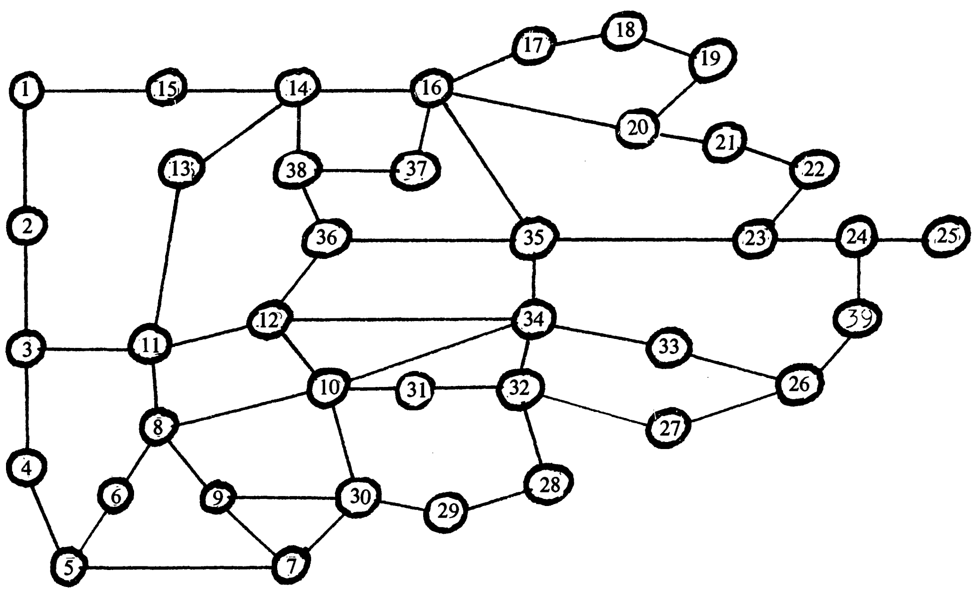
Linear Algebra in Geography: Eigenvectors of Networks

Accessibility of towns on a road map is measured by the principal eigenvector of its adjacency matrix.

PHILIP D. STRAFFIN, JR.
Beloit College
Beloit, WI 53511

Like economics and psychology before it, modern theoretical geography is a discipline in which the use of mathematics has become increasingly important. In this article I would like to discuss one use of linear algebra in geography. The application is elementary enough to be presented to a first undergraduate linear algebra class, although to my knowledge it has not appeared in linear algebra texts except for a brief mention in [2]. It illustrates well the problem of giving meaningful interpretation to the results of mathematical manipulation of physical data.

The geographical problem starts with a transportation network—a map of geographically significant entities (for instance urban centers) connected by transportation routes (for instance railway lines, highways, or scheduled air routes). Such a network can be conveniently repre-



Trade routes in medieval Russia (adapted from [13]).

FIGURE 1.

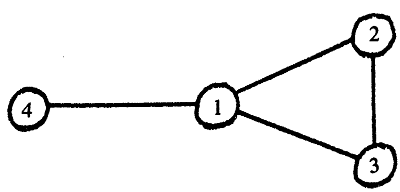
sented by a graph in which the urban centers are vertices and the transportation routes are edges joining pairs of vertices. For example, the graph in FIGURE 1, adapted from [13], represents the major river trade routes in central Russia in the twelfth and thirteenth centuries.

An Index of Accessibility

The problem we address is the development of a suitable index of what geographers have called the accessibility of each vertex in the network. This index should provide a numerical answer to such questions as, "How accessible is this vertex from other vertices in the network? What is its relative geographical importance in the network?" Such an index, once devised, could be used in a number of interesting ways. For instance:

1. Knowledge of which vertices have the highest accessibility could be of interest in itself. For example, the principality of Moscow (number 35 in FIGURE 1) eventually assumed the dominant position in central Russia. Pitts in [13] reviews claims by geographers that this was due to the strategic location of Moscow on medieval trade routes. Does Moscow indeed have the highest accessibility in this network, or must other factors have modified the forces of "situational determinism?"
2. The accessibility of vertices could be statistically correlated to other economic, sociological or political variables to test theoretical hypotheses in geography. Is high accessibility of an urban center in a transportation network associated with high per capita income, a high suicide rate, or a high degree of political awareness?
3. Accessibility indices for the same urban centers in different transportation networks could be compared, as in [4], where the rail network and the interstate highway network are compared for cities in the southeastern United States.
4. Proposed changes in a transportation network could be evaluated in terms of their effect on the accessibility of vertices. Which urban centers would become more, or less, transportationally important?

One solution to the problem of developing a suitable index of accessibility was first proposed by Peter Gould in [6]. Postponing for a moment the justification of his index, let us see how it works. As an example, consider the graph in FIGURE 2.



A simple transportation network.

FIGURE 2.

The **adjacency matrix** of a graph is the square matrix with rows and columns labeled by the vertices, and entries

$$a_{ij}=\begin{cases}1 \text{ if vertices } i \text{ and } j \text{ are joined by an edge,}\\ 0 \text{ if vertices } i \text{ and } j \text{ are not joined by an edge.}\end{cases}$$

It is traditional to define the diagonal entries a_{ii} to be zero, in which case we will denote the adjacency matrix by A. We will have more occasion to use the modified matrix in which the diagonal entries are defined to be one, and we denote this matrix by $B=A+I$. For example, the adjacency matrices for the graph in FIGURE 2 are

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

Gould's definition is based on the eigenvalues of the matrix A. In this example, the characteristic polynomial for A is $\lambda^4 - 4\lambda^2 - 2\lambda + 1 = (\lambda + 1)(\lambda^3 - \lambda^2 - 3\lambda + 1)$, with approximate eigenvalues $\lambda_1 = 2.17$, $\lambda_2 = .31$, $\lambda_3 = -1.00$, $\lambda_4 = -1.48$. Now compute the eigenvector $\mathbf{v}_1$ for the largest, or **principal**, eigenvalue λ_1, normalized in any convenient way. In our example we have $\mathbf{v}_1 = (.32, .27, .27, .14)$, where the normalization has been done so that the components add to one. The components of this eigenvector are Gould's **index of accessibility**. We note that vertex 1 has the

Node	$\mathbf{v}_1$ (Gould index)	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_{19}$
1. Novgorod	.0042	.0439	.0883	−.1278
2. Vitebsk	.0065	.0020	.0990	−.0233
3. Smolensk	.0186	−.0383	.1630	.1231
4. Kiev	.0104	−.0603	.1551	.1651
5. Chernikov	.0176	−.1352	.2309	−.0894
6. Novgorod Severskij	.0208	−.1156	.1687	.0175
7. Kursk	.0303	−.2133	.2623	−.2009
8. Bryansk	.0547	−.1978	.1974	.0930
9. Karachev	.0386	−.2295	.2528	−.0495
10. Kozelsk	.0837	−.1724	−.0801	.1676
11. Dorogobusch	.0477	−.0519	.1598	−.1167
12. Vyazma	.0722	−.0046	−.1081	−.2917
13. "A"	.0207	.0912	.1534	.0519
14. Tver	.0242	.3146	.2296	.1273
15. Vishnij Totochek	.0082	.1245	.1252	−.0028
16. Ksyatyn	.0321	.4337	.1800	−.0708
17. Uglich	.0105	.1872	.0969	−.3397
18. Yaroslavl'	.0044	.1055	.0660	.0014
19. Rostov	.0047	.1165	.0706	.3400
20. "B"	.0121	.2299	.1133	.0680
21. "C"	.0054	.1118	.0370	−.2554
22. Suzdal	.0068	.0921	−.0194	−.1201
23. Vladimir	.0182	.1533	−.0862	.2309
24. Nizhnij Novgorod	.0076	.0721	−.0968	−.0759
25. Bolgar	.0022	.0251	−.0381	−.3719
26. Isad'-Ryazan	.0142	.0122	−.2117	.1015
27. Pronsk	.0164	−.0079	−.2065	−.0825
28. Dubok	.0178	−.0480	−.1123	−.0769
29. Elets	.0193	−.1033	.0275	.1027
30. Mtsensk	.0493	−.2496	.1822	.0979
31. Tula	.0363	−.0720	−.1547	.2413
32. Dedoslavl'	.0429	−.0349	−.3126	−.1183
33. Pereslavl'	.0267	.0138	−.2095	−.0224
34. Kolomna	.0788	.0275	−.3202	−.1061
35. MOSCOW	.0490	.2772	−.1025	.2432
36. Mozhaysk	.0415	.1837	−.0340	−.0044
37. Dmitrov	.0159	.2396	.1198	−.1132
38. Volok Lamskij	.0234	.2563	.1242	.0476
39. Murom	.0063	.0293	−.1215	.1255
	$\lambda_1 = 4.48$	$\lambda_2 = 3.88$	$\lambda_3 = 3.54$	$\lambda_{19} = 1.20$

Eigenvectors for the Russian trade route graph of Figure 1. (Eigenvalues are for the matrix B.)

TABLE 1.

highest index, followed by vertices 2 and 3, followed by vertex 4 with the lowest index, and that these results accord well with intuition. The accessibility indices for vertices in the Russian trade route graph are given in the second column of TABLE 1. Notice that Moscow is not the most highly accessible vertex in the network. In fact it ranks sixth, behind Kozelsk, Kolomna, Vyazma, Bryansk and Mtsensk. The conclusion would be that other sociological and political factors must have been important in Moscow's rise.

It is comforting to have a procedure like the above which seems to reinforce and complement our intuition with numbers carried to several decimal places. However, the first question both geographers and mathematicians must ask is, "What do these numbers mean? Why is it that this manipulation through graphs, matrices, eigenvalues and eigenvectors should produce numbers entitled to the name of an 'accessibility index'?" On this crucial question, Gould and the first users of his index (see [1] for example) were unfortunately vague:

> Vectors representing well-connected towns will not only lie in the middle of a large number of dimensions but will tend, in turn, to lie close to the principal axis of our enveloping oblate spheroid. Towns that are moderately well-connected will not lie in the middle of so many dimensions as the well-connected towns, and will tend to form small structural clusters on their own. ([6], page 66)

Although the geometric intuition in this statement tells us something about why the principal eigenvector might have something to do with accessibility, it certainly does not tell us why its components have a claim to giving a precise index. The goal of this article is to use linear algebra to develop three different models to justify Gould's index. We begin with some background from linear algebra.

The Perron-Frobenius Theorem

First, note that the matrix identity $B=A+I$, where I is the $n\times n$ identity matrix, entails that the eigenvalues of B are exactly one larger than the corresponding eigenvalues for A, and that the eigenvectors of the two matrices are exactly the same. Hence we may use the matrix B instead of A to compute Gould's index. Second, the matrix B is symmetric. Linear algebra tells us that a real symmetric matrix can be diagonalized by an orthogonal matrix. Hence all the eigenvalues of B are real, and we can rank them $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$. The corresponding eigenvectors $\mathbf{v}_1,\ldots,\mathbf{v}_n$ are real and give an orthogonal basis for R^n. Thus the prescription of choosing the largest eigenvalue makes sense, and the components of the principal eigenvector will be real numbers.

The final, crucial piece of information we need is the Perron-Frobenius Theorem for nonnegative square matrices. This theorem is so important for applications of linear algebra in the social sciences that it ought to be at least stated in an undergraduate linear algebra course. Detailed discussions can be found in [3] and [16]. A matrix $M=(m_{ij})$ is **nonnegative** if $m_{ij} \geqslant 0$ for all i,j. A square nonnegative matrix M is said to be **primitive** if there exists a positive integer k such that all the entries of M^k are strictly positive.

PERRON-FROBENIUS THEOREM. *If M is an $n\times n$ nonnegative primitive matrix, then there is an eigenvalue λ_1 such that*

(i) *λ_1 is real and positive, and is a simple root of the characteristic equation,*
(ii) *$\lambda_1 > |\lambda|$ for any eigenvalue $\lambda \neq \lambda_1$,*
(iii) *λ_1 has a unique (up to constant multiples) eigenvector $\mathbf{v}_1$, which may be taken to have all positive entries.*

To apply this theorem to the adjacency matrix of a graph, note that by the definition of matrix multiplication, the ijth entry of A^k counts the number of ways of getting from vertex i to vertex j by paths of length k. The effect of the 1's along the diagonal in $B=A+I$ is to make the ijth entry of B^k count the number of ways of getting from vertex i to vertex j by paths of length k, including possible stopovers at vertices along the way. For a connected graph, the **diameter** of the graph is the smallest integer k such that any vertex may be reached from any other vertex by a path of length less than or equal to k. Hence if our graph is connected and we choose k to be bigger than or equal to its diameter, the entries of B^k will all be positive. In other words, B is primitive, so the Perron-Frobenius Theorem applies. (It is an easy exercise to show that if the underlying graph is **bipartite**, that is, if its vertices can be partitioned into two sets V_1 and V_2 such that no two vertices in V_i are adjacent, $i=1,2$, then the matrix A will *not* be primitive. See **[10]** or **[15]**. It is for this reason that we work with B instead of A.)

Thus if the transportation network is connected, we are guaranteed that the principal eigenvector $\mathbf{v}_1$ is well-defined and has all positive entries. Moreover, consider any vector $\mathbf{x}$ not orthogonal to $\mathbf{v}_1$:

$$\mathbf{x}=\alpha_1\mathbf{v}_1+\alpha_2\mathbf{v}_2+\cdots+\alpha_n\mathbf{v}_n \quad (\alpha_1\neq 0).$$

Then $B^k\mathbf{x}=\lambda_1^k\alpha_1\mathbf{v}_1+\lambda_2^k\alpha_2\mathbf{v}_2+\cdots+\lambda_n^k\alpha_n\mathbf{v}_n$ and, as $k\to\infty$,

$$\frac{B^k\mathbf{x}}{\lambda_1^k}=\alpha_1\mathbf{v}_1+\left(\frac{\lambda_2}{\lambda_1}\right)^k\alpha_2\mathbf{v}_2+\cdots+\left(\frac{\lambda_n}{\lambda_1}\right)^k\alpha_n\mathbf{v}_n\to\alpha_1\mathbf{v}_1,$$

since λ_1 is the eigenvalue of strictly largest modulus. In other words, *the ratios of the components of $B^k\mathbf{x}$ approach the ratios of the components of $\mathbf{v}_1$ as k increases.* This important fact will provide the basis for our three justifications for using the components of $\mathbf{v}_1$ as an accessibility index.

Justifications of Gould's Index

The first justification of Gould's index relies on the fact mentioned above that the ijth entry of B^k counts the number of paths between vertices i and j of length k (allowing stopovers). A highly accessible vertex should have a large number of paths to other vertices. This idea was used to study accessibility before Gould's article in 1967—see **[4]** and **[13]** for example. The idea was to compute B^k (or often A^k) for some suitably large k (often the diameter of the graph), and then use the row sums of its entries as a measure of accessibility. The accessibility index of vertex i would thus be the sum of the entries in the ith row of B^k, and this is the total number of paths of length k (allowing stopovers) from vertex i to all vertices in the graph. One problem with this method is that the integer k seems arbitrary. At exactly what path length should you stop counting? Here our linear algebra comes to the rescue. Let $\mathbf{e}$ be the n-dimensional column vector all of whose components are 1's. Then the vector of row sums of B^k is just the vector $B^k\mathbf{e}$. Since $\mathbf{e}$ and $\mathbf{v}_1$ both have all positive entries, they cannot be orthogonal, so the ratios of the components of $B^k\mathbf{e}$ approach the ratios of the components of $\mathbf{v}_1$ as k increases. As we count longer and longer paths, this measure of accessibility converges to Gould's index.

A second justification for Gould's index was given by Tinkler in **[17]**. Imagine a rumor starting with a teller at some vertex i in the network at time 0. By time 1, the teller has told the rumor to someone at each vertex which is joined to vertex i by an edge (and of course he remembers it himself). By time 2, each person who knew the rumor at time 1 has told it to someone at each vertex which is adjacent to his vertex. As time progresses, the rumor will spread throughout the network, and we might measure the accessibility of a vertex by the number of people at that vertex who know the rumor. If the spread of a rumor seems too frivolous, think of the spread of a technological innovation, or a trade good.

Once again, our linear algebra is applicable. Let $\mathbf{x}_i$ be the column vector with 1 in the ith position and zeros elsewhere. The distribution of the rumor at time 1 is given by $B\mathbf{x}_i$, and the distribution at time k by $B^k\mathbf{x}_i$. Since $\mathbf{x}_i$ has no negative entries, it cannot be orthogonal to $\mathbf{v}_1$, so the vectors $B^k\mathbf{x}_i$ approach a multiple of $\mathbf{v}_1$ as k increases. No matter where the rumor starts, its equilibrium distribution after a large number of time periods is given by Gould's index.

This model also indicates a geographical meaning of the principal eigenvalue λ_1: it gives the equilibrium growth rate of a rumor spreading according to the model. It is thus a measure of what might be called the "total connectivity" of a network: highly connected networks should be those in which rumors can spread quickly. Perron-Frobenius theory tells us some interesting things about λ_1, for instance that adding an edge to a network must always increase λ_1. For other results on λ_1 and the other eigenvalues of graphs, see **[10]** and **[15]**.

Our third justification of Gould's index is based on an idea discussed by J. W. Moon in **[12]** about how we could measure the relative strengths of players in a round-robin tournament. A first order approximation of the strength of player i might be simply the number of players he beat in the tournament. But beating strong players ought to count more heavily than beating weak players. Hence a second order index of the strength of player i might be the sum of the first order strengths of players he beat. And so it goes: we keep iterating and hope for convergence.

If we apply this idea to a transportation network, it works like this. As a first order measure of the geographical importance of vertex i, we simply use its **degree**, the number of vertices adjacent to vertex i. But being adjacent to important vertices should count more heavily than being adjacent to unimportant vertices. Hence as a second order index of the importance of vertex i, we might use the sum of the first order indices of vertices adjacent to vertex i (and let us count it as adjacent to itself). As we continue the iteration, we recognize what will happen. If $\mathbf{x}$ is the vector of degrees, our kth order index of geographical importance is the vector $B^{k-1}\mathbf{x}$, and these vectors converge to Gould's index as k increases. Gould's index gives the equilibrium relative importance of vertices under this iterative procedure.

Summary

Gould's idea was to measure the accessibility, or geographical importance, of nodes in a transportation network by using the components of the principal eigenvector of the adjacency matrix of the corresponding graph. Originally this idea was justified only on fairly vague heuristic grounds. We have seen that linear algebra, in particular the Perron-Frobenius Theorem, allows us to obtain Gould's index by three separate chains of reasoning. The index gives the relative number of paths joining each vertex to all vertices in the graph, the equilibrium distribution of a rumor spreading in the graph from any vertex, and the equilibrium relative importance of vertices calculated according to an iterative scheme. If we believe that any of these three models captures what we would wish to describe as accessibility, Gould's index is appropriate.

Computing eigenvalues and eigenvectors of large matrices, even when they are sparse, is not an easy task. Hence it is useful to recognize that our analysis yields as a by-product an efficient algorithm, well-known to numerical analysts, for approximating $\mathbf{v}_1$ and λ_1 as closely as desired. Label the columns of an array by the vertices. In the zeroth row, enter a 1 in each column. In the jth column of the $(i+1)$th row, enter the sum of the entries in the ith row corresponding to vertices to which vertex j is adjacent (and count it as adjacent to itself). The rows will converge quite rapidly to the Gould index. The ratio of the total of the $(i+1)$th row to the total of the ith row will converge to the principal eigenvalue of the matrix B. (Recall that this will be one larger than Gould's principal eigenvalue, since he used A instead of B.) The procedure is illustrated in TABLE 2 for the simple graph of FIGURE 2.

Extensions and Generalizations

Gould's index, or its cruder predecessor described above, has been used by geographers to study such things as trade routes in Serbia in the reign of Stefan Dušan **[1]**, the road systems of Uganda in 1921 and of Syria in 1963 **[6]**, the U.S. interstate highway system **[4]**, and the growth of the São Paulo economy **[5]**. See **[17]** and **[20]** for other references to studies of the highway network of northern Ontario, internal migration in Hungary in the 1960's, detribalization in Tanzania, urban accessibility in Indianapolis and Columbus, and the evolution of airline routes

k	1	2	3	4	Total
0	1	1	1	1	4
1	4	3	3	2	12
2	12	10	10	6	38
3	38	32	32	18	120
4	120	102	102	56	380

$$\mathbf{v}_1 \approx (.32, .27, .27., .15) \quad \lambda_1 \approx \frac{380}{120} \approx 3.17$$

Iterative approximation of the principal eigenvalue and eigenvector for the graph in FIGURE 2.

TABLE 2.

in the United States and Australia. In looking at applications, we need not limit ourselves to problems in geography. For instance, indices of the type we have been considering have also been used to study the idea of status in sociology **[11]**, **[8]**. In this context vertices might represent individuals and edges represent friendship or acquaintance. We would be interested in identifying the most "well-connected" individuals, and knowing other well-connected people should count more heavily than knowing poorly connected people. Our third model would suggest the Gould index as an appropriate measure of status.

Several generalizations of Gould's index would be natural. For instance, if enough information were available, we might wish to weight the edges of the transportation graph in some suitable way. Since the weighted adjacency matrix would still be symmetric and primitive nonnegative, the analysis would still work. For example, if we could weight the edges in the Russian trade route network by the volume or worth of trade along various routes, it might turn out that Moscow did have the highest weighted Gould index. For this particular example, though, the historical information necessary for such weighting is not available. The Gould index could also be adapted to directed graphs (one-way trade flows), though then the adjacency matrix would no longer be symmetric, and we would have to require that the digraph be "strongly connected" for the adjacency matrix to be primitive.

In addition to using the principal eigenvector, Gould and other geographers have proposed that the non-principal eigenvectors $\mathbf{v}_2, \mathbf{v}_3, \ldots$ might have geographical meaning. The non-principal eigenvectors must be orthogonal to $\mathbf{v}_1$, which has all positive components. Hence they have some positive and some negative components. Thus in a graph, a non-principal eigenvector partitions the vertices into those with positive components in the eigenvector and those with negative components. This partitioning might pick out significant geographical subsystems. In TABLE 1, the eigenvectors $\mathbf{v}_2, \mathbf{v}_3$ and $\mathbf{v}_{19}$ are given. If you draw the corresponding partitions on FIGURE 1, you will find that $\mathbf{v}_2$ partitions the graph into a northern section and a southern section, $\mathbf{v}_3$ gives an east-west partition, and the partition given by $\mathbf{v}_{19}$ is charmingly complicated. Analyses of this type are given in **[6]** and **[1]**. Going beyond the mystical stage in justifying this kind of analysis seems much more complicated than in the case of the principal eigenvector. Tinkler in **[17]** has proposed an interpretation based on the spread of a rumor (positive numbers) and a canceling anti-rumor (negative numbers) through the network. If the initial distribution of rumor and anti-rumor is exactly given by the components of a non-principal eigenvector corresponding to an eigenvalue $\lambda_j > 1$, then both rumor and anti-rumor will be able to grow at a rate λ_j without either forcing the other out. One problem with this interpretation is lack of stability: if the initial distribution $\mathbf{x}$ differs only slightly from being orthogonal to $\mathbf{v}_1$, we know that eventually its $\mathbf{v}_1$ component will dominate and destroy the coexistence of rumor and anti-rumor. Generically, coexistence is impossible. The nature of the significance, if any, of partitions given by non-principal eigenvectors seems to me as yet unjustified by a reasonable model. Discussions of this question may be found in **[17]**, **[9]** and **[18]**.

I would close by suggesting that the reader interested in getting a feel for how the Gould index works might enjoy calculating Gould indices for some simple families of graphs; for

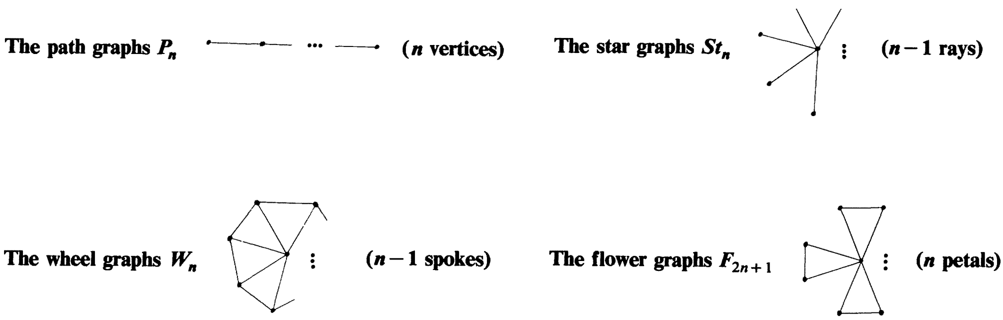


FIGURE 3.

example, those illustrated in FIGURE 3. Some answers appear in [**17**]. Readers interested in exploring other methods of geographical analysis based on graph theory and linear algebra might consult [**7**], [**14**], [**19**] and [**20**].

The author thanks Keith Tinkler and the referees for suggestions which improved the presentation of this paper.

References

[1] F. W. Carter, An analysis of the medieval Serbian oecumene: a theoretical approach, Geografiska Annaler, 51 (1969) 39–56.
[2] T. J. Fletcher, Linear Algebra Through Its Applications, Van Nostrand and Reinhold, London, 1972.
[3] F. R. Gantmacher, The Theory of Matrices, vol. 2, Chelsea, New York, 1959.
[4] William Garrison, Connectivity of the interstate highway system, Papers and Proceedings of the Regional Science Association, 6 (1960) 121–137.
[5] H. L. Gauthier, Transportation and the growth of the São Paulo economy, J. Regional Science, 8 (1968) 77–94.
[6] Peter Gould, The geographical interpretation of eigenvalues, Transactions of the Institute of British Geographers, 42 (1967) 53–85.
[7] Peter Haggert and Richard Chorley, Network Analysis in Geography, St. Martin's Press, New York, 1969.
[8] F. Harary, R. Norman and D. Cartwright, Structural Models: An Introduction to the Theory of Directed Graphs, John Wiley, 1965.
[9] A. Hay, On the choice of methods in the factor analysis of connectivity matrices: a comment, Trans. Inst. British Geog., 66 (1975) 163–167.
[10] A. J. Hoffman, Eigenvalues of graphs, pages 229–245 in Fulkerson (Editor), Studies in Graph Theory, Part II, MAA Studies in Mathematics, 1975.
[11] W. Katz, A new status index derived from sociometric analysis, Psychometrika, 18 (1953) 39–43.
[12] J. W. Moon, Topics on Tournaments, Holt Rinehart and Winston, New York, 1968.
[13] F. R. Pitts, A graph theoretical approach to historical geography, Professional Geographer, 17 (1965) 15–20.
[14] Fred Roberts, Discrete Mathematical Models, Prentice-Hall, New Jersey, 1976.
[15] A. J. Schwenk and R. J. Wilson, On the eigenvalues of a graph, pages 307–336 in Beinecke and Wilson, Editors, Selected Topics in Graph Theory, Academic Press, New York, 1978.
[16] E. Seneta, Nonnegative Matrices, An Introduction to Theory and Applications, John Wiley, New York, 1973.
[17] K. J. Tinkler, The physical interpretation of eigenfunctions of dichotomous matrices, Trans. Inst. British Geog., 55 (1972) 17–46.
[18] ______, On the choice of methods in the factor analysis of connectivity matrices: a reply, Trans. Inst. British Geog., 66 (1975) 168–171.
[19] ______, Introduction to Graph Theoretical Methods in Geography, Geographical Abstracts Ltd., University of East Anglia, Norwich, England, 1977.
[20] ______, Graph theory, Progress in Human Geography, 3 (1979) 85–116.

NOTES

Mathematics and Theology: Game-Theoretic Implications of God's Omniscience

STEVEN J. BRAMS
Department of Politics
New York University
New York, NY 10003

In this note I shall offer an operational definition of the omniscience of God—or some other superior being— in a game, demonstrate in a specific instance how this definition can lead to a paradox in the play of a game, and, finally, suggest three ways in which the paradox can be resolved, only one of which is satisfactory. I shall thereby try to show how mathematical reasoning can help to clarify the central question of theology, man's relationship to God. Although only a single two-person, nonzero-sum game will be used to illustrate the consequences I shall derive from my definition of omniscience, I shall state some more general consequences at the end of this essay and also comment briefly on the role mathematics might play in illuminating questions in religion and theology, subjects considered by many to be inaccessible to formalization and rigorous analysis.

Consider the following two-person game played between man and God, whom I assume to be on a "collision course:"

—If both do not compromise, the result will be disastrous (the worst outcome for both).
—If both compromise, they will reach a satisfactory agreement (the next-best outcome for both).
—If one player compromises and the other does not, the one who does not prevails (best outcome) over the one who does (next-worst outcome).

Although I refer to the players as "man" and "God," this is simply a convenience; man could as well be "woman" or "person," and God could be some more secular "superior being." I shall indicate shortly exactly in what way God is distinguished from man as a player.

To facilitate its analysis, the game can be represented in normal, or matrix, form. Let each player have two strategies, "compromise" and "don't compromise," as shown in FIGURE 1. The

		God	
		Compromise (C)	Don't compromise ($\bar{C}$)
Man	Compromise (C)	(3,3) Cooperation	(2,4) God prevails over man
	Don't compromise ($\bar{C}$)	(4,2) Man prevails over God	(1,1) Disaster

The matrix game of "Chicken" used as a model for conflict between God and man. The outcomes are merely ranked in order of preference, 4 = best, 3 = next-best, 2 = next-worst and 1 = worst. Each outcome, e.g., (4,2), lists the result for the row player (man) first, and the column player (God) second.

FIGURE 1.

choice of a strategy by both players leads to one of four possible outcomes: CC (both compromise); $C\bar{C}$ (man compromises, God does not); $\bar{C}C$ (man does not compromise, God does); $\overline{CC}$ (neither man nor God compromises). Let "4" represent the best outcome for a player, "3" next best, "2" next worst, and "1" worst. Thus, the higher the number, the better the outcome, though I assume no value, or "cardinal utility," is associated with each outcome. Because the numbers represent only preferences, one cannot say how much more a player prefers, say, the outcome he ranks 3 to the outcome he ranks 2.

In a matrix game, it is customary to represent an outcome as an ordered pair of numbers, with the first number being the preference of the row player (man), the second number the preference of the column player (God). Thus, the outcome (2,4) in the matrix of FIGURE 1 is next-worst for man, best for God.

The outcome matrix of FIGURE 1 simply offers a symbolic formulation of the game described informally at the beginning of this note. It is a game in which both players face disaster if they do not compromise, both do relatively well if they do. However, each player can obtain his best outcome only if he does not compromise and the other player does, which hurts the player who does compromise but is not as disastrous for him as the case in which both do not compromise and head for collision.

In the game-theoretic literature, this game is called "Chicken," and its implications, particularly in politics, are discussed in [**4**]. It is a game that has been used to model, among other situations, the Cuban missile crisis of October 1962 in which the United States and the Soviet Union faced each other in a nuclear confrontation. (For a criticism of this application, and a suggested alternative model, see [**2**].)

I do not wish to suggest here that God's relationship to man can invariably be modeled as a game of Chicken. However, there are certainly instances in the Bible in which such figures as Adam and Eve, Cain, Saul, and even Moses confronted God and, in addition, were forewarned or had forebodings that the consequences of their defiance would not be benign.

Moreover, God in these conflicts did not always prevail. For example, after Moses's intercession, He reversed His original position to destroy the Israelites following their idolatry at Mount Sinai and accepted a kind of compromise solution in which only some were killed. Indeed, although one might be hard put to say that disaster befell God after Adam and Eve sinned, Cain murdered his brother, Abel, or Saul disobeyed God's mouthpiece, the prophet Samuel, He certainly did not appear elated by these violations of His commands and precepts. (See [**1**] for further analysis of these stories and this chain of reasoning.)

My point is that while the game of Chicken obviously does not mirror all conflicts and confrontations that man has had or might have with God, it is not implausible to think of man and God as being occasionally on a collision course, with possible results not salutary for either. In fact, most games that the biblical God of the Old Testament played with man were such that both players could, relatively speaking, "win" or "lose" simultaneously. That is, what one player won the other player did not necessarily lose, as is the case in "zero-sum" games, wherein the sum of the payoffs to the players is always zero (if reckoned in cardinal utilities). By contrast, Chicken is a nonzero-sum game, for which there is no generally accepted solution like "minimax" in zero-sum games.

But God is no ordinary game player, so standard solutions, even if they did exist, might not be applicable anyway. For example, God is often thought to possess omniscience, implications of which I shall focus on in the remainder of this essay.

In a two-person game, I define **omniscience** to be the ability of one player to predict the strategy choice of the other player before the play of the game—that is, before both players actually make their strategy choices, which determine the outcome. Given this ability, I assume that the omniscient player (God) acts, on the basis of his prediction, to achieve the best possible outcome for himself. (In game theory, God's omniscience can be conceptualized as moving second in a game, with man's prior choice of a strategy already known.)

This apparent superior ability of God, however, may not be an unmixed blessing *if man is aware that God possesses it* (which I henceforth assume he is). Specifically, in the game of Chicken, if man has this awareness, he should not compromise, because he knows that God will predict this choice on his part. Thereby God will be forced to compromise Himself to prevent His worst outcome (1) from occurring and to salvage instead His next-worst outcome (2).

Clearly, then, if man is aware of God's omniscience, he would prevail over Him, yielding the outcome (4,2) in Chicken. The fact that God's superior ability, and man's awareness of it, help man and hurt God I call the *paradox of omniscience*. It is a paradox, I believe, because one would not expect this superior ability of God to impede His position—the outcome He can ensure—in a game. Yet, it is precisely God's omniscience which ensures that man obtains his best outcome and that God does not.

In order to avoid misunderstanding, I want to stress that I have made only the usual assumptions about the play of the game I have described: both players (i) have complete information about the outcome matrix, (ii) make simultaneous choices, (iii) cannot communicate with each other, and (iv) are rational in the sense of seeking to achieve their best possible outcomes. The only assumptions I have added to these in the play of Chicken is that (v) God is omniscient and (vi) man is aware of His omniscience.

To try to resolve the paradox of omniscience, I want first to relax assumption (iii) and then allow for the possibility of retaliation by God, or retaliation by both players in a sequence of moves and countermoves. I suggest that this reformulation of the game permits at least three possible ways for God to counter the paradox:

I. *Before play of the game, God can threaten man*: "*If you compromise, I will*; *otherwise, I won't*." This threat is not credible for two reasons: (i) if man compromises, and God predicts this, it is not in God's interest to compromise, too, obtaining His next-best outcome (3) at CC; rather, God should not compromise and obtain His best outcome (4) at $C\bar{C}$; (ii) if man does not compromise, and God predicts this, God has no alternative but to compromise to prevent His worst outcome (1) at $\overline{CC}$ from occurring. Therefore, man should not compromise.

In the next two possible solutions, I assume that not only is communication permissible—making threats possible—before the play of the game, but also that one or both players can retaliate after the play of the game. In solution II below, I assume that only God can retaliate by switching from His strategy of C to $\bar{C}$, or vice versa. In solution III, I assume that both players can retailiate by switching strategies in a strict sequence—first one, then the other, and so on—without the possibility of retracting a move once it is made. When one player has no interest in making a further move, the game terminates and its outcome is where play stops. Here are the two possible solutions, with an analysis of their merits:

II. *Before play of the game, God can threaten man*: "*Choose the strategy that you wish, but I reserve the right to retaliate by switching my strategy after the play of the game.*" This is again an incredible threat because, by virtue of His omniscience, God can predict man's choice and make His own best response to it. Reserving the right to move after the play of the game gives God no additional advantage.

III. *Before play of the game, God can threaten man*: "*Choose the strategy that you wish, but we each will have the right to move and countermove an indefinite number of times until one of us stops, whence the outcome reached is implemented.*" I shall now show that outcome (3,3) is the *only* outcome from which neither player will have a desire to move if it is chosen. Moreover, if any of the other three outcomes is chosen, moves and countermoves will terminate at an outcome inferior to (3,3) for at least one player, thereby establishing the choice of compromise by one, and then of necessity the other, as their only rational choice. Hence, (3,3) is engendered under the specified rules.

Assume (4,2) is chosen, which is the outcome in the original game and creates the paradox of omniscience. Then God, by moving the process to (1,1), would induce man to move it to (2,4), where the process would stop since it is best for God. But this outcome is next-worst for man.

Assume (2,4) is chosen. Then analogous moves by man and God, respectively, would move the process to (4,2), which is next-worst for God.

Assume (1,1) is chosen. Then the player who moves first (man or God) would get his next-worst outcome [(2,4) or (4,2)], and the other player would obtain his best outcome and have no incentive to countermove. In either event, or if no player moved initially, at least one player would obtain an outcome inferior to his next-best outcome 3.

Now assume (3,3) is chosen. Then if God initiates the first move, the process will proceed as follows: $(3,3)\to(2,4)\to(1,1)\to(4,2)$, so the final outcome, (4,2), is inferior to (3,3) for God. Similarly, if man initiates the first move, the process will proceed as follows: $(3,3)\to(4,2)\to(1,1)\to(2,4)$, so the final outcome, (2,4), is inferior to (3,3) for man. Hence, neither man nor God will, if they both choose compromise, have an incentive to deviate initially from (3,3).

Because the three other outcomes will (eventually) result in an outcome's being implemented that is inferior to 3 for the player who does *not* choose compromise, each player will have an incentive to choose compromise. Since the resulting (3,3) outcome, given the rules I have specified in the third solution, is stable, rational players who desire to ensure their best possible outcome will choose it.

This solution needs to be slightly qualified. If both players choose not to compromise, resulting in (1,1), the player who can hold out longer, forcing his opponent to move first, obtains his best outcome 4. Presumably this would be God in most situations. But whoever it is, the other player has a foolproof counterstrategy—compromise. For if the recalcitrant player (say, God) does not, His best outcome, (2,4), is transformed by one move and one countermove into (4,2), which is next-worst for Him. Thus, He should compromise, too.

More generally, as long as moves are strictly sequential and no backtracking is allowed, the unique stable (and desirable) outcome for both players is (3,3), whoever should deviate first. Even if one player believes he can hold out longer at (1,1), should it be chosen, a common perception of this fact by both players will lead one—and hence both, for reasons given above —to choose compromise. Only if perceptions of the two players differ—each thinks he can force the other to capitulate first when they are at (1,1)—will compromise not be the rational choice of both players. But I assume this is not the case—the players' perceptions agree.

Thus, given the rules of solution III, the outcome (3,3) is implemented. Paradoxically, perhaps, the possibility of moves and countermoves by both players—not just God's ability to retaliate unilaterally—enables God to counter the paradox His omniscience induces in the play of Chicken. Note that both the verbal threat of solution I, and the threat of possible retaliation after the play of the game of solution II, are insufficient to resolve the paradox. Only when reprisals by *both* players are permitted is God able to obliterate the disadvantage His omniscience encumbered Him with in the original play of Chicken.

A formalization of solution III, including an analysis of its existence and properties in different classes of games, is given in **[8]**, and available from the author on request.

With the possible exception of recent work on "Newcomb's problem," much of which is cited in **[3]**, **[5]**, and **[9]**, I know of no serious applications of game theory or decision theory to theological questions. Although Newcomb's problem involves some controversial probabilistic assumptions, the aforementioned works strongly suggest that it is essentially a Prisoners' Dilemma, game theory's most famous game.

In my opinion, there is nothing uniquely theological about the pathology of Prisoners' Dilemma. Similarly, the game of Chicken, which, like Prisoners' Dilemma, is one of 78 distinct 2×2 ordinal games (two-person games in which each player has two strategies and can order the resulting four outcomes from best to worst), has no special theological properties, but the assumption of omniscience very definitely does. Although it says nothing about the existence of God—which I regard as a question no scientific or mathematical theory can resolve—it does show up what I believe is a surprising consequence of omniscience: the player endowed with this special quality in a game may be aggrieved because he possesses it.

If this player is God, it is not evident how a relaxation of the rules of the game can eliminate His problem, as the first two suggested solutions to the paradox of omniscience demonstrated. But one set of rules, allowing for moves and countermoves by *both* God and man, does offer a solution, implying that God may need man as much as vice versa to help resolve certain difficulties. There is, in my view, an implicit recognition of this fact in the Bible: endowing man with free will, God expects him to act not just in his own interest but in God's as well. Put another way, the expectation is that games will be nonzero-sum, whereby all players can benefit by cooperating with each other.

An examination of the 78 2×2 games, which are enumerated by Rapoport and Guyer in **[10]** and Brams in **[2]**, reveals that six (including Chicken) are vulnerable to the paradox of omniscience. These six (Nos. 64–69 in Rapoport and Guyer, Nos. 68–73 in Brams) are listed in FIGURE 2. In each case, the row player (man), by choosing his second strategy, forces the omniscient column player (God) to choose a strategy that results in the best outcome for man but an inferior outcome for God. Although there are other games in which an outcome best for man is inferior for God, these six games are the only ones in which neither player has a dominant strategy associated with it and, therefore, the omniscience of God, and man's awareness of it, are required to guarantee its choice.

There are two other 2×2 games (Nos. 55 and 72 in **[10]**, Nos. 65 and 76 in **[2]**), listed in FIGURE 3, in which man can force God to choose an outcome of lower rank than his, but this outcome, (3,2), is next-best, not best, for man. Moreover, in each of these games, there is another outcome, (4,3), better for *both* players but, paradoxically, unattainable if God is omniscient and man knows this.

(3,4)	(2,1)
(1,2)	(4,3)

(2,4)	(3,1)
(1,2)	(4,3)

(2,4)	(3,3)
(1,1)	(4,2)

(3,4)	(2,3)
(1,1)	(4,2)

(3,4)	(2,2)
(1,1)	(4,3)

(3,4)	(1,1)
(2,2)	(4,3)

The six 2×2 games vulnerable to the paradox of omniscience. In each case, as in FIGURE 1, the row player is man and the column player is an omniscient God.

FIGURE 2.

(1,4)	(4,3)
(2,1)	(3,2)

(2,4)	(4,3)
(1,1)	(3,2)

Two games in which man can beat God but not obtain his best outcome.

FIGURE 3.

The second of these games has been used to model the 1974 confrontation between President Nixon and Supreme Court Justices Burger and Blackmun over the release of Watergate tapes. There was no omniscient player in this game but, by responding to the Supreme Court decision, Nixon acted with knowledge of the Justices's prior strategy choice, which resulted in the (3,2) outcome (see **[6]** and **[7]**).

Only Chicken has a stable "cooperative" solution in the sense of solution III—both players, by choosing it, do better than at least one would do if any other outcome were chosen. Hence, among the six games in FIGURE 2 that are vulnerable to the paradox of omniscience, only in Chicken (third game in the first row) is there an incentive for both players to compromise their differences and settle on the cooperative outcome, even though it is not the best outcome for either player.

This conclusion applies whether one player has omniscience or not. But what makes solution III especially poignant for God is that without it His omniscience, and man's awareness of it, would assure Him of the unsatisfactory outcome of 2.

I believe that game theory can help render more precise the consequences that arise from games man may play with God. It may thereby not only shed light on certain theological questions but also expose better the nonobvious implications of different assumptions various religions make about man's relationship to God. This, I venture to say, would not be a mean achievement for mathematics.

References

[1] Steven J. Brams, Biblical Games: A Strategic Analysis of Stories in the Old Testament, MIT Press, Cambridge, MA, 1980.
[2] ______, Deception in 2×2 games, Journal of Peace Science, 2 (Spring 1977) 171–203.
[3] ______, Paradoxes in Politics: An Introduction to the Nonobvious in Political Science, Free Press, New York, 1976.
[4] ______, Game Theory and Politics, Free Press, New York, 1975.
[5] ______, Newcomb's problem and prisoners' dilemma, Journal of Conflict Resolution, 19 (December 1975) 596–612.
[6] ______, The Presidential Election Game, Yale University Press, New Haven, CT, 1978.
[7] ______ and Douglas Muzzio, Unanimity in the Supreme Court: A game-theoretic explanation of the decision in the White House tapes case, Public Choice, 32 (Winter 1977) 67–83.
[8] ______ and Donald Wittman, Nonmyopic equilibria, mimeographed, 1980.
[9] David Lewis, Prisoners' dilemma is a Newcomb problem, Philosophy & Public Affairs, 8 (Spring 1979) 235–240.
[10] Anatol Rapoport and Melvin Guyer, A Taxonomy of 2×2 Games, General Systems: Yearbook of the Society for General Systems Research, 11 (1966) 203–214.

A Budget of Boxes

JOHN B. FRIEDLANDER
JOHN B. WILKER
University of Toronto
West Hill, Ontario
Canada M1C 1A4

So, Nat'ralists observe, a Flea
Hath smaller Fleas that on him prey,
And these have smaller Fleas to bite 'em,
And so proceed ad infinitum.
—Jonathan Swift (1667–1745)
On Poetry: A Rhapsody (1733)

"You have a square sheet of cardboard with edge length 1. What is the volume of the largest open box you can manufacture by cutting squares out of the corners of your sheet and folding up the flaps?"

This question must be answered nearly a million times a year by calculus students from every corner of the globe. Seeing such a large audience of interested persons we naturally felt encouraged to write a paper on the subject. The topical theme of conservation occurred to us because there is a needless waste of hypothetical cardboard and pedagogical opportunity if four million square corners are being thrown away every year. In order to end this waste we would like to propose a scheme for recycling these squares once, twice or even infinitely often.

games man may play with God. It may thereby not only shed light on certain theological questions but also expose better the nonobvious implications of different assumptions various religions make about man's relationship to God. This, I venture to say, would not be a mean achievement for mathematics.

References

[1] Steven J. Brams, Biblical Games: A Strategic Analysis of Stories in the Old Testament, MIT Press, Cambridge, MA, 1980.
[2] ______, Deception in 2×2 games, Journal of Peace Science, 2 (Spring 1977) 171–203.
[3] ______, Paradoxes in Politics: An Introduction to the Nonobvious in Political Science, Free Press, New York, 1976.
[4] ______, Game Theory and Politics, Free Press, New York, 1975.
[5] ______, Newcomb's problem and prisoners' dilemma, Journal of Conflict Resolution, 19 (December 1975) 596–612.
[6] ______, The Presidential Election Game, Yale University Press, New Haven, CT, 1978.
[7] ______ and Douglas Muzzio, Unanimity in the Supreme Court: A game-theoretic explanation of the decision in the White House tapes case, Public Choice, 32 (Winter 1977) 67–83.
[8] ______ and Donald Wittman, Nonmyopic equilibria, mimeographed, 1980.
[9] David Lewis, Prisoners' dilemma is a Newcomb problem, Philosophy & Public Affairs, 8 (Spring 1979) 235–240.
[10] Anatol Rapoport and Melvin Guyer, A Taxonomy of 2×2 Games, General Systems: Yearbook of the Society for General Systems Research, 11 (1966) 203–214.

A Budget of Boxes

JOHN B. FRIEDLANDER
JOHN B. WILKER
University of Toronto
West Hill, Ontario
Canada M1C 1A4

So, Nat'ralists observe, a Flea
Hath smaller Fleas that on him prey,
And these have smaller Fleas to bite 'em,
And so proceed ad infinitum.
—Jonathan Swift (1667–1745)
On Poetry: A Rhapsody (1733)

"You have a square sheet of cardboard with edge length 1. What is the volume of the largest open box you can manufacture by cutting squares out of the corners of your sheet and folding up the flaps?"

This question must be answered nearly a million times a year by calculus students from every corner of the globe. Seeing such a large audience of interested persons we naturally felt encouraged to write a paper on the subject. The topical theme of conservation occurred to us because there is a needless waste of hypothetical cardboard and pedagogical opportunity if four million square corners are being thrown away every year. In order to end this waste we would like to propose a scheme for recycling these squares once, twice or even infinitely often.

Just to be sure we are proceeding with a common understanding and an agreed notation, let us solve the original problem one more time. If we cut squares of edge length λ, $0<\lambda<\frac{1}{2}$, from the given sheet, we can form an open box of volume $V_1(\lambda)=\lambda(1-2\lambda)^2$. For a maximum,

$$\frac{dV_1}{d\lambda}=12\lambda^2-8\lambda+1=0,$$

so

$$\lambda=\frac{8\pm\sqrt{64-48}}{24}=\frac{1}{6}\text{ or }\frac{1}{2}.$$

This gives $\lambda_1=1/6=.166\ldots$ and a maximum volume of $\bar{V}_1=V_1(\lambda_1)=2/27=.074\ldots.$

Now, what about the two-stage problem? Here we begin by cutting squares of edge length λ from our sheet to make a box of volume $\lambda(1-2\lambda)^2$. But then, instead of throwing away the squares of edge length λ, we manufacture four more boxes by cutting squares of edge length $\mu_i\lambda$, $i=1,2,3,4$, from the ith of these residual squares. The objective, of course, is to make the total volume of our five boxes as large as possible.

We observe that each of the μ_i must be equal to λ_1 because at the second stage of the two-stage problem we are confronted with four scaled-down versions of the one-stage problem. The best yield from the second stage is therefore $4\lambda^3\bar{V}_1$ and the total volume which can be produced by working the two-stage problem this way is

$$\begin{aligned}V_2(\lambda)&=\lambda(1-2\lambda)^2+4\lambda^3\bar{V}_1\\&=4(\bar{V}_1+1)\lambda^3-4\lambda^2+\lambda.\end{aligned}$$

For a maximum

$$\frac{dV_2}{d\lambda}=12(\bar{V}_1+1)\lambda^2-8\lambda+1=0,$$

so

$$\lambda=\frac{2\pm\sqrt{1-3\bar{V}_1}}{6(\bar{V}_1+1)}.$$

Both these roots lie in the interval $0<\lambda<\frac{1}{2}$, but the smaller root gives the maximum because $V_2(\lambda)$ is a cubic with leading coefficient $4(\bar{V}_1+1)>0$. This gives $\bar{V}_2=V_2(\lambda_2)$, where

$$\lambda_2=\frac{2-\sqrt{1-3\bar{V}_1}}{6(\bar{V}_1+1)}$$

optimizes the first stage.

Now we move along to the n-stage problem. Here we are interested in maximizing the total volume of $1+4+4^2+\cdots+4^{n-1}=\frac{1}{3}(4^n-1)$ boxes cut out in n stages. When we cut out the first box of volume $\lambda(1-2\lambda)^2$ we leave four scaled-down $(n-1)$-stage problems with a maximum yield of $4\lambda^3\bar{V}_{n-1}$. The total volume which can be achieved by working the n-stage problem this optimal way is therefore

$$V_n(\lambda)=4(\bar{V}_{n-1}+1)\lambda^3-4\lambda^2+\lambda;$$

for a maximum, we solve

$$\frac{dV_n}{d\lambda}=12(\bar{V}_{n-1}+1)\lambda^2-8\lambda+1=0,$$

yielding $\bar{V}_n=V_n(\lambda_n)$ where

$$\lambda_n=\frac{2-\sqrt{1-3\bar{V}_{n-1}}}{6(\bar{V}_{n-1}+1)}.$$

It is clear that the sequence $\bar{V}_n$ is increasing so that $\bar{V}_n \geq \bar{V}_1 = 2/27$. On the other hand, the fact that λ_{n+1} is real implies that $\bar{V}_n \leq 1/3$, and so $2/27 \leq \bar{V}_n \leq 1/3$. The same bounds apply to $\bar{V}$, the limit of the increasing sequence $\bar{V}_n$.

It is intuitively plausible that λ_n also increases with n. The larger n gets, the more it pays to leave material for subsequent stages and not be too greedy for volume in making the first box. One way to verify this conjecture is to use the equation

$$12(\bar{V}_{n-1}+1)\lambda_n^2 - 8\lambda_n + 1 = 0$$

in the formula

$$\bar{V}_n = 4(\bar{V}_{n-1}+1)\lambda_n^3 - 4\lambda_n^2 + \lambda_n$$

to show that

$$\bar{V}_n = \frac{2}{3}\lambda_n(1-2\lambda_n).$$

Since $y = \frac{2}{3}x(1-2x)$ is a parabola opening downwards with its vertex at $(1/4, 1/12)$, we get the improved bound $2/27 \leq \bar{V}_n \leq 1/12$. This allows us to show that

$$\lambda_n \leq \frac{2-\sqrt{1-\frac{1}{4}}}{6\left(\frac{2}{27}+1\right)} = \frac{9(4-\sqrt{3})}{116} < \frac{1}{4},$$

from which it follows that λ_n and $\bar{V}_n$ increase together along the parabola. Hence $\frac{1}{6} \leq \lambda_n < \frac{1}{4}$. The same bounds apply to $\bar{\lambda}$, the limit of the increasing sequence λ_n.

In order to find $\bar{\lambda}$ and $\bar{V}$ we note that

$$\bar{V}_n = \frac{2}{3}\lambda_n(1-2\lambda_n)$$

implies that

$$\bar{V} = \frac{2}{3}\bar{\lambda}(1-2\bar{\lambda})$$

and

$$\bar{V}_n = 4(\bar{V}_{n-1}+1)\lambda_n^3 - 4\lambda_n^2 + \lambda_n$$

implies that

$$\bar{V} = 4(\bar{V}+1)\bar{\lambda}^3 - 4\bar{\lambda}^2 + \bar{\lambda}$$

from which it follows that

$$\bar{V} = \bar{\lambda}(1-2\bar{\lambda})^2(1-4\bar{\lambda}^3)^{-1}.$$

Comparing the two expressions for $\bar{V}$, we find that $8\bar{\lambda}^3 - 6\bar{\lambda} + 1 = 0$; some elementary curve tracing shows that this cubic has a unique zero in the required interval. The roots of $x^3 + px + q = 0$ are given by Tartaglia's formula

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}$$

where the cube roots are paired so that their product is $-p/3$. In particular, the roots of $8\bar{\lambda}^3 - 6\bar{\lambda} + 1 = 0$ are $-\cos 20°$, $\cos 80° = \sin 10°$ and $\cos 40°$ and so

$$\bar{\lambda} = \sin 10° = .1736\ldots$$

and

$$\bar{V} = \frac{2}{3}\bar{\lambda}(1-2\bar{\lambda}) = .0756\ldots.$$

An alternative way to determine $\bar{\lambda}$ is to use the equation

$$12(\bar{V}_{n-1}+1)\lambda_n^2-8\lambda_n+1=0$$

to show that

$$\bar{V}_{n-1}=(6\lambda_n-1)(1-2\lambda_n)(12\lambda_n^2)^{-1}$$

and then to compare this equation with

$$\bar{V}_n=\frac{2}{3}\lambda_n(1-2\lambda_n)$$

to obtain the instructive formula

$$\bar{V}_n=8\lambda_n^3(6\lambda_n-1)^{-1}\bar{V}_{n-1}.$$

In the limit, $8\bar{\lambda}^3(6\bar{\lambda}-1)^{-1}=1$ and we obtain the same cubic equation for $\bar{\lambda}$ as we had before.

Finally, let us consider infinite recycling: instead of stopping at stage n, we continue to make new boxes out of the scraps **ad infinitum**. We seek a scheme which maximizes the total volume of the infinitely many boxes produced this way.

An infinite scheme can be described by a tree opening downwards and splitting four ways at each node. The nodes correspond to boxes and are marked with a ratio $0<\mu<\frac{1}{2}$. The product of the ancestral μ's gives the size of the square from which a box is cut and the μ on the node itself describes the actual cutting of that particular box. If the nodes are labelled by indices $i \in I$ which are ordered so that $j<i$ if j is an ancestor of i, then the total volume of the boxes produced by the scheme is given by

$$V=\sum_{i\in I}\Big(\prod_{j<i}\mu_j\Big)^3\mu_i(1-2\mu_i)^2.$$

By rearranging this into a sum over successive stages, we get the estimate

$$V\leqslant\sum_{n=1}^{\infty}4^{n-1}\left[\left(\frac{1}{2}\right)^{n-1}\right]^3\frac{2}{27}$$
$$=\frac{2}{27}\sum_{n=0}^{\infty}\left(\frac{1}{2}\right)^n=\frac{4}{27}.$$

This shows that the volumes are bounded above and have a finite sup. It would appear to be rather difficult to prove directly that this sup is attained. However our preliminary work on the n-stage problem will help us to see that the sup is actually $\bar{V}$ and it is attained by the single scheme with $\mu_i=\bar{\lambda}$ for all i. This contrasts sharply with the optimal scheme for the n-stage problem where the cutting ratios λ_i are strictly decreasing in successive stages.

To prove our results about the infinite-stage problem, we first of all let $V(\lambda)$ be the volume produced by the scheme with all the μ_i equal, say $\mu_i=\lambda$, $0<\lambda<\frac{1}{2}$. By considering the first stage of these schemes and noticing that they leave four schemes similar to the original, we obtain

$$V(\lambda)=\lambda(1-2\lambda)^2+4\lambda^3V(\lambda),$$

or

$$V(\lambda)=\lambda(1-2\lambda)^2(1-4\lambda^3)^{-1}.$$

For a maximum,

$$\frac{dV}{d\lambda}=(1-2\lambda)(1-6\lambda+8\lambda^3)(1-4\lambda^3)^{-2}=0,$$

and since the only acceptable root is $\lambda=\bar{\lambda}$, we obtain $V(\bar{\lambda})=\bar{V}$ for our maximum. This shows that $\bar{V}$ is attained as described.

Now consider an arbitrary scheme with volume V. For any $\epsilon>0$ we can find an integer n so that the boxes from the first n stages of this scheme have volume at least $V-\epsilon$. Since these boxes belong to an n-stage problem, their volume is at most $\bar{V}_n$ and therefore less than $\bar{V}$. It follows that $V<\bar{V}+\epsilon$ and, since ϵ is arbitrary, $V\leqslant\bar{V}$. This shows that $\bar{V}$ is indeed the maximum volume in the infinite-stage problem.

Finally, if a scheme produces the maximum volume $\bar{V}$, then all of the subschemes belonging to any stage must produce $\bar{V}$ times their scaling factor or they could be replaced with a definite improvement. Without loss of generality we consider the μ of the first stage. It must satisfy

$$\bar{V}=\mu(1-2\mu)^2+4\mu^3\bar{V}$$

or

$$4(\bar{V}+1)\mu^3-4\mu^2+\mu-\bar{V}=0,$$

and this equation suffices to prove that $\mu=\bar{\lambda}$. This follows because $\bar{\lambda}$ is a double root and therefore the remaining root must be

$$\frac{\bar{V}}{4(\bar{V}+1)\bar{\lambda}^2}=.5827\ldots>\frac{1}{2}.$$

This completes the proof that the only scheme with volume $\bar{V}$ is the one in which $\mu_i=\bar{\lambda}$ for all i.

The Uniformity Assumption in the Birthday Problem

GEOFFREY C. BERRESFORD
Long Island University
Greenvale, NY 11548

The birthday problem, to find the probability that in a group of n people some two will share a common birthday, has occurred frequently in the literature since having been proposed in 1939 by von Mises. It is easily solved under the assumptions that each person's birthday is determined independently, and that the 365 possible birthdays (ignoring leap years) are equally likely. Under these independence and uniformity assumptions it is easy to show that the probability of a shared birthday reaches $\frac{1}{2}$ as soon as the size of the group reaches 23.

The reason for the uniformity assumption is interesting. Depending on the population, it may or may not be a reasonable approximation to reality, but in any case it is enormously convenient. To see this, let us consider the problem without assuming uniformity. To take full account of the 365 probabilities of being born on different days of the year, we let p_i be the probability of being born on day i, $i=1,\ldots,365$, and obtain the (complementary) probability of n independently chosen people all having different birthdays as

$$P(n)=n!\sum_{i_1<\cdots<i_n} p_{i_1}\cdots p_{i_n}, \tag{1}$$

the sum being over all n-subsets of $\{1,2,\ldots,365\}$ such that $i_1<i_2<\cdots<i_n$. The difficulty is that the sum has $\binom{365}{n}$ terms, and for group size $n=23$ this is $\binom{365}{23}\approx 10^{36}$ terms, which even the fastest computer would need 10^{20} centuries to calculate.

Nevertheless, two observations can be made, one theoretical, one empirical. *In a group of n people, the probability of a shared birthday is least for the uniform distribution.* Therefore, regardless of the actual distribution of birthdays, a group size of 23 is sufficient to make a shared birthday more probable than not. There are proofs of this in the literature, but the following

Now consider an arbitrary scheme with volume V. For any $\epsilon > 0$ we can find an integer n so that the boxes from the first n stages of this scheme have volume at least $V-\epsilon$. Since these boxes belong to an n-stage problem, their volume is at most $\bar{V}_n$ and therefore less than $\bar{V}$. It follows that $V < \bar{V}+\epsilon$ and, since ϵ is arbitrary, $V \leqslant \bar{V}$. This shows that $\bar{V}$ is indeed the maximum volume in the infinite-stage problem.

Finally, if a scheme produces the maximum volume $\bar{V}$, then all of the subschemes belonging to any stage must produce $\bar{V}$ times their scaling factor or they could be replaced with a definite improvement. Without loss of generality we consider the μ of the first stage. It must satisfy

$$\bar{V} = \mu(1-2\mu)^2 + 4\mu^3\bar{V}$$

or

$$4(\bar{V}+1)\mu^3 - 4\mu^2 + \mu - \bar{V} = 0,$$

and this equation suffices to prove that $\mu = \bar{\lambda}$. This follows because $\bar{\lambda}$ is a double root and therefore the remaining root must be

$$\frac{\bar{V}}{4(\bar{V}+1)\bar{\lambda}^2} = .5827\ldots > \frac{1}{2}.$$

This completes the proof that the only scheme with volume $\bar{V}$ is the one in which $\mu_i = \bar{\lambda}$ for all i.

The Uniformity Assumption in the Birthday Problem

GEOFFREY C. BERRESFORD
Long Island University
Greenvale, NY 11548

The birthday problem, to find the probability that in a group of n people some two will share a common birthday, has occurred frequently in the literature since having been proposed in 1939 by von Mises. It is easily solved under the assumptions that each person's birthday is determined independently, and that the 365 possible birthdays (ignoring leap years) are equally likely. Under these independence and uniformity assumptions it is easy to show that the probability of a shared birthday reaches $\frac{1}{2}$ as soon as the size of the group reaches 23.

The reason for the uniformity assumption is interesting. Depending on the population, it may or may not be a reasonable approximation to reality, but in any case it is enormously convenient. To see this, let us consider the problem without assuming uniformity. To take full account of the 365 probabilities of being born on different days of the year, we let p_i be the probability of being born on day i, $i = 1,\ldots,365$, and obtain the (complementary) probability of n independently chosen people all having different birthdays as

$$P(n) = n! \sum_{i_1 < \cdots < i_n} p_{i_1} \cdots p_{i_n}, \tag{1}$$

the sum being over all n-subsets of $\{1,2,\ldots,365\}$ such that $i_1 < i_2 < \cdots < i_n$. The difficulty is that the sum has $\binom{365}{n}$ terms, and for group size $n = 23$ this is $\binom{365}{23} \approx 10^{36}$ terms, which even the fastest computer would need 10^{20} centuries to calculate.

Nevertheless, two observations can be made, one theoretical, one empirical. *In a group of n people, the probability of a shared birthday is least for the uniform distribution.* Therefore, regardless of the actual distribution of birthdays, a group size of 23 is sufficient to make a shared birthday more probable than not. There are proofs of this in the literature, but the following version of Munford's 1977 proof [3] is perhaps the simplest. We will show that the (complementary) probability $P(n)$ of n independently chosen people having different birthdays is greatest for the uniform distribution. We assume, of course, that $n \geq 2$ and that at least n of the p_i's are nonzero, so that n different birthdays are possible. We will show how $P(n)$ changes if we alter the values of two unequal probabilities, say p_1 and p_2, by replacing them with their common mean, $\frac{1}{2}(p_1+p_2)$. Let us partition the sum (1) into three parts: those terms containing both p_1 and p_2, those terms containing just one of p_1 and p_2, and those terms containing neither p_1 nor p_2. Let us also factor out p_1 and p_2 whenever they occur. We may then express $P(n)$ as

$$P(n)=n!\Bigg(p_1p_2\sum_{2<i_1<\cdots<i_{n-2}} p_{i_1}\cdots p_{i_{n-2}}+(p_1+p_2)\sum_{2<i_1<\cdots<i_{n-1}} p_{i_1}\cdots p_{i_{n-1}} + \sum_{2<i_1<\cdots<i_n} p_{i_1}\cdots p_{i_n}\Bigg). \tag{2}$$

If we now replace both p_1 and p_2 by their common mean, $\frac{1}{2}(p_1+p_2)$, thus leaving their sum unchanged, only the first term in (2) changes, in which we must replace p_1p_2 by $(\frac{1}{2}(p_1+p_2))^2$. But since $p_1p_2<(\frac{1}{2}(p_1+p_2))^2$, (taking the square root of each side, this is merely the statement that the geometric mean of two unequal numbers is less than their arithmetic mean) this replacement will only increase the value of $P(n)$. Thus since the probability $P(n)$ of different birthdays can be increased by this operation whenever two p_i's are unequal, it must be greatest when all the p_i's are equal, that is, for the uniform distribution.

The second observation involves comparing the uniformity assumption with actual data. FIGURE 1 is a graph of the empirical probabilities of a birth occurring on any day of the year 1977 for the 239,762 live births in New York State (source: New York State Health Department). I leave it to the reader to surmise reasons for the obvious weekly cyclical component. The empirical probabilities vary from a low of .002135 (on Sunday, December 11th) to a high of .003478 (on Wednesday, July 6th), a variation of almost 27% from the mean of 1/365. Thus for a population born in a given year (the type of population from which most school classes are drawn) the assumption of uniformity is not valid. However, because a given birthday will fall on different days of the week in different years (since 365 is relatively prime to 7) in a population of mixed ages the weekly cycle will be averaged out. For such a population uniformity will be a reasonable assumption, as is shown by the graph in FIGURE 2, in which the data is as in FIGURE 1 except that each daily probability has been averaged with the six following it to remove the weekly cycle. The variation here is only about 10% above and below the mean of 1/365.

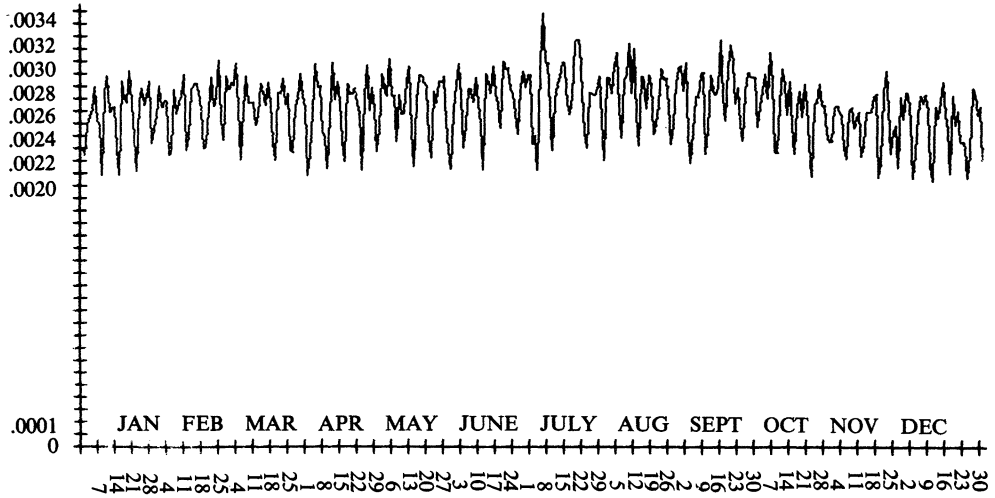


FIGURE 1.

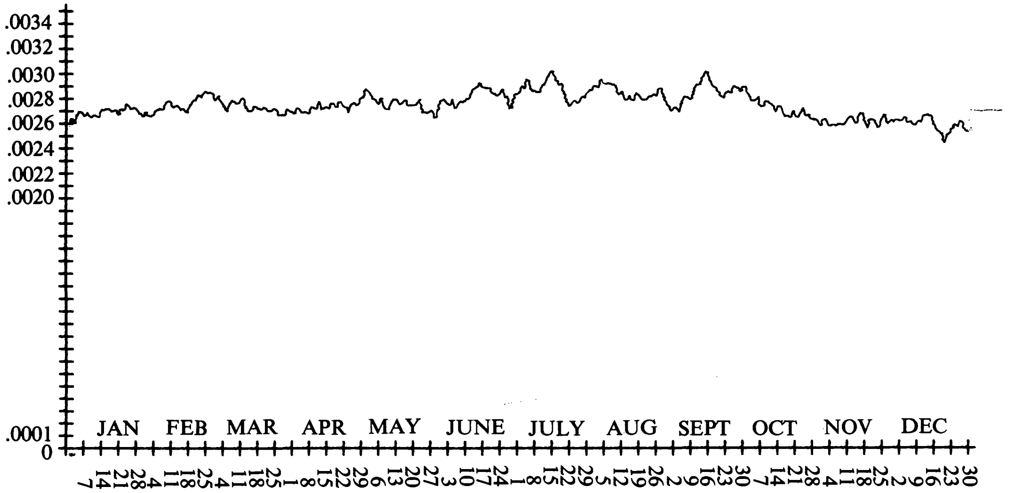


FIGURE 2.

Although the birthday problem with 365 different probabilities is hopelessly intractable, it becomes manageable if one allows a smaller number, m, of different probabilities, $p_1, \ldots, p_m$, and rounds each of the actual empirical probabilities to the nearest of these. Then if d_i is the number of days being assigned probability p_i, subject to the obvious consistency relations $\Sigma d_i = 365$, $\Sigma d_i p_i = 1$, the probability of n people all having different birthdays is

$$n! \sum_{n_1 + \cdots + n_m = n} \prod_{i=1}^{m} \binom{d_i}{n_i} p_i^{n_i},$$

where the sum is over the $\binom{m+n-1}{n}$ different m-tuples $(n_1, \ldots, n_m)$ of nonnegative integers satisfying $n_1 + \cdots + n_m = n$ (see Feller [2], p. 38).

Size (n)	Probability of a Shared Birthday	
of Group	Uniform Case	Non-Uniform Case
12	.1670	.1683
15	.2529	.2537
18	.3469	.3491
21	.4437	.4463
22	.4757	.4783
23	.5073	.5101

TABLE 1.

A computer calculation (performed on a Univac 1100/82 with 18 digit precision) using $m = 10$ different probabilities based on the empirical probabilities from the 1977 New York State data graphed in FIGURE 1 shows that the probability of a shared birthday is surprisingly robust: a group size of 23 is still required to raise the probability above one-half (see TABLE 1).

I would like to thank the C. W. Post Computer Center and the C. W. Post Research Committee for their assistance.

References

[1] D. M. Bloom, A birthday problem, Amer. Math. Monthly, 80 (1973) 1141–1142.
[2] W. Feller, An Introduction to Probability Theory and Its Applications, vol. 1, 3rd ed., John Wiley, New York, 1968.
[3] A. G. Munford, A note on the uniformity assumption in the birthday problem, Amer. Statist., 31, no. 3 (1977) 119.

The Geometry of Binocular Visual Space

WAYNE M. ZAGE
Ball State University
Muncie, IN 47306

If one were to ask what geometry best describes binocular visual space, most people would immediately reply Euclidean geometry. We do not often ask if the world we perceive by sight (with normal vision in both eyes) is congruent to the physical world around us. And yet daily experiences suggest that these two worlds are not congruent. For example, from the time a person is a young child, he probably notices that parallel railroad tracks appear to converge in the distance. Also, in contexts where we may take physical space to be unbounded, visual space is bounded. Thus the sky appears dome-like, and astronomical objects like the sun or the moon appear to be at relatively small distances from the observer [**9**]. In this paper, which is based on [**14**, Ch. II], we relate the results of experiments in binocular vision to geometric models to arrive at the conclusion that the geometry of binocular visual space is non-Euclidean, and more specifically, hyperbolic.

Certain experiments in binocular vision indicate that the way in which objects are located and described by physical measurement differs from the way in which objects are perceived in visual space. In 1902 F. Hillebrand, who was aware that physically parallel lines are perceived as converging in the distance, conducted an experiment to determine the physical configuration of two lines which appear parallel to a person [**7**]. To discuss this experiment, we specify the position of a point P in physical space by (x,y,z) within an associated cartesian coordinate system fixed with respect to the subject. The subject is positioned so that his left eye (L) is at $+1$ and right eye (R) is at -1 on the y-axis. With this orientation, the x-z plane is the "median plane" between the eyes and the x-y plane is the horizontal plane containing the eyes. In what follows, we are primarily concerned with the horizontal x-y plane. In this plane, the x-axis (as in FIGURE 1) is the "median line" between the eyes.

Hillebrand instructed the subject to help arrange a number of lights into two rows flanking the x-axis, such that the rows were perceived as parallel. To eliminate other visual cues, this experiment was conducted in a darkroom. The subject's head was held fixed, but he was free to vary the point of fixation of his eyes. The results of this experiment are consistent among subjects and are typically of the form shown in FIGURE 1: physically, the curves are neither straight nor parallel.

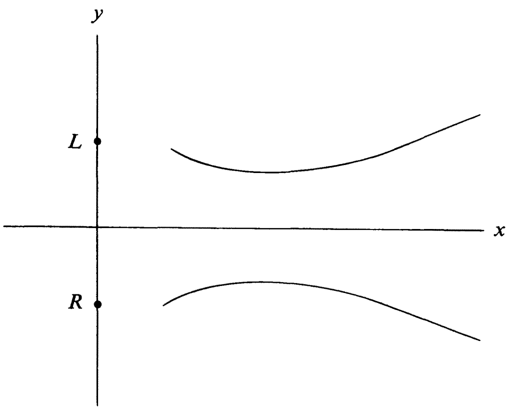
In 1913, W. Blumenfeld modified Hillebrand's experiment and conducted (as they are often called) the classic alley experiments [**3**]. As before, the experiments were conducted in a darkroom, with the subject's head held fixed. However, in this experiment Blumenfeld fixed two light sources P and Q symmetric to the median line in the horizontal plane of the eyes and used two sets of instructions. As other lights were introduced in pairs at points nearer to the subject and flanking the median, the subject was instructed to

(i) "Adjust the lights until the two rows of lights appear to be straight, parallel to each other and parallel to the median."

(ii) "With only the fixed lights P and Q left on, set successive pairs of lights symmetric to the median and at the same apparent separation as the two fixed lights." [**6**]

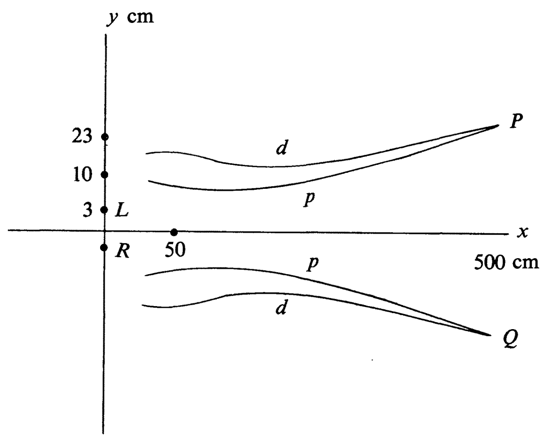
The rows obtained from (i) form the "parallel alley," while those obtained from (ii) form the "distance alley." In the construction of the parallel alley, all of the lights were left on while the subject completed the experiment, whereas in the construction of the distance alley, only P and Q and one other pair of lights were on at any one time.

The typical results of Blumenfeld's alley experiments are shown in FIGURE 2. (These curves, as well as those in FIGURE 1, are adapted from [**10**].) Note first that the parallel curves (p) are not the same as the distance curves (d), and secondly that the parallel curves lie closer to the



Curves resulting from Hillebrand's experiment.

FIGURE 1.



Blumenfeld's parallel and distance alleys.

FIGURE 2.

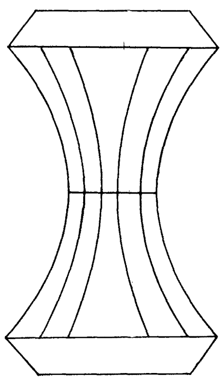
median line than the distance curves. Thus in binocular vision, parallel lines are not equidistant lines. We conclude that *the geometry of binocular visual space is non-Euclidean*. If the geometry were Euclidean, then the two sets of instructions (i) and (ii) would lead to the same experimental result, for only in Euclidean geometry are parallel lines everywhere equidistant.

It is interesting to relate the mathematics of perspective drawings to the results of the alley experiments. When two lines are positioned to give the effect of parallelism in a perspective drawing, they are also meant to be equidistant. Thus perspective drawings are constructed in a Euclidean framework, even though this framework does not agree with the experimental results of the alley experiments. Although lines representing parallels in pictures are drawn as straight, the experimental evidence shows that lines perceived as parallel through binocular vision are indeed curved. However, in spite of these disparities, straight lines seem to be reasonable approximations to the parallel and distance curves, as evidenced by the fact that their use in perspective drawings does give a viewer the expected three-dimensional impression.

We shall use models of hyperbolic and elliptic plane geometry as well as the usual Euclidean geometry of the plane to determine which geometry best fits the results of the parallel and distance alley experiments. In considering only these three cases, we are presupposing that visual space is a space of constant curvature. This claim has been supported experimentally **[2]** and, in fact, is consistent with certain theoretical considerations. Given an object in binocular visual space, we can perceive another configuration congruent to it at a different location in our field of vision. There is the psychological conviction that the visual shape and size of objects can indeed be repeated in any position and orientation **[9]**. In other words, the shape and form of an object are considered as qualities which are independent of location, and thus there is a free mobility of objects in binocular visual space. It is shown in differential geometry that free mobility is possible if and only if the space has constant curvature **[8**, p. 138]. It follows that the curvature of visual space must be constant and that the geometry of binocular visual space can be represented on a model for either hyperbolic, elliptic or Euclidean plane geometry.

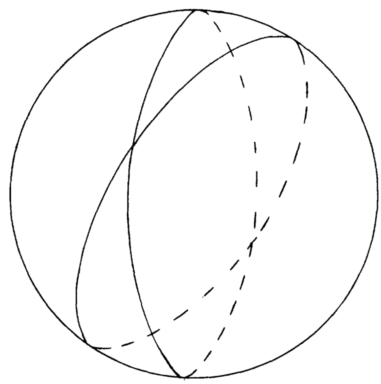
The two-dimensional hyperbolic and elliptic geometries can be represented, at least locally, on surfaces in three-dimensional Euclidean space (E^3). Through any two points on a given surface, or model, there will pass one arc of shortest length belonging to the surface. The proverbial inhabitant of the surface sees this shortest arc as a line in that it is for him the most direct route between the two points. These shortest arcs, when extended in both directions, produce curves of shortest arc length known as **geodesics**. The geodesics of the model are then regarded as the lines of the geometry. In addition, the measure of the angle between two curves on the surface is defined as the measure of the Euclidean angle between the tangents to the curves. In this way, the surface in E^3 becomes a model for hyperbolic or elliptic plane geometry in that all of the postulates of the geometry are realized on that particular surface.

The surface sketched in FIGURE 3, the hyperbolic pseudospherical surface of revolution, or more briefly, the **hyperbolic pseudosphere**, is one model for hyperbolic plane geometry. (The common term "pseudosphere" usually denotes the parabolic pseudospherical surface of revolution.) One family of geodesics on this surface is the family of axial cross sections, which is the set of intersections of the hyperbolic pseudosphere with planes containing its axis. Some of these geodesics are shown in FIGURE 3. The only cross section perpendicular to the axis of this surface which results in a geodesic is the cut which bisects the surface to produce the "horizontal" circle at the "waist." The difference in length between the geodesic segment connecting two points on any other horizontal circle and the circular arc connecting those points increases as the points move away from the "waist" of the surface.



Hyperbolic pseudosphere, with axial cross-sections.

FIGURE 3.



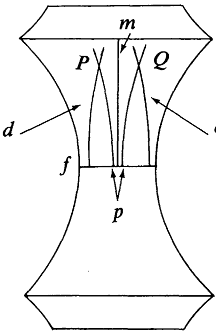
A model for elliptic geometry.

FIGURE 4.

The simplest model locally representing elliptic plane geometry is the surface of the sphere, where the geodesics are the great circles. The postulate in elliptic geometry that no two lines are parallel clearly holds in this model since each pair of great circles intersect (FIGURE 4).

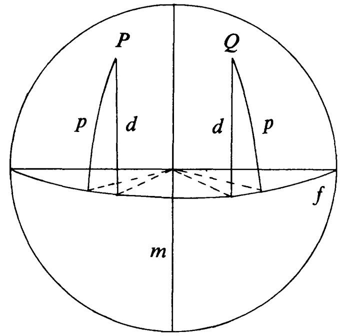
In order to exhibit the results of the alley experiments in geometric models, we must characterize the parallel and distance curves. We assume that the curves which the subject chooses as the parallels are the shortest ones (in his perceptual space) from P and Q to the subjective frontal plane containing his eyes. This assumption allows us to characterize the parallel curves as visual geodesics perpendicular to the frontal plane [**1**]. The distance curves, on the other hand, are characterized as the set of points of constant perceived distance (measured along geodesics) from the median line. Subsequent experiments have supported these interpretations [**1**].

With the above characterizations of the parallel and distance curves, one can correlate experimental data and mathematical considerations. In the hyperbolic model, let the perceived frontal plane f be the geodesic around the "waist" of the hyperbolic pseudosphere, and let the points P and Q be symmetric to the median line of sight m, as shown in FIGURE 5. To construct the distance curve d through P and Q, we slide P and Q along the surface towards the frontal plane f, at a constant distance from m. As we move P and Q towards the "thin" part of the hyperbolic pseudosphere, the distance curves will appear on the model as curves bending away from m, since the distance from d to m must remain constant. On the other hand, the parallel curves p through P and Q are axial cross-sections since these are the geodesics perpendicular to f. These curves will move toward m as they reach f. Therefore, in this model for hyperbolic plane geometry, as the parallel and distance curves reach f, the parallel curves p lie inside the distance curves d.



The parallel and distance curves on a model for hyperbolic geometry.

FIGURE 5.



The parallel and distance curves on a model for elliptic geometry.

FIGURE 6.

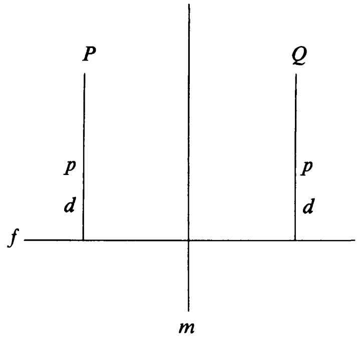
The above intuitive argument leading to the result that p lies inside d on the surface of a hyperbolic pseudosphere may be made rigorous as follows. It can be shown that in the hyperbolic plane two lines cannot have more than one common perpendicular, that this perpendicular is the shortest distance between them, and that the perpendicular distance from a point on one line to the other line increases as the point moves away in either direction from the common perpendicular [**13**, Ch. 4]. In our model, m and p have the common perpendicular f. Let the perpendicular distance from Q to m be ρ. Then by the previous argument, the distance between m and p is less than ρ for all points on p between Q and the line f. On the other hand, all of the points on the distance curve d are at the constant distance ρ from m. Therefore, once again in the model for hyperbolic plane geometry, as p and d reach f, p must lie inside d.

In the elliptic model, let the frontal plane f be the equator of the sphere and let the points P and Q be symmetric to the median line of sight m (a great circle), as shown in FIGURE 6. To construct the distance curves d through P and Q, we slide P and Q down the surface towards the frontal plane f, at a constant geodesic distance from m. Thus the distance curves d will consist of segments of circles (not great circles) perpendicular to the equator. On the other hand, the parallel curves p through P and Q will consist of segments of two great circles perpendicular to f and passing through the poles. Therefore, in a model for elliptic plane geometry, the parallel curves p lie outside the distance curves d. (Note that these "parallel curves," when extended past P and Q, actually intersect. However, recall that the choice of which curves to take as the parallel curves is made not by finding parallel lines through P and Q, but rather by taking the lines through P and Q which are perpendicular to the frontal plane.)

In the Euclidean plane, let the frontal plane f be the x-axis and let the points P and Q be symmetric to the median line of sight m (the y-axis), as shown in FIGURE 7. To construct the distance curves d through P and Q, we slide P and Q down the surface towards the frontal plane f, at a constant distance from m. Thus the distance curves are line segments perpendicular to f. The parallel curves p consist of segments of the geodesics through P and Q perpendicular to f. These are also line segments perpendicular to f. Since in Euclidean plane geometry there exists a unique perpendicular from a given point to a given line, it follows that the parallel curves are exactly the same as the distance curves.

Only in hyperbolic geometry does the parallel alley lie inside the distance alley as was found experimentally. Therefore, we conclude that *the geometry of binocular visual space is best described as hyperbolic geometry*.

Although the alley experiments offer dramatic support for the hyperbolic nature of binocular visual space (under the assumption of constant curvature), researchers are not without disagreement concerning this theory. In fact the views range from those who scarcely want to admit a



The parallel and distauce curves on a model for Euclidean geometry.

FIGURE 7.

geometry of visual space and contend that visual space is a standard Euclidean space congruent to physical space, to those who have been motivated by research such as the alley experiments to produce a mathematical treatment of binocular space perception. At this latter end of the spectrum was Rudolph K. Luneburg who sought an explicit description of the visual space of an individual by using the principle that physical space and visual space may be treated as abstract mathematical systems connected by a transformation. The details of Luneburg's theory can be found in [**14**], which is an exposition of [**9**].

Other researchers, such as Foley, contend that binocular visual space is not a space of constant curvature for all subjects and, therefore, cannot be classified simply as hyperbolic [**4**]. Shipley has questioned the choice of the geodesic for the parallel curves [**11**]. He suggests an alternate definition of p which results in the geometry of binocular visual space being elliptic. The interested reader can find a discussion of these contentions in [**14**, pp. 95–99].

Gogel has studied the tendency for objects, in the absence of effective distance cues, to appear visually at the same distance as each other from the subject [**5**]. Suppes remarks that this "equidistance tendency" is evidence that the geometry of binocular visual space is strongly contextual in character and that the complexity of contextual and individual differences must be accounted for in any theory of visual space. For an excellent survey of these and other theories on the geometry of binocular visual space the reader is referred to [**12**].

Research will undoubtedly continue on the geometry of the space we see. For those who accept the premise that visual space is a space of constant curvature, the alley experiments give conclusive evidence that the geometry of binocular visual space is hyperbolic.

References

[1] A. A. Blank, The Luneburg theory of binocular visual space, J. Opt. Soc. Amer., vol. 43, no. 9 (September, 1953) 717–727.

[2] ______, The Luneburg Theory of Binocular Space Perception, S. Koch (Editor), Study 1, vol. 1: Psychology: A Study of a Science, McGraw-Hill, New York, 1959, pp. 395–426.

[3] W. Blumenfeld, Untersuchungen über die schienbare Grösse in Sehraume, Z. fuer Psychologie, vol. 65, 1913, pp. 241–404.

[4] J. M. Foley, Desarguesian property in visual space, J. Opt. Soc. Amer., vol. 54, no. 5 (May, 1964) 684–692.

[5] W. C. Gogel, Equidistance tendency and its consequences, Psychological Bull., vol. 64, no. 3 (September, 1965) 153–163.

[6] L. H. Hardy, G. Rand, M. C. Rittler, A. A. Blank, and P. Boeder, The Geometry of Binocular Space Perception, New York, Knapp Memorial Laboratories, Institute of Opthalmology, Columbia Univ. College of Physicians and Surgeons, 1953.

[7] F. Hillebrand, Theorie der Scheinbaren Grösse bei binocularem Sehen. Denkschr Acad. Wiss. Wien, Math-Nat. Kl., vol. 72, 1902, pp. 255–307.

[8] D. Laugwitz, Differential and Riemannian Geometry, Academic Press, New York, 1965.

[9] R. K. Luneburg, Mathematical Analysis of Binocular Vision, Princeton Univ. Press, Princeton, NJ, 1947.
[10] ______, Metric Methods in Visual Perception, Studies and Essays, Courant Anniversary Volume, Interscience Publishers, New York, 1948, pp. 215–239.
[11] T. Shipley, Convergence function in binocular visual space II, experimental report, J. Opt. Soc. Amer., vol. 47, no. 9 (September, 1957) 804–821.
[12] P. Suppes, Is visual space Euclidean? Synthese, vol. 35 (1977) 397–421.
[13] H. E. Wolfe, Introduction to Non-Euclidean Geometry, Dryden Press, New York, 1945.
[14] W. M. Zage, Hyperbolic Geometry and Luneburg's Theory of Binocular Visual Space, D. A. dissertation, Univ. of Illinois, Chicago, IL, 1977. (Available through Xerox Univ. Microfilms, order no. 77–22, 586.)

Degenerate Critical Points

THEODORE S. BOLIS
State University College at Oneonta
Oneonta, NY 13820

Up to now the tail of the Taylor series has wagged the dog. One can now tell, at least locally, how much is dog, how much tail, and let the dog do the wagging.

—E. C. Zeeman

We investigate here certain aspects of the local behavior of a function at a degenerate critical point. Primarily we seek ways of determining whether such a point is a local maximum, a local minimum, or a saddle point. The criteria developed do not cover all contingencies, but they are general enough to contain the well-known case of nondegenerate critical points. The basic tool for determining the local behavior of a function at a point of its domain is Taylor's theorem. The problem whose solution we pursue is the following: *Given a point a in the domain of a sufficiently smooth function f, what is the least degree of a Taylor polynomial of f at a that adequately describes the local behavior of f at a?*

Let us begin by establishing our notation and reviewing the standard results which form the first part of a response to our problem. Let f be a sufficiently smooth real valued function defined on some open nonempty subset G of R^n. A point $a \in G$ is a **critical point** of f if all partial derivatives of f at a vanish. This is equivalent to saying that the gradient of f at a is the zero vector.

Before going on we observe that if a point a is *not* a critical point, the nonzero gradient gives rise to a linear function of $x-a$ which is a good local approximation. The formula which expresses this is the formula for the tangent (hyper)plane to the graph of f in R^{n+1} at the point $(a,f(a))$. In fact, under a suitable (curvilinear) change of coordinates, f can be made locally equal to the projection on the first coordinate axis. In short, at a noncritical point the gradient (tangent hyperplane) tells all.

More interesting things happen at a critical point. A first stage classification of a critical point a of f is obtained by looking at the quadratic form given by the **Hessian matrix** $H=(D_iD_jf(a))$, where D_i denotes partial differentiation with respect to x_i. For any symmetric matrix (quadratic form) there are two nonnegative integers associated with it: the **index** λ, which is the number of negative eigenvalues of the matrix, and the **nullity** ν, which is the number of zero eigenvalues (the multiplicity of the eigenvalue zero). Sylvester's law of inertia (cf., for example, p. 387 of **[2]**) states that the pair (λ, ν) is invariant under linear coordinate transformations (i.e., under change

[9] R. K. Luneburg, Mathematical Analysis of Binocular Vision, Princeton Univ. Press, Princeton, NJ, 1947.
[10] ______, Metric Methods in Visual Perception, Studies and Essays, Courant Anniversary Volume, Interscience Publishers, New York, 1948, pp. 215–239.
[11] T. Shipley, Convergence function in binocular visual space II, experimental report, J. Opt. Soc. Amer., vol. 47, no. 9 (September, 1957) 804–821.
[12] P. Suppes, Is visual space Euclidean? Synthese, vol. 35 (1977) 397–421.
[13] H. E. Wolfe, Introduction to Non-Euclidean Geometry, Dryden Press, New York, 1945.
[14] W. M. Zage, Hyperbolic Geometry and Luneburg's Theory of Binocular Visual Space, D. A. dissertation, Univ. of Illinois, Chicago, IL, 1977. (Available through Xerox Univ. Microfilms, order no. 77–22, 586.)

Degenerate Critical Points

Theodore S. Bolis
State University College at Oneonta
Oneonta, NY 13820

Up to now the tail of the Taylor series has wagged the dog. One can now tell, at least locally, how much is dog, how much tail, and let the dog do the wagging.

—E. C. Zeeman

We investigate here certain aspects of the local behavior of a function at a degenerate critical point. Primarily we seek ways of determining whether such a point is a local maximum, a local minimum, or a saddle point. The criteria developed do not cover all contingencies, but they are general enough to contain the well-known case of nondegenerate critical points. The basic tool for determining the local behavior of a function at a point of its domain is Taylor's theorem. The problem whose solution we pursue is the following: *Given a point a in the domain of a sufficiently smooth function f, what is the least degree of a Taylor polynomial of f at a that adequately describes the local behavior of f at a?*

Let us begin by establishing our notation and reviewing the standard results which form the first part of a response to our problem. Let f be a sufficiently smooth real valued function defined on some open nonempty subset G of R^n. A point $a \in G$ is a **critical point** of f if all partial derivatives of f at a vanish. This is equivalent to saying that the gradient of f at a is the zero vector.

Before going on we observe that if a point a is *not* a critical point, the nonzero gradient gives rise to a linear function of $x-a$ which is a good local approximation. The formula which expresses this is the formula for the tangent (hyper)plane to the graph of f in R^{n+1} at the point $(a,f(a))$. In fact, under a suitable (curvilinear) change of coordinates, f can be made locally equal to the projection on the first coordinate axis. In short, at a noncritical point the gradient (tangent hyperplane) tells all.

More interesting things happen at a critical point. A first stage classification of a critical point a of f is obtained by looking at the quadratic form given by the **Hessian matrix** $H=(D_iD_jf(a))$, where D_i denotes partial differentiation with respect to x_i. For any symmetric matrix (quadratic form) there are two nonnegative integers associated with it: the **index** λ, which is the number of negative eigenvalues of the matrix, and the **nullity** ν, which is the number of zero eigenvalues (the multiplicity of the eigenvalue zero). Sylvester's law of inertia (cf., for example, p. 387 of **[2]**) states that the pair (λ,ν) is invariant under linear coordinate transformations (i.e., under change of basis for R^n). We define the **index** and **nullity** of a critical point a of f to be the index and nullity of the Hessian matrix at a, and these numbers are invariant under smooth coordinate transformations of neighborhoods of a because such a transformation T transforms the Hessian quadratic form by means of the Jacobian matrix of T. (This loose statement is justified by nothing more than the chain rule.) A **nondegenerate** critical point is a critical point of zero nullity. Of course, a **degenerate** critical point is a critical point of positive nullity. Since $0 \leqslant \lambda \leqslant n-\nu$ and $0 \leqslant \nu \leqslant n$, it follows that this (λ, ν) classification produces $\frac{1}{2}(n+1)(n+2)$ types of critical points, of which only $n+1$ are nondegenerate. Our first stage classification, as usually shown in advanced calculus, states that a nondegenerate critical point of a twice differentiable function f on an open set G in R^n is a local maximum if and only if $\lambda = n$; a local minimum if and only if $\lambda = 0$; and a saddle point if and only if $0 < \lambda < n$. Most standard textbooks point out that at a degenerate critical point anything can happen and back it up by easily concocted examples.

Our task remains that of shedding some light on the case of degenerate critical points. For this we need some more terminology and notation. A subset K of R^n is a **cone** if $x \in K$ and $t \in R$ imply $tx \in K$, i.e., if the set contains every line through the origin determined by any point in it. A homogeneous function $g: K \to R$ is a function such that $g(tk) = t^p g(k)$ for every $t \in R$ and some fixed p. A homogeneous function is **positive definite**, if for every nonzero $x \in K$ we have $g(x) > 0$; **negative definite**, if $-g$ is positive definite; **nondefinite**, if it is neither positive nor negative definite; **positive semi-definite**, if for every $x \in K$ we have $g(x) \geqslant 0$; **negative semi-definite**, if $-g$ is positive semi-definite; **nonsemi-definite**, if there exist x and $y \in K$ such that $g(x) > 0$ and $g(y) < 0$. For example, if g is a quadratic form on R^n defined by a symmetric matrix of index λ and nullity ν, then g is: positive definite iff $\lambda = 0$ and $\nu = 0$; negative definite iff $\lambda = n$ and $\nu = 0$; nondefinite iff either $\nu > 0$ or $\nu = 0$ and $0 < \lambda < n$; positive semi-definite iff $\lambda = 0$; negative semidefinite iff $\lambda = n - \nu$; and nonsemi-definite iff $0 < \lambda < n - \nu$. If K is a cone in R^n and $\epsilon > 0$, the **ϵ-conic neighborhood** $K(\epsilon)$ of K is defined to be the cone generated by the ϵ-neighborhood in S^{n-1} of $K \cap S^{n-1}$, where S^{n-1} denotes the unit sphere in R^n. This represents a fattening of K by ϵ at radius one; to remain a cone the actual fattening elsewhere is proportional to the distance from the origin.

Finally we introduce a useful notation for Taylor polynomials in many variables. We begin with the terms of degree k which are described by the summation

$$D_x^k f(a) = \sum (k!/i_1!\,i_2! \cdots i_n!) D_1^{i_1} D_2^{i_2} \cdots D_n^{i_n} f(a) x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n},$$

where the summation is carried over all nonnegative indices $i_1, \ldots, i_n$ with $i_1 + \cdots + i_n = k$. Here D_j^0 is the identity operator, $D_j^i = \partial^i/(\partial x_j)^i$ and f is assumed k times differentiable. We can now express Taylor's Theorem in a form suitable for stating and studying our original question clearly (we do not strive for weakest possible hypotheses). Let $F_k(x)$ denote $(1/k!) D_x^k f(a)$. We call $F_k(x)$ the kth Taylor form.

TAYLOR'S THEOREM. *Let $f \in C^{m+1}(G)$, $a \in G$ and $x \in R^n$ with $a + x \in G$. Then $f(a+x) = f(a) + \sum_{k=1}^m F_k(x) + \|x\|^m R_m(x)$, where $\lim_{x \to 0} R_m(x) = 0$.*

For a proof we refer to [1]. Actually, if $P_m(x)$ is a polynomial of degree less than m such that $\lim_{x \to 0}(f(a+x) - P_m(x))/\|x\|^m = 0$, then $P_m(x) = f(a) + \sum_{k=1}^m F_k(x)$. In other words, the Taylor polynomial of f at a with terms of degree less than or equal to m is the unique **order m approximation** to $f(a+x)$.

In what follows, a **saddle point** of f is a critical point of f that is neither a local maximum nor a local minimum for f. Also a will be a critical point of f which is not a **flat point**, i.e., there is a positive integer k such that $F_k \neq 0$, where F_k is defined in the statement of Taylor's Theorem. F_p will denote the first, F_s the second, and F_t the third nonzero Taylor form of f at a critical point a. Obviously $2 \leqslant p < s < t$, and s or t might not exist. The cones which are the zero sets of these homogeneous forms will be denoted by K_p, K_s and K_t respectively, i.e., $K_i = \{x \in R^n : F_i(x) = 0\}$, $i = p, s, t$.

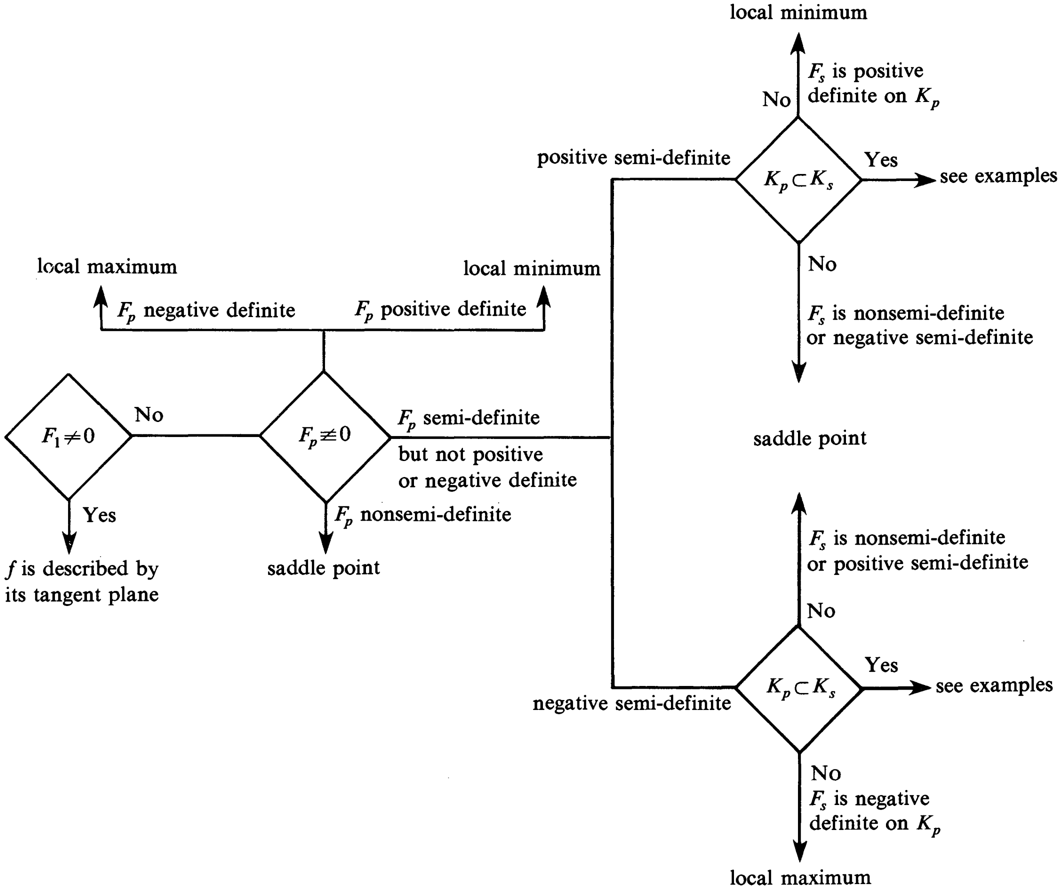


FIGURE 1.

We shall use the Taylor forms F_p, F_s and F_t to develop a procedure for classifying the critical points for f; this procedure is outlined schematically in FIGURE 1. As a preview of our approach, let us examine the function $F(x, y) = (x^2 - y^2)^2 + x^6 + y^6$. Here $F_4(x, y) = x^4 - 2x^2y^2 + y^4$ and $F_6(x, y) = x^6 + y^6$, while all other F_k are zero. So $p = 4$, $s = 6$ and t does not exist. The cone K_4 consists of the two lines $x = y$ and $x = -y$, and K_6 is just the origin. Thus, by FIGURE 1, since F_4 is semi-definite, $K_4 \not\subset K_6$, and F_6 is positive definite on K_4, we see that the origin is a local minimum for F. The propositions which follow provide details for this classification procedure.

PROPOSITION 1. *If F_p is nonsemi-definite, then a is a saddle point of f.*

Proof. Without loss of generality we may (and we shall throughout the sequel) assume that $a = 0$, $f(a) = 0$ and G is a ball centered at 0. From Taylor's Theorem we have

$$f(x) = F_p(x) + \|x\|^p R_p(x), \tag{1}$$

where

$$\lim_{x \to 0} R_p(x) = 0. \tag{2}$$

Let $F_p(h) \neq 0$. Because of (2), there exists a suitable $\delta > 0$ such that, if $0 < t < \delta$, then

$$\|h\|^p |R_p(th)| < \tfrac{1}{2} |F_p(h)|.$$

Since by (1) we have

$$f(th)=t^p\left[F_p(h)+\|h\|^p R_p(th)\right],$$

we see that $f(th)$ has the sign of $F_p(h)$ if $0<t<\delta$. Since F_p takes both positive and negative values, this finishes the proof.

Note that if p is odd, then $F_p(-h)=-F_p(h)$ since F_p is homogeneous of degree p and hence F_p is nonsemi-definite. Then by Proposition 1, a is a saddle point.

PROPOSITION 2. *If F_p is positive definite, then a is a local minimum of f. If F_p is negative definite, then a is a local maximum of f.*

Proof. Let $m=\inf\{F_p(x):\|x\|=1\}$. Since F_p is positive definite and continuous in R^n and since the unit sphere in R^n is compact, we have $m>0$. Because of homogeneity, we get $F_p(x)\geqslant m\|x\|^p$ for all $x\in R^n$. Because of (2), there exists a $\delta>0$ such that $|R_p(x)|<\frac{1}{2}m$ provided that $0<\|x\|<\delta$. By virtue of (1), we get that if $0<\|x\|<\delta$, then $f(x)>m\|x\|^p-\frac{1}{2}m\|x\|^p>0$, i.e., 0 is a local minimum for f. In order to prove the second part of the proposition, it suffices to replace f by $-f$ and apply the first part.

The classification given by Propositions 1 and 2 includes the classical (advanced calculus) case of a nondegenerate critical point. But it remains to treat the case when F_p is positive (respectively negative) semi-definite but not positive (respectively negative) definite.

PROPOSITION 3. *Suppose that F_p is positive semi-definite but not positive definite and suppose that K_p is not a subset of K_s. If F_s is either nonsemi-definite or negative semi-definite on K_p, then a is a saddle point of f. If F_s is positive definite on K_p, then a is a local minimum for f.*

Proof. If F_s is nonsemi-definite on K_p, then by Taylor's Theorem $f(x)=F_s(x)+\|x\|^sR_s(x)$, where $\lim_{x\to 0}R_s(x)=0$ if $x\in K_p\cap G$, and the argument of Proposition 1 applies. If F_s is negative semi-definite on K_p, we can use the argument of Proposition 1 to obtain $h_1\in K_p$ and a $\delta>0$ such that $f(th)<0$ if $0<t<\delta$. Since F_p is positive definite on $K_p^c\cup\{0\}$ (the superscript c denotes complement), by the same argument we can find $h_2\in K_p^c$ and $\delta_2>0$ such that $f(th_2)>0$ if $0<t<\delta_2$. Thus the point 0 is a saddle point of f. Now, suppose that F_s is positive definite on K_p and let $m=\inf\{F_s(x):x\in K_p\cap S^{n-1}\}$. Obviously $m>0$. Let $\epsilon>0$ be such that $F_s(x)>(3/4)m$, provided that x belongs to the ϵ-neighborhood of $K_p\cap S^{n-1}$. By the homogeneity of F_s, we have $F_s(x)\geqslant\frac{3}{4}m\|x\|^s$ for all $x\in K_p(\epsilon)$. Since F_p is positive definite on the cone $C=[K_p(\epsilon)]^c\cup\{0\}$, the argument of Proposition 2 shows that there exists a $\delta>0$ such that $f(x)>0$ if $0<\|x\|<\delta$ and $x\in G\cap C$. Let $\delta_1>0$ be a positive number such that $|R_s(x)|<(1/2)m$, if $0<\|x\|<\delta_1$. Then, for $x\in K_p(\epsilon)$ and $0<\|x\|<\delta_1$, we get

$$f(x)\geqslant F_s(x)+\|x\|^sR_s(x)>(3/4)m\|x\|^s-(1/2)m\|x\|^s=(1/4)m\|x\|^s>0.$$

Thus $f(x)>0$ for all x satisfying $0<\|x\|<\min\{\delta,\delta_1\}$, i.e., 0 is a local minimum for f.

We note that the case when F_p is negative semi-definite but not negative definite is entirely analogous. Also if F_p and F_s restricted to K_p are positive semi-definite but not positive definite, we have to look into the behavior of F_t on K_s and make a similar statement.

It remains to investigate the case $K_p\subset K_s$ under the hypothesis that F_p is positive semi-definite but not positive definite. Here no simple conclusion applies as we will show with a number of examples. Obviously, if the first form nonvanishing identically on K_p takes a negative value on K_p, the argument of Proposition 3 shows that a must be a saddle point of f. However, if the first nonzero form on K_p is positive definite on K_p, we may not assert that a is a local minimum. The following example demonstrates this. [In this and the following examples we denote a point in the plane by the more traditional (x,y) instead of (x_1,x_2)].

EXAMPLE 1. Let $f(x,y)=x^2-3xy^2+2y^4=(x-y^2)(x-2y^2)$ (see FIGURE 2). Obviously the critical point $(0,0)$ is a saddle point of f, since $f(x,y)>0$ for $x>2y^2$ or $x<y^2$, and $f(x,y)<0$ for $y^2<x<2y^2$. Here $p=2$, $s=3$, $t=4$, $K_p=$ "the y-axis" and $K_s=$ "x-axis" $\cup$ "y-axis." Thus $K_p\subset K_s$

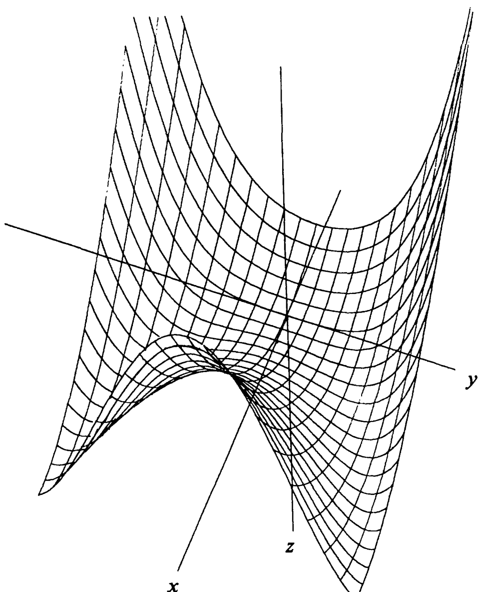


FIGURE 2.

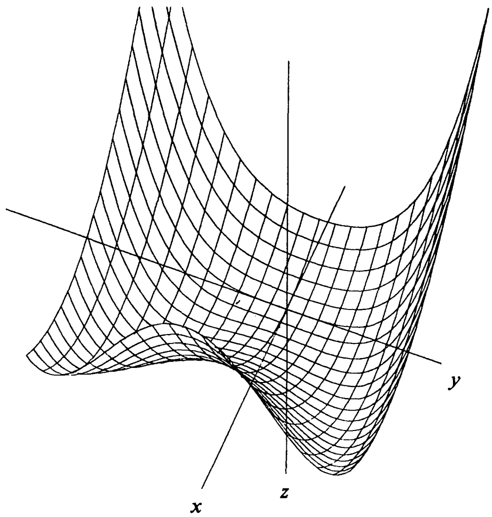


FIGURE 3.

and the form $F_4(x,y)=2y^4$ is positive definite on K_p. On the other hand, the function $g(x,y)=x^2-2xy^2+2y^4=(x-y^2)^2+y^4$ has the same situation of cones and sign definiteness, but $(0,0)$ is a minimum (see FIGURE 3).

Indeed, in the case that all forms of order higher than p vanish on K_p, even if s is odd, the following examples show that there can be either a saddle point or a local minimum.

EXAMPLE 2. Let $f(x,y)=x^2-xy^2=x(x-y^2)$ (see FIGURE 4). Obviously the critical point $(0,0)$ is a saddle point of f, since $f(x,y)>0$ for $x>y^2$ or $x<0$, but $f(x,y)<0$ for $0<x<y^2$. Here $p=2$, $s=3$ and the cones are the same as in Example 1. Moreover, F_m vanishes on K_p for all m.

EXAMPLE 3. Let $f(x,y)=x^2-x^2y+e^{-1/y^2}$ for $(x,y)\neq(0,0)$, and $f(0,0)=0$. Then $f(x,y)>0$ for $y<1$ and hence the critical point $(0,0)$ is a minimum for f (see FIGURE 5). Here again, $p=2$, $s=3$ and the cones K_m are the same as in Examples 1 and 2. Moreover, F_m vanishes on K_p for all m.

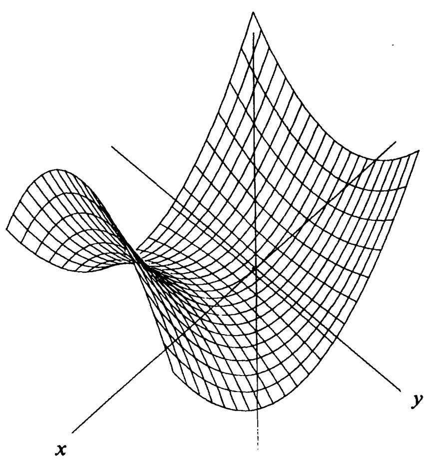


FIGURE 4.

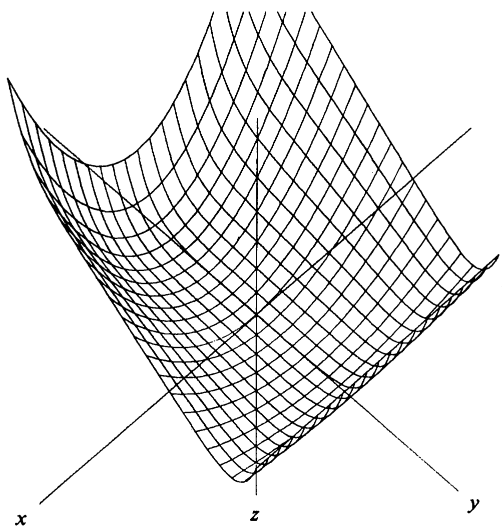


FIGURE 5.

We conclude with some further informal remarks about critical points which show how our ideas extend to and pervade more general geometry. It is well known that degenerate critical points of C^2 functions are atypical in the sense that there are small changes of the function which render the critical point nondegenerate. However, nondegenerate critical points a of a C^2 function f are isolated (only one in a sufficiently small open set) and for every small enough C^2 function ϵ which has a as a critical point, the point a will be a nondegenerate critical point of the function $f+\epsilon$. Here "smallness" is measured by the least upper bound of the function and its partial derivatives of order $\leqslant 2$ in a neighborhood of a. The proofs of these properties of critical points is based on the simple fact that in the space of symmetric $n \times n$ matrices which can be thought of as R^m with $m=\frac{1}{2}n(n+1)$, the singular ones form a closed cone without any interior points. This cone is obviously $\{A:\det A=0\}$. Once in this cone a small step in most directions will take you out of the cone, whereas if you are outside of this cone, you have a whole neighborhood to walk on and still stay away from singular matrices.

Functions with no degenerate critical points are called **Morse functions**. They are very important tools in the study of the geometry of smooth manifolds at large. A C^k-smooth manifold of dimension $n(k=0,1,\ldots,\infty)$ is a separable Hausdorff space on which we can introduce local coordinates so that at points in overlapping domains of such coordinates we can go from one system to the other by means of smooth coordinate transformations; see **[4]** for a technical definition.

A common type of C^k manifold $(k \geqslant 1)$ is the subset M of R^n defined by $M=\{x \in R^n: g(x)=0\}$, where $g: R^n \to R^m$ is a C^k mapping such that at each point of M the Jacobian matrix has rank $m(m<n)$. The dimension of M is $n-m$ and the local coordinates are provided by the implicit function theorem. Nevertheless, there are more complicated manifolds (like projective spaces) and it takes the famous theorem of Whitney to show that every n-dimensional manifold can be actually considered as a (smooth if necessary) submanifold of R^{2n+1}. For details we refer to **[7]**.

By means of the local coordinates we can introduce smooth mappings between smooth manifolds. For example, a function f defined on some open set G of a manifold M is smooth, if it is smooth in each local coordinate system. A point a is critical for f, if it is critical for some (and, by the chain rule, for all) local coordinate system. Thus, the classification of critical points can be immediately extended to C^k-functions defined on C^k-manifolds (k sufficiently high) because of the local character of the classification. One might as well work in R^n. This is usually done in the case of optimizing a smooth function under smooth constraints (Lagrange multipliers). The constraints usually define a manifold and the objective function is restricted to that manifold.

It was mentioned earlier that the role of critical point theory is not restricted to obtaining local results only. The theory of Morse functions has been very successful in obtaining beautiful and far-reaching global properties of manifolds. On this, the reader is referred to **[3]**, **[4]** and **[5]**. A particularly beautiful result of this kind is the following characterization of the n-dimensional sphere S^n due to Reeb: If M is a compact C^2 n-dimensional manifold and if f is a C^2-smooth function on M with exactly two critical points, then M is homeomorphic to S^n (cf., p. 25 of **[3]**).

References

[1] C. H. Edwards, Advanced Calculus of Several Variables, Academic Press, New York, 1973.
[2] S. Mac Lane and G. Birkhoff, Algebra, Macmillan, New York, 1967.
[3] J. Milnor, Morse Theory, Annals of Mathematics Studies #51, Princeton Univ. Press, Princeton, 1963.
[4] J. Milnor, Topology from the differentiable viewpoint, Univ. Press of Virginia, Charlottesville, 1965.
[5] M. Morse and S. S. Cairns, Critical Point Theory in Global Analysis and Differential Topology, Academic Press, New York, 1969.
[6] T. Poston and I. N. Stewart, Taylor expansions and catastrophes, Res. Notes Math. #7, Pitman, New York, 1976.
[7] S. Sternberg, Lectures on Differential Geometry, Prentice-Hall, Englewood Cliffs, 1964.

PROBLEMS

DAN EUSTICE, Editor
LEROY F. MEYERS, Associate Editor
The Ohio State University

Proposals

To be considered for publication, solutions should be mailed before May 1, 1981.

1107. Determine the maximum value of $\sin A_1 \sin A_2 \cdots \sin A_n$ if $\tan A_1 \tan A_2 \cdots \tan A_n = 1$. [*M. S. Klamkin, University of Alberta.*]

1108.* For n odd, let $C[n]$ be the number of cycles in the permutation of $\{0,1,\ldots,n-1\}$ sending $i \to 2i \pmod{n}$. Prove that $C[3(2^j-1)] = C[5(2^j-1)]$ for all odd positive integers j. [*James Propp, Harvard University.*]

1109. March 21 is commonly considered as the first day of spring (the date of the vernal equinox)—a tradition dating from the Council of Nicaea in 325 A.D. The most recent year in which this was in fact true was 1979, when the vernal equinox occurred at 12:22 a.m. EST on March 21.

When will be the next year in which spring begins as late as March 21 in the United States? (The average interval between vernal equinoxes—the tropical year—is to be taken as 365 days, 5 hours, 48 minutes, and 46 seconds.) [*Thomas R. Nicely, Lynchburg College.*]

1110. A certain mathematician, in order to make ends meet, moonlights as an apprentice plumber. One night, as the mathematician contemplated a pile of straight pipes of equal lengths and right-angled elbows, the following question occurred to this mathematician: "For which positive integers n could I form a closed polygonal curve using n such straight pipes and n elbows?" [*Gerald Wildenberg, University of Hartford.*]

1111. Let R be a finite ring. Evaluate $\Sigma_{r \in R^*} r$ and $\Sigma_{r \in R} r$, where R^* is the group of units of R. [*Douglas Lewan, Brown University.*]

ASSISTANT EDITORS: DON BONAR, *Denison University;* WILLIAM A. MCWORTER, JR., *The Ohio State University. We invite readers to submit problems believed to be new. Proposals should be accompanied by solutions, when available, and by any information that will assist the editors. Solutions to published problems should be submitted on separate, signed sheets. An asterisk (*) will be placed by a problem to indicate that the proposer did not supply a solution. A problem submitted as a Quickie should be one that has an unexpected succinct solution. Readers desiring acknowledgment of their communications should include a self-addressed stamped card. Send all communications to this department to Dan Eustice, The Ohio State University, 231 W. 18th Ave., Columbus, Ohio 43210.*

Quickies

Solutions to Quickies appear at the conclusion of the Problems section.

Q666. If w and z are complex numbers, prove that

$$2|w|\cdot|z|\cdot|w-z| \geqslant (|w|+|z|)\,|w|z|-z|w||.$$

[*M. S. Klamkin, University of Alberta.*]

Solutions

Closure is R **September 1979**

1079. Define $a_0=1$ and $a_{n+1}=(a_n-2)/a_n$ for $n\geqslant 0$.
(a) Show that the set $\{a_n: n=0,1,2,\dots\}$ is unbounded.
(b) There exists a real number α such that $\{n: a_n \geqslant 1\}=\{[k\alpha]: k=0,1,2,\dots\}$. Find α.
(c)* Find the closure of the set defined in part (a).
[*James Propp, student, Harvard University.*]

Solution: Setting $a_n=u_{n+1}/u_n$ leads to the equation $u_{n+2}-u_{n+1}+2u_n=0$, whose general solution can be obtained by the method of generating functions as found in books on difference equations, probability theory, and combinatorics. This solution together with the condition $a_0=1$ yields

$$a_n=\frac{1+\sqrt{7}\,\cot(n+1)\theta}{2},\ \text{where } \theta=\tan^{-1}\sqrt{7}\,.$$

The solution of (a) will follow from solving (c). To solve (b) we see that $a_n\geqslant 1$ is equivalent to $\cot(n+1)\theta\geqslant 1/\sqrt{7}$, and this will hold if and only if for some integer k,

$$k\pi<(n+1)\theta\leqslant k\pi+\cot^{-1}\frac{1}{\sqrt{7}}=k\pi+\theta.$$

Thus $n\leqslant k\pi/\theta<n+1$ and $[k\pi/\theta]=n$, and so $\alpha=\pi/\theta$.

Solution of (c). As is well known, the numbers $(n+1)\theta$, modulo π, will be dense in the interval $(0,\pi)$ if θ/π is irrational, which we will show to be true. It then follows that the values of $\cot(n+1)\theta$, and hence of a_n, are dense in the reals.

Now θ/π is irrational if and only if $\sin n\theta\neq 0$ for all positive integers n. By De Moivre's formula, this is equivalent to the assertion that $(\cos\theta+i\sin\theta)^n$ is not a real number. Since $2^{3n/2}(\cos\theta+i\sin\theta)^n=(1+i\sqrt{7}\,)^n$, it suffices to show that the latter is not real. But this follows from the fact, easily proved by induction, that $(1+i\sqrt{7}\,)^n$ can be put in the form $2^{n-1}(a+4M+ai\sqrt{7}\,)$, where a and M are integers and a is odd.

HUGH NOLAND
California State University at Chico

Also solved by Nicolas K. Artémiadis (Greece), Bern Problem Solving Group (Switzerland), Robert Clark, Gordon Fisher, Hans Kappus (Switzerland), L. Kuipers (Switzerland), Mark F. Kruelle, P. J. Pedler, Ken Yocum, and the proposer. Not all of the solvers found the closure of the set.

Int and sgn

November 1979

1080. Some calculators have an "int" key. The "integral part of x" is given by $\operatorname{int}x=[|x|]\operatorname{sgn}x$, where $[x]$ denotes the greatest integer not greater than x and where $\operatorname{sgn}x$ is -1 when $x<0$, 0 when $x=0$, and $+1$ when $x>0$.

We have $|x|=x\operatorname{sgn}x$ and $\max(x,y)=(x+y+|x-y|)/2$ as examples of familiar functions which can be expressed in terms of "sgn" together with the operations $\{+,-,\times,\div\}$. Show that these functions can be similarly expressed in terms of "int." [*Marlow Sholander, Case Western Reserve University.*]

Solution: Since $x/(x^2+1)$ is always smaller than 1 in absolute value and has the same sign as x,

$$\operatorname{sgn}x \quad = \quad \operatorname{int}\left(\frac{x}{x^2+1}+1\right) \quad + \quad \operatorname{int}\left(\frac{x}{x^2+1}-1\right).$$

KARL HEUER & G. A. HEUER
Concordia College

Also solved by Richard Beigel, Clayton W. Dodge, Gordon Fisher, Mark F. Kruelle, Peter Schumer, Harry Sedinger, Lawrence Somer, Wen-King Su, Douglas H. Underwood, and the proposer.

Two Solutions

November 1979

1081. Find all real t such that for all $x>y>0$,

$$(x-y)^t(x+y)^t=(x^t-y^t)^t(x^t+y^t)^{2-t}.$$

[*Edwin P. McCravy, Midlands Technical College.*]

Solution: Assume $t\neq 0$. Write the equation as

$$(x^2-y^2)^t(x^t+y^t)^t=(x^t-y^t)^t(x^t+y^t)^2$$

and divide both sides by x^{t^2+2t} to get

$$\left(1-(y/x)^2\right)^t\left(1+(y/x)^t\right)^t=\left(1-(y/x)^t\right)^t\left(1+(y/x)^t\right)^2.$$

Putting $u=y/x$, this gives the equivalent equation

$$(1+u^t)^{t-2}=((1-u^t)/(1-u^2))^t$$

for $0<u<1$. Taking the limit of both sides as u goes to 1 from below, using L'Hôpital's rule, we get $2^{t-2}=(t/2)^t$ or $4^{1/t}t=4$. Clearly $t=1$ and $t=2$ are solutions of this equation. They are the only solutions. In fact, set $f(t)=4^{1/t}t-4$. Then $f'(t)=4^{1/t}(1-((\ln 4)/t))=0$ if and only if $t=\ln 4$. Since $f''(\ln 4)>0$, the only extremum is a minimum for $t=\ln 4$, where f is negative and $1<\ln 4<2$. It follows that f has exactly the two zeros $t=1$ and $t=2$, and that these are the t required.

GORDON FISHER
James Madison University

Also solved by Anders Bager (Denmark), Duane M. Broline, Karl Heuer & G. A. Heuer, L. Kuipers (Switzerland), P. Ramankutty, J. M. Stark, Edward T. H. Wang, and the proposer.

A Trigonometric Inequality

November 1979

1082. Prove that $\tan\sin x>\sin\tan x$ for $0<x<\pi/2$. [*C. S. Gardner, University of Texas.*]

Solution: Case I. $0<x<\operatorname{Arctan}\pi/2$. The proposed inequality is implied by $(d/dx)\tan\sin x>$

$(d/dx)\sin\tan x$, which is equivalent to

$$\frac{2}{3}\log\sec\sin x+\frac{1}{3}\log\sec\tan x>\log\sec x.$$

The function logsec is convex and increasing since its first and second derivatives are positive. Hence it suffices to prove

$$\frac{2}{3}\sin x+\frac{1}{3}\tan x>x.$$

But this is true since

$$\frac{d}{dx}\left(\frac{2}{3}\sin x+\frac{1}{3}\tan x-x\right)=\frac{1}{3}\sec^2 x(1+2\cos x)(1-\cos x)^2>0.$$

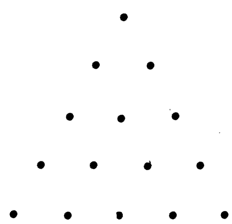
Case II. $\operatorname{Arctan}\pi/2\leqslant x<\pi/2$. Then $\sin x>\pi/4$ and $\tan\sin x>1\geqslant\sin\tan x$.

C. S. GARDNER
University of Texas

Also solved by Chico Problem Group of California State University, Gordon Fisher, and Freddy J. Maynard & Gerald Thompson.

Lacing a Lattice **November 1979**

1083. Given an equilateral point lattice with n points on a side, it is easy to draw a polygonal path of n segments passing through all the $n(n+1)/2$ lattice points. Show that it cannot be done with less than n segments. [M. S. Klamkin and A. Liu, University of Alberta.]



$n=5$

Solution: Proof by induction. This is obvious for $n=2$. Assume the proposition for $n-1$. If any segment is contained completely in an edge, then at least $n-1$ segments are required in order to cover the remaining points by the inductive hypothesis. Otherwise, each segment intersects an edge in at most one point. Since each edge has n points, the proposition holds. Clearly, the proof shows that the segments need not be connected.

RICHARD BEIGEL
Stanford University

Also solved by Walter Bluger (Canada), Duane M. Broline, Chico Problem Group of California State University, Cliff Corzatt, Gordon Fisher, P. K. Garlick, Joseph Heisler, R. L. McKinney (Canada), and the proposers.

Four Different Zeros **November 1979**

1085. Consider the polynomial $P(x)=x^4-14x^2+x+38$. Find a function $g=g(x;\epsilon_1,\epsilon_2)$, where ϵ_1 and ϵ_2 are ±1, such that the recursive sequence $x_{n+1}=g(x_n)$ converges to a different zero of $P(x)$ for each of the four distinct values of (ϵ_1,ϵ_2). [*Bert Waits, The Ohio State University.*]

Solution: Four successful iterating functions are $g(x;\ \epsilon_1,\epsilon_2)=\epsilon_1\sqrt{7+\epsilon_2\sqrt{11-x}}$, where $\epsilon_1,\epsilon_2=\pm1$. Since the limit of any sequence of iterates, $x_{n+1}=g(x_n)$, satisfies $x=g(x)$, limiting values necessarily satisfy $x=\epsilon_1\sqrt{7+\epsilon_2\sqrt{11-x}}$ and hence $P(x)=x^4-14x^2+x+38=0$. Furthermore, convergence of each $g(x;\ \epsilon_1,\epsilon_2)$ to one of the four roots of $P(x)=0$ is guaranteed by the following theorem from numerical methods: If $g(x)$ is continuously differentiable over $[a,b]$, the range $g([a,b])$ is a subset of the domain $[a,b]$, and the maximum absolute value of $g'(x)$ over $[a,b]$ is less than or equal to some constant which is strictly less than 1, then the sequence $\{x_n | x_{n+1} = g(x_n)\}$ converges for all initial values $x_0 \in [a, b]$ to a unique limit in $[a, b]$. All four functions $g(x_n; \epsilon_1, \epsilon_2)$ are continuously differentiable on $(-38, 11)$, and when restricting this interval to $[-25, 10]$ the maximum absolute value of $g'(x)$ is bounded above by $1/4$. Since also the range $g([-25, 10])$ is contained in the domain $[-25, 10]$, the theorem implies that the limit as $n \to \infty$ of $g(x; \epsilon_1, \epsilon_2)$ exists and is one of the roots of $P(x) = 0$ regardless of initial choice of $x_0 \in [-25, 10]$. Specifically the roots obtained are 3.1313125^+ for $\epsilon_1 = \epsilon_2 = 1$; 2.0000000 for $\epsilon_1 = 1$, $\epsilon_2 = -1$; -1.8481265^+ for $\epsilon_1 = \epsilon_2 = -1$; and -3.2831860^- for $\epsilon_1 = -1$, $\epsilon_2 = +1$.

PROBLEM SOLVING GROUP
University of Hartford

Also solved by G. A. Heuer, J. M. Stark, Michael Vowe (Switzerland), and the proposer.

Group of Transformations **November 1979**

1086. Consider the following transformations on 4×4 matrices. Let R move the top row to the bottom and the other rows cyclically up; let D be the reflection across the main diagonal; let S be the interchange of the 1st and 2nd rows followed by the interchange of the 1st and 2nd columns. What is the order of the group generated by R, D, and S? [*Barbara Turner, California State University, Long Beach.*]

Solution: The order of the group generated by R, D, and S is 1152.

Proof. Let the group be denoted by G. The following property of each of R, D, and S holds for every element T of G: If T takes any row of the matrix A into a row (column) of the image, then T takes every row of A into a row (column) or the image, and T takes every column of A into a column (row) of the image.

Direct calculations show that R^2SRS interchanges rows 2 & 3, where the transformations are applied from left to right, R^3SRSR^3 interchanges rows 3 & 4, and $RSRSR$ interchanges rows 1 & 2. A standard theorem in permutation theory then shows that all permutations of rows can be achieved by using R & S.

Similarly, if T performs a permutation of rows, DTD performs the same permutation of columns.

Given any matrix A, we can take row 1 into any row or, by using D, into any column. Hence there are 8 choices for the image of row 1. By the property in the first paragraph, there are then 3 choices for the image of row 2, 2 choices for the image of row 3, and 1 choice for the image of row 4. Having mapped the rows of A into (say) rows, we can permute the columns of the resulting matrix in $4! = 24$ ways. Since there were $8 \cdot 3 \cdot 2 \cdot 1$ ways of mapping the rows, there are $24 \cdot 48 = 1152$ ways of mapping A in all, and all may be accomplished by elements of G. From paragraph 1, these are the only ways in which A may be mapped by elements of G. Hence G has order 1152.

DOUGLAS UNDERWOOD
Whitman College

Also solved by Peter R. Atwood, Duane Broline, Chico Problem Group of California State University, and the proposer.

Irrational Coefficients **November 1979**

1087. Let $\sum_{k=-\infty}^{k=+\infty} a_k z^k$ be the Laurent series of $e^{z+\frac{1}{z}}$ for $0 < |z| < \infty$.

(a) Show that each a_k is an irrational number.

(b) Show that the set $\{a_k : k \geqslant 0\}$ is linearly dependent over the rationals. [*Barbara Turner, California State University, Long Beach.*]

Solution: To obtain the Laurent series of $f(z)=\exp(z+1/z)$, $0<|z|<\infty$, we may consider the product of the series expansions of $\exp z$ and $\exp(1/z)$. The coefficient a_k of z^k is then seen to be

$$a_k=a_{-k}=\sum_{n=0}^{\infty}\frac{1}{n!(k+n)!},\ k\geqslant 0. \tag{1}$$

Since $m!>2^{m-1}$ for $m>2$, we have

$$\sum_{n=2}^{\infty}\frac{1}{n!(k+n)!}<\sum_{n=2}^{\infty}(1/2)^{2n+k-2}=\frac{1}{3}\cdot 2^{-k}. \tag{2}$$

Hence

$$\frac{1}{k!}+\frac{1}{(k+1)!}<a_k<\frac{1}{k!}+\frac{1}{(k+1)!}+\frac{1}{3}\cdot 2^{-k}. \tag{3}$$

From 3 we infer that $2<a_0<3$, $1<a_1<2$, and $0<a_k<1$ for $k\geqslant 2$. Therefore a_k is not an integer for any k. Assume now that $a_k=p/q$, where p and q are integers and $q\geqslant 2$. Then $q!(k+q)!a_k$ is an integer. On the other hand it follows from (1) that $q!(k+q)!a_k=g+r$, where g is an integer and

$$0<r<\sum_{m=1}^{\infty}\prod_{i=1}^{m}\frac{1}{(q+i)(q+k+i)}<\sum_{m=1}^{\infty}(q+1)^{-2m}\leqslant\frac{1}{8}.$$

This contradiction proves (a).
Furthermore we have

$$f'(z)=(1-z^{-2})f(z).$$

Since termwise differentiation is permitted, we obtain by comparing coefficients that

$$ka_k-a_{k-1}+a_{k+1}=0,\ k\geqslant 1.$$

Thus we have infinitely many linear relations in the set $\{a_k|k\in\mathbb{Z}\}$ with integer coefficients, which proves (b).

REMARK. $a_k=I_k(2)$, where I_k denotes the kth Bessel function with purely imaginary argument.

HANS KAPPUS
Switzerland

Also solved by R. P. Boas, Chico Problem Group of California State University, Mark F. Kruelle, J. M. Stark, and the proposer.

Answers

Solutions to the Quickies which appear near the beginning of the Problems section.

Q666. Let $w=re^{i\alpha}$ and $z=se^{i\beta}$. Then we have to show that $2|re^{i\alpha}-se^{i\beta}|\geqslant(r+s)|e^{i\alpha}-e^{i\beta}|$ or

$$4\{(r\cos\alpha-s\cos\beta)^2+(r\sin\alpha-s\sin\beta)^2\}\geqslant(r+s)^2\{(\cos\alpha-\cos\beta)^2+(\sin\alpha-\sin\beta)^2\}.$$

This is equivalent to

$$2(r^2+s^2-2rs\cos(\alpha-\beta))\geqslant(r^2+s^2+2rs)(1-\cos(\alpha-\beta))$$

or finally $(r-s)^2(1+\cos(\alpha-\beta))\geqslant 0$. There is equality if $r=0$ or $s=0$ or $r=s$ or $\alpha-\beta=\pm\pi$.

REVIEWS

Assistant Editor: Eric S. Rosenthal, Princeton University. Articles and books are selected for this section to call attention to interesting mathematical exposition that occurs outside the mainstream of the mathematics literature. Some reviews of books are adapted from the Telegraphic Reviews in the American Mathematical Monthly.

Kolata, Gina Bari, *Hua Lo-Keng shapes Chinese Math*, Science 210 (24 Oct. 1980) 413-414.

A profile of China's leading mathematician, now visiting in the United States, relating his research, his life, and his recent effort to popularize mathematics.

Singmaster, David, Notes on Rubik's Magic Cube, Fifth Edition, Logical Games (4509 Martinwood Dr., Haymarket, VA 22069), 1980; 75 pp, $5 (P).

Invented by the Hungarian artist E. Rubik, the Magic Cube has sold more than a million copies under various tradenames. Included is an easy step-by-step solution to restore the cube to original position in at most 200 moves, as well as the more complicated algorithm of Thistlethwaite to do it in 52 or fewer. The booklet also explains, in language students in an abstract algebra course will understand, the group theoretical underpinnings of the puzzle.

Morris, Scott, *Games: Ernö Rubik's magic cube*, Omni (September 1980) 128-129; *Inside Rubik's cube* (October 1980) 193.

Popular description of the Magic Cube, comparing its impact to that of Sam Loyd's 15 Puzzle (1873)--as described by Loyd himself: "People became infatuated with the puzzle and ludicrous tales are told of shopkeepers who neglected to open their stores...Pilots are said to have wrecked their ships, engineers rush their trains past stations, and businesses became demoralized."

Kahan, William M., *Handheld calculator evaluates integrals*, Hewlett-Packard J. 31:8 (August 1980) 23-32.

Description of the ingeniously robust numerical integration algorithm incorporated into the HP-34C calculator. The article is excellent reading for a course in numerical analysis.

Lawler, E.L., *The great mathematical Sputnik of 1979*, The Sciences 20 (Sept. 1980) 12-15, 34.

A review of media reaction to Khachiyan's work on elliptical algorithms for linear programming problems--an interesting case study of how mathematical misinformation is propagated by the popular press.

Berresford, G.C., et al., *Khachiyan's algorithm. Part 1: A new solution to linear programming problems; Part 2: Problems with the algorithm*, Byte 5 (August 1980) 198-208; (September 1980) 242-255.

Explanation of Khachiyan's paper and analysis of the geometry of the algorithm, accompanied by a Basic program that simulates the algorithm.

MacLane, Saunders, *Mathematics*, Science 209 (4 July 1980) 104-110.

Selective summary of results in mathematics achieved in the last 10 to 15 years. A must for researchers and graduate students, but difficult reading all the same.

Miser, Hugh J., *Operations research and systems analysis*, Science 209 (4 July 1980) 139-147.

History and prospects, with detailed case studies of blood-banking and school busing.

Simon, Herbert A., *The behavioral and social sciences*, Science 209 (4 July 1980) 72-78.

Nobel Prize winner reviews progress in technique, then discusses frontiers: 1) evolutionary theory, including economic growth, survival of profit maximizing firms, and modeling of altruism and egoism; 2) human rational choice, including the theory of subjective expected utility, and Prisoners' Dilemma; and 3) "cognitive science," the revolution in psychology brought on by the information processing paradigm and the digital computer.

Kadane, Joseph B. and Kairys, David, *Fair numbers of per-emptory challenges in jury trials*, J. American Statistical Association 74 (December 1979) 747-753.

Investigates application of mathematics to the problem of eliminating biased jurors.

Grattan-Guinness, I. (Ed.), From the Calculus to Set Theory, 1630-1910: An Introductory History, Duckworth, 1980; 306 pp, $49.50.

Six chapters trace the emergence of mathematical analysis and set theory, with the object of introducing the reader to the historical development of those subjects. The author's goal is to enhance the reader's understanding of why the mathematical theory developed and took the *form* it did. The book is a superb means to that end. Libraries, teachers, and students alike, however, will regret that the end does not justify setting such a high price on the means.

Cook, Theodore Andrea, The Curves of Life, Dover, 1979; xxx + 479 pp, $5.95.

"Being an account of spiral formations and their application to growth in nature, to science and to art; with special reference to the manuscripts of Leonardo da Vinci. With 11 plates and 415 illustrations." Unabridged reprint of 1914 edition--this is *the* book on spirals in nature, architecture and art.

Steen, Lynn Arthur, *A monstrous piece of research*, Science News 118 (27 Sept. 1980) 204-206.

A popular summary of Robert Griess' confirmation of the existence of the "monster" sporadic group, which led to the recent completion of the classification of the finite simple groups.

Bacon, P.J. and MacDonald, D.W., *To control rabies--Vaccinate foxes,* New Scientist 87 (29 Aug. 1980) 640-645.

Uses mathematical models to show why vaccination of foxes may be the most efficient method for controlling the spread of rabies.

Brush, Lorelei R., Encouraging Girls in Mathematics, Abt Books, 1980; xv + 163 pp, $16.

Results of a 3-year longitudinal study of 1500 students indicate increasing negative attitudes toward mathematics from 6th to 12th grade among both boys and girls, but girls perceive mathematics as more difficult and anxiety-producing. Brush's "remedial strategies"--relaxed classrooms, 4 years of mathematics for all students, new courses, career information, and promoting the usefulness of mathematics--are sound but likely to be ineffective, particularly in addressing attitudes and sex role definitions deeply embedded in our culture.

Armstrong, Jane M., *Achievement and participation of women in mathematics: An overview,* Education Committee of the States (now available from ERIC Document Reproduction Service as ED #184 878), 1980, xi + 35 pp.

Report of a two-year study to "identify the most important factors related to the problem of women's (low level of) participation in mathematics." Results indicate that females start high school on a par with males--better, in fact, at computation and spatial visualization; but by 12th grade, males have overtaken females.

Brams, Steven J., Biblical Games: A Strategic Analysis of Stories in the Old Testament, MIT Pr, 1980; xi + 196 pp, $15.

"The thesis...is that both God and human biblical characters acted rationally in a series of games played in the Old Testament. ...Elementary tools from the theory of games are used. ...[A] detailed assessment of God's character and motivations is offered at the end, with reasons given for His frequently wrathful behavior."

Dershem, Herbert L. and Smith, David A., Computers in Teaching: 1979 State of the Art Report of Instructional Computing: Mathematics and Statistics, Conduit, 1979; 40 pp, $4 (P).

Notes that instructors need not each "reinvent the wheel:" as opposed to a few years ago, today assorted texts and program packages are available for use in mathematics courses. The main force holding back more common use of the computer, the authors assert, is lack of funds for equipment and software.

Schoijet, Mauricio, *Who's afraid of a vector?* Bull. Atomic Scientists 36:6 (June 1980) 60-62.

Three provinces of Argentina have banned the teaching of modern mathematics because it is considered subversive, claiming: "Some themes of mathematics use words such as vector and matrix, which are typical of a Marxist or typically subversive vocabulary. The same happens with set theory [*teoría de conjuntos; conjunto* also means ensemble] which evidently tends to 'massify' [*masificar*] and to evoke multitudes." Moreover, "through the teaching of modern mathematics the postulates of formal logic are denied. From there on a dangerous breach for subversive action would be opened."

Freiman, Grigori, It Seems I Am A Jew: A Samizdat Essay, Southern Illinois U. Press, 1980; xvi + 97 pp, $9.95.

Damning indictment of anti-Semitism among Soviet mathematicians that names

names: I.M. Vinogradov, Yu. L. Ershov, and others. Appendices include an assessment of the situation by emigré Russian mathematicians, plus examples of "Jewish problems:" extraordinarily difficult problems given to Jews on examinations in order to fail them.

Miller, Rupert G., et al., Biostatistics Casebook, Wiley, 1980; xi + 238 pp, $14.95 (P).

Collection of intriguing and enlightening case studies, well worth the attention of any college teacher of statistics. The original data are included, so that readers may try their hands at reanalysis.

Paulos, John Allen, Mathematics and Humor, University of Chicago Press, 1980; 116 pp, $12.95.

"[S]ome qualities inherent in a good mathematical proof are similar to qualities inherent in good humor: cleverness and economy, playfulness, combinatorial ingenuity, and logic (particularly reductio ad absurdum)." This book is not a collection of jokes with a mathematical flavor, but a serious study of the nature of both humor and mathematics. The author focuses on "perceived incongruity with a point" as a key element of humor, and illustrates his comparison with explanations of both catastrophe theory and self-reference paradoxes as models for humor.

Mathematical sciences, and computer science, Annual Report, National Science Foundation, 1979, pp. 8-13.

Broad summaries in lay language of a few areas of recent research success: Connelly's result that a convex triangulated polyhedron is rigid, progress in robust statistical estimation, attack on the three dimensional problem of tomography image reconstruction, methodology for testing of computer programs, and understanding human vision through computer-based electro-optical imaging systems.

Papert, Seymour, *New cultures from new technologies*, Byte 5 (September 1980) 230-240.

Hints at a vision of a profound change in how children can learn, in a future where computers could tend to make schools as we know them obsolete. "It is time we learned to think in terms of a computer for every child."

Judson, Horace Freeland, *The rage to know*, Atlantic Monthly (April 1980) 112-117.

A paean to the curiosity, inspiration, and aesthetic satisfaction that motivate scientists: "Science has several rewards, but the greatest is that it is the most interesting, difficult, pitiless, exciting, and beautiful pursuit that we have yet found. Science is our century's art."

Rush, Jean C., *On the appeal of M.C. Escher's pictures*, Leonardo 12 (1979) 48-50.

Recounts some of the influences on Escher and the belated popularity of his work. Ironically, Escher's work was first ignored by artists and art critics because of its structured intellectual content, and because it was devoid of emotional expression; then later he was popularized by the psychedelic subculture.

McKeown, K.R. and Badler, N.I., *Creating polyhedral stellations*, Computer Graphics 14 (July 1980) 19-24.

A description of a computer program for generating stellations of solids. Some very nice graphics; the program has produced pictures of stellations "not previously drawn or enumerated."

NEWS & LETTERS

CONFERENCE ON UNDERGRADUATE MATHEMATICS

The sixth annual Conference on Undergraduate Mathematics, sponsored by the *Journal of Undergraduate Mathematics*, will be held at Hendrix College, Conway, Arkansas, on April 10-11, 1981. The program of the meeting will include presentation of papers written by students during their undergraduate careers, and talks by R.H. Bing, Paul R. Halmos, Burton W. Jones, M.Z. Nashed, and John W. Neuberger.

All student papers submitted before February 15, 1981 for publication in *JUM* will be considered for presentation at the Conference. Notification of acceptance will be made before March 15, 1981. For those participants presenting papers, board and lodging will be paid and limited travel support will be available.

For information regarding the Conference, contact J.R. Boyd, Department of Mathematics, Guilford College, Greensboro, NC 27410 (919-292-5511, Ext. 276) or Robert C. Eslinger, Department of Mathematics, Hendrix College, Conway, AR 72032 (501-450-1254).

1981 SHORT COURSE ON NUMERICAL LINEAR ALGEBRA

A Short Course in Numerical Linear Algebra will be held June 16-19, 1981 at the Ohio State University. The course, sponsored by the Ohio section of the M.A.A., will examine some of the basic theoretical ideas of linear algebra in a practical, computer-oriented context.

-- How can the rank, nullity, range and nullspace of a matrix be computed reliably?

-- How can systems of linear equations be solved efficiently and accurately? How meaningful will the results be?

-- How can the eigenvalues and eigenvectors of a given matrix be determined? What about the Jordan Canonical Form?

The Lectures will be presented by Bootwick Wyman of Ohio State.

The first part of the course will be a "consciousness-raising" experience with many examples. Some fundamental ideas of linear numerical analysis, and the main ideas, if not all the details, of some important algorithms will be presented. Intelligence (and, to some extent, skeptical) use of available mathematical software will be emphasized.

Participants will be invited to use the "Speakeasy" interpretive processor available at the Ohio State University Instruction and Research Computer Center. This powerful and easily learned language is particularly well suited to linear algebra calculations and supplies automatic access to the Argonne Laboratory LINPACK and EISPACK packages.

Registration fee is $25; room and board in University housing costs approximately $65. For further information contact Professor Barbara Miller, Department of Mathematics, Lorain County Community College, Elyria, Ohio 40035.

OLYMPIAD NEWS

The Brazilian Mathematical Society plans to sponsor the First Pan-American Mathematical Olympiad, probably early in 1982, in either Sao Paulo or Rio de Janeiro. Countries interested in participating should contact Professor João Bosco Pitombeira, Departmento do Matematica, Pontificia Universidade Católica do Rio de Janeiro, Rua Marquês de São Vicente 225, 22453 Rio de Janeiro, RJ, Brazil.

THE 1980 U.S.A. MATHEMATICAL OLYMPIAD

The ninth U.S.A. Mathematical Olympiad took place on May 6, 1980, and the problems were published that same month in this column. The following sketches of solutions were adapted by Loren Larson from Samuel Greitzer's pamphlet "The U.S.A. Math Olympiad for 1980." This pamphlet, containing more complete discussion of many of the problems, may be obtained for $.50 from Dr. Walter Mientka, Executive Director, MAA Committee on High School Contests, 917 Oldfather Hall, University of Nebraska, Lincoln, NE 68588.

1. A two pan balance is inaccurate since its balance arms are of different lengths and its pans are of different weights. Three objects of different weights A, B and C are each weighed separately. When placed on the left hand pan, they are balanced by weights A_1, B_1, C_1, respectively. When A and B are placed on the right hand pan, they are balanced by A_2 and B_2, respectively. Determine the true weight of C in terms of A_1, B_1, C_1, A_2 and B_2.

Sol. Let the lengths of the right arm of the balance be k times the length of the left arm, and let L and R denote the left and right pan weights, respectively. We are given:

(1) $A + L = k(A_1+R)$

(2) $B + L = k(B_1+R)$

(3) $C + L = k(C_1+R)$

(4) $A_2 + L = k(A+R)$

(5) $B_2 + L = k(B+R)$.

From (1) and (2), $(A-B) = k(A_1-B_1)$, and from (4) and (5), $(A_2-B_2) = k(A-B)$. It follows that $k^2 = (A_2-B_2)/(A_1-B_1)$.

From (1) and (4), $(A-A_2) = k(A_1-A)$ and therefore $A = (kA_1+A_2)/(k+1)$.

From (1) and (3), $C - A = k(C_1-A_1)$, and from this we can solve for C and use the expressions from the two preceding paragraphs to eliminate A and k. After some algebraic manipulation we arrive at the result: C equals

$$\frac{C_1\sqrt{(A_1-B_1)(A_2-B_2)}+C_1(A_2-B_2)+(A_1B_2-A_2B_1)}{\sqrt{(A_1-B_1)(A_2-B_2)}+(A_1-B_1)}$$

2. Determine the maximum number of different three term arithmetic progressions which can be chosen from a sequence of n real numbers $a_1 < a_2 < \ldots < a_n$.

Sol. Let T_n denote this maximum number, and let A_i denote the number of three-term arithmetic progressions with middle term a_i $(1<i<n)$. Let n be odd, $n = 2k + 1$. Then for $1 < i \leq k$, $A_i \leq i$ i-1 and for $k < i < n$, $A_i \leq n-i$. Therefore

$$T_n = \sum_{i=2}^{n-1} A_i \leq \sum_{i=2}^{k} (i-1) + \sum_{i=k+1}^{n-1} (n-i)$$

$$= \frac{(k-1)k}{2} + \frac{k(k+1)}{2} = \frac{(n-1)^2}{4} .$$

If n is even, we proceed similarly, and find that $T_n \leq ((n-1)^2-1)/4$. These bounds are realized when $a_1,\ldots,a_n$ is an arithmetic progression.

3. Let

$$F_r = x^r\sin(rA) + y^r\sin(rB) + z^r\sin(rC),$$

where x, y, z, A, B, C are real and $A+B+C$ is an integral multiple of π. Prove that if $F_1 = F_2 = 0$, then $F_r = 0$ for all positive integral r.

Sol. For each nonnegative integer r, let

$$G_r = x^re^{irA} + y^re^{irB} + z^re^{irC} .$$

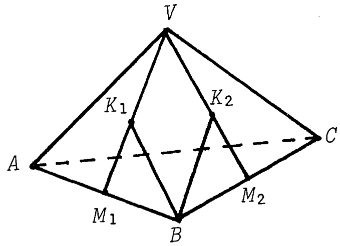
The imaginary part of G_r is equal to F_r. We induct on r. Suppose that $F_k = 0$, $k = 1,\ldots,r$. Then

$$\begin{aligned} G_r\cdot G_1 - G_{r+1} &= xye^{i(A+B)}\left[x^{r-1}e^{i(r-1)A} + y^{r-1}e^{i(r-1)B}\right] \\ &+ xze^{i(A+C)}\left[x^{r-1}e^{i(r-1)A} + z^{r-1}e^{i(r-1)C}\right] \\ &+ yze^{i(B+C)}\left[y^{r-1}e^{i(r-1)B} + z^{r-1}e^{i(r-1)C}\right] \\ &= xye^{i(A+B)}\left[-z^{r-1}e^{i(r-1)C}\right] \\ &+ xze^{i(A+C)}\left[-y^{r-1}e^{i(r-1)B}\right] \\ &+ yze^{i(B+C)}\left[-x^{r-1}e^{i(r-1)A}\right] \\ &= -xyze^{i(A+B+C)}G_{r-2} . \end{aligned}$$

The result ($F_{r+1}=0$) follows after equating the imaginary parts of each side of this equality.

4. The inscribed sphere of a given tetrahedron touches each of the four faces of the tetrahedron at their respective centroids. Prove that the tetrahedron is regular.

Sol. Let K_1 and K_2 be centroids of the triangles VAB and VBC respectively.



Since tangents to a sphere from an external point are equal, we have $VK_1 = VK_2$, $BK_1 = BK_2$, and since $VB = VB$, ΔVBK_1 and ΔVBK_2 are congruent. Now $\angle BVK_1 = \angle BVK_2$, and since $VM_1 = \frac{3}{2}VK_1$ and $VM_2 = \frac{3}{2}VK_2$, it follows that $\Delta BVM_1 = \Delta BVM_2$ are congruent. Therefore $BM_1 = BM_2$, and hence also $BA = BC$.

In the same way, we find that $AB = AC$ so that ΔABC is equilateral. The result follows after repeating the argument with A as vertex, and then with B as vertex.

5. If $1 \geq a,b,c \geq 0$, prove that

$$\frac{a}{b+c+1} + \frac{b}{c+a+1} + \frac{c}{a+b+1} + (1-a)(1-b)(1-c) \leq 1 .$$

Sol. Without loss of generality, $0 \leq a \leq b \leq c \leq 1$. Therefore

$$\frac{a}{b+c+1} + \frac{b}{a+c+1} + \frac{c}{a+b+1} + (1-a)(1-b)(1-c)$$
$$\leq \frac{a+b+c}{a+b+1} + (1-a)(1-b)(1-c)$$
$$= 1 + (c-1)\left[\frac{1}{a+b+1} - (1-a)(1-b)\right] .$$

The result follows after observing that

$$(1+a+b)(1-a)(1-b) \leq (1+a+b+ab)(1-a)(1-b)$$
$$= (1+a)(1+b)(1-a)(1-b)$$
$$= (1-a^2)(1-b^2) \leq 1 .$$

FABRIC REFERENCE

Here are some additional references concerning our paper "Satins and Twills: An Introduction to the Geometry of Fabrics" (this *Magazine*, May 1980, pp. 139-161):

(a) Certain decorative tilings by black and white tiles can be interpreted as designs for twillins and color-alternate twillins; see D. Wade, *Pattern in Islamic Art*, The Overlook Press, Woodstock, NY, 1976, p. 16;

(b) "Norwegian squares," which are closely related to the designs of twillins, are discussed by E.S. Selmer, "Doubly Periodic Arrays," *Computers in Number Theory*, Proceedings of the S.R.C. Atlas Symposium No. 2, Oxford, 1969, edited by A.O.L. Atkin and B.J. Birch, Academic Press, London and New York, 1971;

(c) An efficient algorithm for deciding whether a diagram represents a fabric which "hangs together" and some related results are described in "When a Fabric Hangs Together" by C.R.J. Clapham (to appear in the *Bulletin of the London Mathematical Society*).

Branko Grünbaum
University of Washington
Seattle
Washington 98195

Geoffrey C. Shephard
University of East Anglia
University Plain
Norwich NR4 7TJ

TURKEY ROAST

Readers of "Cooking a Turkey" by John Ruebush and Robert Fisk (this *Magazine*, September 1980, pp. 237-239) may be interested in the note by Murray Klamkin "On Cooking a Roast" in *SIAM Review* (1961) 167-169. Roasting a turkey, it seems, is just like roasting a roast.

Paul J. Campbell
Beloit College
Beloit
Wisconsin 53511

COBWEB CORRECTION

An interesting error appears in Elwyn Davis' proof of the Cobweb Theorem ("Cauchy and Economic Cobwebs," this *Magazine*, May 1980, p. 172). The derivative $(s^{-1})'(q) = 1/s'(p)$ must be evaluated at $p_s = p^{-1}(q)$, whereas the derivative $(d^{-1})'(q) = 1/d'(p)$ must be evaluated at $p_d = d^{-1}(q)$. Except at equilibrium, p_s and p_d are different; therefore $|s'(p)| < |d'(p)|$ for all p does not imply $|(d^{-1})'(q)| < |(s^{-1})'(q)|$ for all q.

As a counterexample to the argument in the proof, consider the functions $d(p) = 100 - e^p$, and $s(p) = .9e^p - .9$, defined on $0 \leq p \leq \ln 100$. Using the Cobweb Theorem notation, observe that $q_E \doteq 46.89$, let $q_1 = 89$, and choose $z = 80$. Then $(d^{-1})'(80) = -1/20 = -.05$, but $(s^{-1})'(80) = 1/80.9 \doteq .012$.

Clifford H. Wagner
Pennsylvania State University
Middletown, PA 17057

A DISCONNECTED ARGUMENT

I just recently received the May 1980 issue of *Mathematics Magazine*, and was glad to see the article "Advanced Plane Topology From an Elementary Standpoint" by Donald E. Sanderson. I have always felt that much of mathematics can be explained in simple ways that give insight into why things happen.

While I like the article very much--if I didn't, I would probably not have been concerned enough to write in like this--I must point out a flaw that it contains. On page 86, just before the statement of the Alexander Duality Theorem, the author argues that it suffices to show the Theorem for S^2 and it will also apply to E^2. While this is true for the theorem as stated, the author continues on--just after the proof of the theorem--and argues that one can show in a similar fashion the 1-connectedness of S^2 with an arc removed. Although this is true, the argument that it will also be true for E^2 breaks down. E^2 is not 1-connected when an arc is removed. (Actually the author doesn't adequately define what is meant by a surface "spanning" a square or a circle for the definition of 1-connectedness, but any reasonable definition will have E^2 not being 1-connected when an arc is removed.) Unfortunately, this "fact" is used in the proof given for the Jordan Curve Theorem, by allowing the use of formula (1) on E^2 with an arc removed.

A more minor problem is in the proof of the Duality Theorem, in the extension to show 1-connectedness of S^2. The Alexander Addition Theorem in higher dimensions, as stated in the article, is used. However, the two surfaces, S_1 and S_2, used here are arbitrary surfaces. In the theorem the surfaces were homeomorphic to a square region. It is not *a priori* clear that an arbitrary surface spanning a square is homeomorphic to a square region, if true at all.

I felt that the above points should be brought to your attention. I continue to look forward to reading more of the interesting and well-written mathematical expositions that I find in *Mathematics Magazine*.

Richard E. Stone
Stanford University
Stanford
California 94305

CLASSICAL REFERENCE FOR MATRIX ROOTS

I recently became aware that the result in my paper "The Square Root of a 2 x 2 Matrix" (this *Magazine*, September 1980, pp. 222-224) is contained in Cayley's original paper "A Memoir on the Theory of Matrices," (*Phil. Trans. R. Soc. London* 148 (1858), *Collected Mathematical Papers*, V. 2, 475-497). This is where the Cayley-Hamilton theorem apparently was first stated explicitly (although Cayley only proved it for 2 x 2 matrices). The present result was presented as a powerful application of the theorem.

Bernard W. Levinger
Colorado State University
Fort Collins
Colorado 80523

PERIODS OF UNIQUE PRIMES

In "The Mystique of Repunits" (this *Magazine*, January 1978, pp. 22-28), it was implied that every natural number is the period length of at least one prime. This could be shown as a corollary of a theorem by G.D. Birkhoff and H.S. Vandiver that appeared in 1904 in *Annals of Mathematics*.

One of the number theoretical lists that I maintain shows every natural number that is known to be the period length of *exactly* one prime. Some of the very large primes have been contributed by participants in a current project known as the Cunningham project. In particular, J. Brillhart, D.H. Lehmer, S. Wagstaff, and H.C. Williams have graciously permitted my use of these numbers.

If a repunit $R_n = (10^n-1)/9$ is prime, its subscript is prime and is the period length of exactly one prime, the repunit itself. This occurs when the subscript is 2, 19, 23, or 317. Prof. H.C. Williams of the University of Manitoba, who discovered in 1977 that R_{317} is prime, is attempting to prove the primality of R_{1031}. He has found that R_{1031} is the only pseudoprime (in several bases) repunit above R_{317} and below R_{2003}. The probability that it is prime is very high.

The subscript of a repunit is the period length of exactly one prime only if the primitive cofactor is prime (that is, the factor of the repunit that is not divisible by any smaller repunit). A repunit is divisible by a smaller repunit if and only if the subscript of the smaller repunit divides the subscript of the larger repunit. In the case of a prime repunit, the subscript is necessarily prime, and the repunit is the same as its primitive cofactor. In the case of a composite repunit, its primitive cofactor is, of course, smaller. The number of digits in the primitive cofactor of repunit R_n is very close to, and frequently equal to Euler's function, $\phi(n)$. The appearance of the primitive cofactor follows a predictable pattern that is also a function of (the factors of) n. We now know 26 numbers that are the period lengths of exactly one prime each. They are shown in the table below.

n	The Only Prime With Period n
1	3
2	$11 = R_2$
3	37
4	101
9	333667
10	9091
12	9901
14	909091
19	1111111111 111111111 = R_{19}
23	1111111111 1111111111 111 = R_{23}
24	99990001
36	9999990000 01
38	9090909090 90909091
39	9009009009 0099099099 0991
48	9999999900 000001
62	9090909090 9090909090 9090909091
93	10 900's followed by 9 990's and a 991
106	25 90's followed by a 91
120	1000099999 9989998999 9000000010 001
134	32 90's followed by a 91
150	1000009999 9999989999 8999990000 0000010000 1
196	9999999999 9999000000 0000000099 9999999999 9900000000 0000009999 9999999999 0000000000 0001
294	1428571571 4285714285 6999999985 7142857142 8585714285 7142855714 2855714285 7142857285 7143
317	R_{317}
586	145 90's followed by a 91
597	66 900's followed by 65 990's and a 991.

Samuel Yates
157 Capri-D, Kings Point
Delray Beach
Florida 33445

KUDOS

This is the last issue of *Mathematics Magazine* that we are editing. Our work as editors has benefited immeasurably from the outstanding work of three persons who deserve special thanks: Raoul Hailpern, Managing Editor, who overcame delays, strikes, and computer chaos with unfailing good humor; Greg Metcalf, St. Olaf art student (now graduated) who provided so many illustrations that gave the *Magazine* its distinctive style; and Mary Kay Peterson, office secretary, who managed correspondence and retyped countless manuscripts with extraordinary speed and precision.

Lynn Arthur Steen
J. Arthur Seebach

ACKNOWLEDGEMENTS

In addition to our associate editors, the following mathematicians have assisted the *Magazine* by refereeing papers during the past year. We appreciate their special efforts.

Albertson, Michael O., *Smith College.*
Albig, David, *Radford College.*
Alexanderson, Gerald L., *University of Santa Clara.*
Andrews, George E., *Pennsylvania State University.*
Anton, Howard, *Drexel University.*
Arrison, John D., *Millikin University.*
Barbeau, Edward J., *University of Toronto.*
Beals, Richard W., *Yale University.*
Berggren, J.L., *Simon Fraser University.*
Berndt, Bruce C., *University of Illinois at Urbana-Champaign.*
Borden, Robert S., *Knox College.*
Buck, R. Creighton, *University of Wisconsin, Madison.*
Burton, David M., *University of New Hampshire.*
Callahan, James, *Smith College.*
Carlson, Kent, *St. Cloud State College.*
Chakerian, G.D., *University of California, Davis.*
Cohen, Joel, *Rockefeller University.*
Coleman, Courtney, *Harvey Mudd College.*
Connelly, Robert, *Cornell University.*
Conroy, John E., *Texas Instruments.*
Corbet, James J., *Radford College.*
Dershem, Herbert, *Hope College.*
Diaconis, Persi W., *Bell Labs.*
Ellison, Robert J., *Hamilton College.*
Emerson, John, *Middlebury College.*
England, James, *Swarthmore College.*
Epp, Susanna, *DePaul University.*
Finkbeiner, Daniel, *Kenyon College.*
Fishback, William T., *Earlham College.*
Gaal, Lisl, *University of Minnesota.*
Gallian, Joseph A., *University of Minnesota, Duluth.*
Grunbaum, Branko, *University of Washington.*
Haimo, Deborah T., *University of Missouri at St. Louis.*
Henkin, Leon, *University of California, Berkeley.*
Herman, Eugene, *Grinnell College.*
Hersh, Reuben, *University of New Mexico.*
Heuer, Gerald, *Concordia College.*
Kleber, Richard, *St. Olaf College.*
Konhauser, Joseph D.E., *Macalester College.*
Larson, Loren, *St. Olaf College.*
Leonard, William, *California State University, Fullerton.*
Loud, Warren, *University of Minnesota.*
Malkevitch, Joseph, *York College (CUNY).*
Merrill, Samuel, *Wilkes College.*
Mills, George H., *St. Olaf College.*
Moore, David S., *Purdue University.*
Moore, Gregory, *University of Toronto.*
Niven, Ivan, *University of Oregon.*
Nummela, Eric, *New England College.*
Packel, Edward W., *Lake Forest College.*
Parry, Charles, *Virg. Poly. Inst.*
Pedersen, Jean J., *Univ. of Santa Clara.*
Peterson, Bruce B., *Middlebury College.*
Prichett, Gordon D., *Sudbury, Mass.*
Richards, Ian, *University of Minnesota.*
Robertson, Timothy J., *Univ. of Iowa.*
Savage, Thomas R., *St. Olaf College.*
Schneider, Dennis M., *Knox College.*
Schue, John, *Macalester College.*
Shapiro, Harold N., *New York University.*
Snow, Wolfe, *Brooklyn College (CUNY).*
Stern, Frederick, *San Jose State University.*
Strotman, Timothy, *St. Olaf College.*
Swetz, Frank, *Pennsylvania State University.*
Tanis, Elliot, *Hope College.*
Tymoczko, Thomas, *Smith College.*
Vanden Eynden, Charles, *Illinois State University, Normal.*
VanIwaarden, John L., *Hope College.*
Weinstein, Michael L., *Passaic, New Jersey.*
Wessels, Linda, *Indiana University, Bloomington.*
Wolf, Frank, *Carleton College.*

U.S. POSTAL SERVICE

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

(Required by 39 U.S.C. 3685)

1. TITLE OF PUBLICATION: MATHEMATICS MAGAZINE
A. PUBLICATION NO.:
2. DATE OF FILING: Sept. 11, 1980
3. FREQUENCY OF ISSUE: every two months except July and August
A. NO. OF ISSUES PUBLISHED ANNUALLY: Five
B. ANNUAL SUBSCRIPTION PRICE: $18.00
4. LOCATION OF KNOWN OFFICE OF PUBLICATION *(Street, City, County, State and ZIP Code) (Not printers)*: 1529 18th Street, N.W., Washington, D.C. 20036
5. LOCATION OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS *(Not printers)*: 1529 18th Street, N.W., Washington, D.C. 20036
6. NAMES AND COMPLETE ADDRESSES OF PUBLISHER, EDITOR, AND MANAGING EDITOR
PUBLISHER *(Name and Address)*: The Mathematical Association of America, 1529 18th St., N.W., Washington, DC 20036
EDITOR *(Name and Address)*: J. Arthur Seebach & Lynn A. Steen, St. Olaf College, Northfield, MN 55057
MANAGING EDITOR *(Name and Address)*: Dr. Raoul Hailpern, SUNY at Buffalo, Dept. of Mathematics, Buffalo, NY 14222
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 giving. If the publication is published by a nonprofit organization, its name and address must be stated.)*

NAME	ADDRESS
The Mathematical Association of America, Inc.	1529 18th Street, N.W. Washington, D.C. 20036

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)*

NAME	ADDRESS
None	

9. FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 132.122, PSM)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

[X] HAVE NOT CHANGED DURING PRECEDING 12 MONTHS [ ] HAVE CHANGED DURING PRECEDING 12 MONTHS *(If changed, publisher must submit explanation of change with this statement.)*

10. EXTENT AND NATURE OF CIRCULATION	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO. COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A. TOTAL NO. COPIES PRINTED *(Net Press Run)*	10,100	9,500
B. PAID CIRCULATION 1. SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES	none	none
2. MAIL SUBSCRIPTIONS	8,283	8,155
C. TOTAL PAID CIRCULATION *(Sum of 10B1 and 10B2)*	8,283	8,155
D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	40	40
E. TOTAL DISTRIBUTION *(Sum of C and D)*	8,323	8,195
F. COPIES NOT DISTRIBUTED 1. OFFICE USE, LEFT OVER, UNACCOUNTED, SPOILED AFTER PRINTING	1,777	1,305
2. RETURNS FROM NEWS AGENTS	none	none
G. TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	10,100	9,500

11. I certify that the statements made by me above are correct and complete. SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER

12. FOR COMPLETION BY PUBLISHERS MAILING AT THE REGULAR RATES *(Section 132.121, Postal Service Manual)*

39 U. S. C. 3626 provides in pertinent part: "No person who would have been entitled to mail matter under former section 4359 of this title shall mail such matter at the rates provided under this subsection unless he files annually with the Postal Service a written request for permission to mail matter at such rates."

In accordance with the provisions of this statute, I hereby request permission to mail the publication named in Item 1 at the phased postage rates presently authorized by 39 U. S. C. 3626.

SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER

PS Form 3526 *(Page 1)* Aug. 1978 (See instructions on reverse)

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

MATHEMATICS MAGAZINE VOL. 53, NO. 5, NOVEMBER 1980