INDEX

AUTHORS

TITLES

P R O B L E M S

Proposals, Solutions and Quickies are indexed below by means of the code letters P, S, *and* Q, *respectively. Page numbers are given in parentheses. Thus,* P1046(194) *refers to proposal number* 1046 *which appears on page* 194.

N E W S A N D L E T T E R S

MATHEMATICS MAGAZINE

Vol. 51, No. 5
November, 1978
CODEN: MAMGA8

ABERRANCY • TOTITIVES
EUCLIDEAN CONSTRUCTION

THE BICENTENNIAL TRIBUTE TO AMERICAN MATHEMATICS

Edited by DALTON TARWATER

This volume is based on the papers presented at the Bicentennial Program of the Association on January 24–26, 1976. In addition to the major historical addresses, the papers cover the following panel discussions: Two-Year College Mathematics in 1976; Mathematics in Our Culture; The Teaching of Mathematics in College; A 1976 Perspective for the Future; The Role of Applications in the Teaching of Undergraduate Mathematics.

The following is a list of the Panelists and the Authors: Donald J. Albers, Garrett Birkhoff, J. H. Ewing, Judith V. Grabiner, W. H. Gustafson, P. R. Halmos, R. W. Hamming, I. N. Herstein, Peter J. Hilton, Morris Kline, R. D. Larsson, Peter D. Lax, Peter A. Lindstrom, R. H. McDowell, S. H. Moolgavkar, Shelba Jean Morman, C. V. Newsom, Mina S. Rees, Fred S. Roberts, R. A. Rosenbaum, S. K. Stein, Dirk J. Struik, Dalton Tarwater, W. H. Wheeler, A. B. Willcox, W. P. Ziemer.

Individual members of the Association may purchase one copy of the book for $7.50; additional copies and copies for nonmembers are priced at $13.00 each. (Orders for under $10.00 must be accompanied by payment. Prepaid orders will be delivered postage and handling free.)

Orders should be sent to:

MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W. Washington, D.C. 20036

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.: MAGA8
2. DATE OF FILING: Sept. 12, 1978	
3. FREQUENCY OF ISSUE: every two months except July and August	A. NO. OF ISSUES PUBLISHED ANNUALLY: Five
	B. ANNUAL SUBSCRIPTION PRICE: $12.00

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)*
Mathematical Association of America, 1529 18th Street, N.W., Washington, D.C. 20036

6. NAMES AND COMPLETE ADDRESSES OF PUBLISHER, EDITOR, AND MANAGING EDITOR

PUBLISHER *(Name and Address)*
Mathematical Association of America, 1529 18th Street, N.W., Washington, D.C. 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 given.)*

NAME	ADDRESS
Mathematical Association of America	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)

☐ 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)*	10,400	10,000
B. PAID CIRCULATION 1. SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES	none	none
2. MAIL SUBSCRIPTIONS	8,472	7,981
C. TOTAL PAID CIRCULATION *(Sum of 10B1 and 10B2)*	8,472	7,981
D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	21	21
E. TOTAL DISTRIBUTION *(Sum of C and D)*	8,493	8,002
F. COPIES NOT DISTRIBUTED 1. OFFICE USE, LEFT OVER, UNACCOUNTED, SPOILED AFTER PRINTING	1,907	1,998
2. RETURNS FROM NEWS AGENTS	none	none
G. TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	10,400	10,000

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

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 Mar. 1977 *(Page 1)* (See instructions on reverse)

Sixth Edition 1975

GUIDEBOOK

TO
DEPARTMENTS IN THE
MATHEMATICAL SCIENCES
IN THE
UNITED STATES AND CANADA

. . . intended to provide in summary form information about the location, size, staff, library facilities, course offerings, and special features of both undergraduate and graduate departments in the Mathematical Sciences . . .

100 pages, 1350 entries.

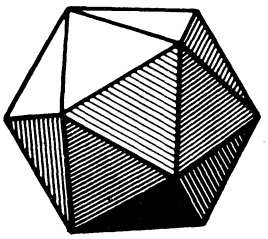
Price: $3.00

Orders with remittance should be sent to:

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

MATHEMATICS MAGAZINE

Volume 51, Number 5, November 1978



COVER: Getting around on one-way downtown streets is a famous graph theory problem. If, like the late Herbert Hoover, you don't like making left turns, it becomes a much more subtle problem which is analyzed on p. 288.

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

INDEX

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.

Editorial correspondence should be sent to: Mathematics Magazine, Department of Mathematics, St. Olaf College, Northfield, Minnesota 55057. 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.

Authors planning to submit manuscripts may find it helpful to obtain the more detailed statement of guidelines available from the editorial office.

ABOUT OUR AUTHORS

Steven H. Schot ("Aberrancy: Geometry of the Third Derivative") is a professor of mathematics at the American University in Washington, D.C. where he has taught since obtaining his Ph.D. from the University of Maryland in 1958. His research interests lie in the areas of partial differential equations and integral equations and their applications to fluid dynamics. In addition, his interest in the history of mathematics has led him and a number of students to study the history of aberrancy and to develop additional results which correct errors in earlier works. This paper grew out of an invited address given to the Maryland-D.C.-Virginia Section meeting of the M.A.A. in April 1977.

Steven Galovich ("Unique Factorization Rings with Zero Divisors") grew up in the Great Central Valley of California. He did his undergraduate work at the University of California-Davis and received his Ph.D. from Brown. Since 1971 he has taught at Carleton College, Northfield, Minnesota and has just completed a term of service as Secretary-Treasurer of the North Central Section of the M.A.A. His research interests are in the areas of algebra and algebraic number theory. This article grew out of his attempts to expose his number theory students to rings which for various reasons failed to possess the unique factorization property. He is spending the 1978–79 academic year as a Visiting Professor at Davis.

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 $12 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.

Aberrancy: Geometry of the Third Derivative

The third derivative is represented geometrically in terms of the quantity called aberrancy, which measures the asymmetry of a curve about its normal.

STEVEN H. SCHOT
American University
Washington, DC 20016

The first and second derivatives play a fundamental role in the study of the behavior of a curve near a point. Their geometrical interpretation in terms of the tangent and curvature at the point is familiar to every student of elementary calculus. The geometrical significance of the third derivative in terms of the osculants to the curve is much less well-known and does not seem to be discussed in any current textbook on analysis. (The most recent calculus text I could find which treats the aberrancy is Edwards [**10**], the last (3rd) edition of which was published in 1900!) Yet, the third derivative—and its interpretation as the so-called aberrancy of a curve at a point—sheds important light on the local behavior of the curve and requires for its study only elementary concepts and results from the calculus. It is the purpose of this paper to present the important results on aberrancy in an elementary, explicit, and comprehensive way, so as to encourage exploration of this neglected topic and to communicate a number of new results.

To fix the ideas, we review briefly the geometrical meaning of the first two derivatives and see how these concepts may be extended to the third derivative. The standard interpretation of the first derivative employs the slope of the tangent line to the curve. If $y=f(x)$ is a function, the first derivative

$$y' = \tan\theta = m, \tag{1}$$

where prime denotes differentiation with respect to x, computes the slope of the curve representing the function at a point P, namely the inclination which the tangent line to the curve at P makes with the x-axis. The tangent line at P may be thought of as the limit of a secant line through P and a neighboring point P_1 on the curve as P_1 approaches P. Thus the tangent line is said to make **2-point contact** with the curve at P. Alternatively, since the tangent line has the same y and y' as the curve at P, it may also be said to make **1st-order contact** with the curve at this point. Many of the fruitful applications of the differential calculus are based on the first derivative. Indeed, in one of the earliest treatises on this subject—Johannes Kepler's *Stereometria doliorum* (1615)—the author uses the first derivative to design various nearly cylindrical barrels so as to hold a maximum amount of wine under given constraints. Apparently Kepler was prompted to write this book on his honeymoon because of an unsatisfactory purchase of a barrel of wine for his wedding festivities!

The second derivative y'' may be interpreted geometrically in terms of the curvature. The **curvature** κ of a curve at a point is generally defined to be the rate of change of direction with respect to arclength, $d\theta/ds$, at the point. Expressing the direction of the curve in terms of the slope angle θ, we may use (1) and the Pythagorean relation for ds in terms of dx and dy to express the curvature as

$$\kappa=\frac{d\theta}{ds}=\frac{d\theta}{dx}\cdot\frac{dx}{ds}=\frac{y''}{1+(y')^2}\cdot\frac{1}{\sqrt{1+(y')^2}}=\frac{y''}{[1+(y')^2]^{3/2}}. \tag{2}$$

It is shown in calculus that the curvature is equal to the reciprocal of the so-called **radius of curvature** ρ, the radius of the circle which has the same curvature at the point as the curve itself. The circle of curvature at P is at the same time the **osculating circle** at this point, i.e., the limit of the circles through P and two neighboring points P_1 and P_2 on the curve as P_1 and P_2 approach P (**3-point contact**). Equivalently, the osculating circle has y, y', and y'' equal to those of the curve at P (**2nd-order contact**). Again, many useful applications of the second derivative and the associated geometrical concept of curvature are known from the differential calculus. Perhaps the earliest of these is the Dutch physicist, Christiaan Huygens', design of an isochronous clock in his *Horologium oscillatorium* (1673).

How may these ideas now be extended to yield a geometrical representation for the third derivative? Merely interpreting y''' as the rate of change of y'' with respect to x yields little useful information. What is needed is a geometrical quantity, much like the slope for y' or the curvature for y'', from which immediate insight can be gained about the behavior of the curve representing $y=f(x)$ at the point. Such a quantity is the **aberrancy** of a curve at a point P, namely the tangent of the angle δ formed between the normal at P and the limiting position of a line drawn from P to the midpoint of a chord parallel to the tangent line at P as the chord approaches P. By carrying out this limiting process, we shall show below that the aberrancy is given by

$$\tan\delta=\frac{\dot{\rho}}{3\rho}=y'-\frac{1+(y')^2}{3(y'')^2}y''', \tag{3}$$

where dot denotes differentiation with respect to the slope angle θ and prime means differentiation with respect to x. Hence aberrancy is a quantity which indeed depends on the third derivative. Clearly, this quantity is the tangent of an angle which is easily constructed and visualized. The aberrancy at a point P of a curve is thus a measure of the asymmetry of the curve with respect to the normal line through P.

As in the case of the lower-order derivatives, it is possible to give an alternate definition of aberrancy in terms of the osculants to the curve. This time the appropriate osculant is the **osculating parabola**, i.e., the parabola at P which is the limit of the parabolas through P and three neighboring points P_1, P_2, and P_3 of the curve as P_1, P_2, and P_3 approach P (**4-point contact**). Again, the osculating parabola is also the parabola whose first three derivatives are equal to those of the curve at the point (**3rd-order contact**). The aberrancy at P may now be defined alternatively as the tangent of the angle between the axis of the osculating parabola and the normal to the curve at P. It should be mentioned that in addition to the geometrical applications presented in this paper, the third derivative and the aberrancy also have applications in mechanics, where the third derivative of distance with respect to time is known as the **jerk**. (Added in proof: The author's paper "Jerk: The Time Rate of Change of Acceleration" on this topic will appear in the January 1979 issue of the *Amer. J. of Physics*.) Before deriving the aberrancy and quantities related to it, we first review briefly the interesting history of the aberrancy concept.

History of the Concept of Aberrancy

The concept of aberrancy—but no name for it—appears for the first time [2] in a short analytical addendum to Lazare N. M. Carnot's 1803 treatise on synthetic geometry [7]. Lazare Carnot (1753-1823), father of the celebrated physicist Sadi Carnot after whom the ideal heat engine cycle is

named, introduced this quantity in connection with his search for intrinsic coordinates [9], i.e., elements in terms of which a curve may be expressed independently of any external frame of reference. He recognized the radius of curvature and the arclength as two such elements and considered as another: "the angle which is formed, at the point describing the curve, by the tangent and the line which bisects the infinitesimally small chords drawn through the curve parallel to this tangent." Carnot then derived an expression for this angle, which he denoted by z, i.e., the complement of the modern angle of aberrancy. His result is incorrect however, as Carda [6] has pointed out. It does contain enough of the correct elements, though, for Carnot to be able to show that for a circle, this angle is 90° (i.e., in modern terminology, that the aberrancy vanishes at every point of a circle). Moreover, as an illustration of the use of z as an intrinsic coordinate, Carnot showed that the parabola whose Cartesian equation is $y^2=px$ assumes the intrinsic form $\rho=(1/2)p\ \csc^3 z$, where ρ is the radius of curvature.

In 1841 another French geometer, Abel Transon [16], who does not mention Carnot's work and may not have been aware of it, recognized the importance of aberrancy in the study of plane curves and surfaces and makes extensive use of it in his construction of the osculating conics to a curve at a point. In place of Carnot's angle z, he uses its complement δ and coins the term "*déviation*" for $\tan\delta$, since he considers it to be a measure of the departure of a curve from its circle of curvature. Transon derives the two expressions in (3) for the aberrancy and points out its dependence on the third derivative. In addition, Transon defines the related notions of axis, radius, and center of aberrancy and derives expressions for the first two of these. He then uses these concepts to construct geometrically the third- and fourth-order osculants to a curve at a point. In particular, he shows that the members of the family of conics which make third-order (4-point) contact with the given curve at a point all have their centers on the axis of aberrancy, and that the unique central conic which makes fourth-order (5-point) contact with the curve—the so-called osculating conic—has the center of aberrancy for its center.

Transon's term *déviation* was first translated into English as "aberrancy" by George Salmon (1819-1904), professor of mathematics and professor of divinity(!) at Trinity College in Dublin and well-known for his popular texts on conic sections, solid analytical geometry, and higher algebra. The inclusion of the topic aberrancy in his book on higher plane curves [14] and the discussion of the osculating conics there may well have contributed to the subsequent appearance of this concept in British calculus texts of the time [10] and its popularity on mathematical tripos examinations [8].

In a 1915 address [19] as retiring chairman of the Chicago Section of the American Mathematical Society, Ernest J. Wilczynski stressed the importance of Transon's basic investigation for the differential geometry of plane curves. In this paper he also outlines how the osculants of orders five through eight may be defined in terms of certain cubic curves. From about this time onward the concept of aberrancy is absorbed and developed in the subject of affine differential geometry. Thus the axis of aberrancy now appears as the "affine normal" and the curve of aberrancy becomes the "affine evolute" in modern texts on differential geometry [1], [12], [13]. An interesting application of the classical aberrancy appears in a paper by Walker [17], who finds this concept indispensable in deriving the intrinsic differential equations of certain plane curves.

Derivations of Aberrancy, Radius of Aberrancy, and Center of Aberrancy

In order to derive explicit expressions for the aberrancy and related concepts, it is convenient to have available the so-called canonical expansion of a curve near a point [1]. This expansion may be obtained from the parametric Maclaurin series expansion of the curve at the point P in terms of arclength s along the curve:

$$x=a_0+\frac{a_1}{1!}s+\frac{a_2}{2!}s^2+\frac{a_3}{3!}s^3+\cdots$$

$$y=b_0+\frac{b_1}{1!}s+\frac{b_2}{2!}s^2+\frac{b_3}{3!}s^3+\cdots.$$

Fixing P to be the point from which arclength is measured by imposing $s=0$ at $x=0=y$, we have

$$a_0=0, \qquad b_0=0,$$

and aligning the tangent direction along the curve with the x-axis by requiring that $\theta=0$ at $x=0=y$, we obtain

$$a_1=\left(\frac{dx}{ds}\right)_0=1, \qquad b_1=\left(\frac{dy}{ds}\right)_0=0,$$

where the subscripts on the derivatives denote evaluation at $s=0$. The higher-order coefficients in the expansion may be expressed in terms of the radius of curvature ρ and its derivatives with respect to θ by differentiating $dx/ds=\cos\theta$ and $dy/ds=\sin\theta$, using $d\theta/ds=1/\rho$, $d^2\theta/ds^2=-\dot{\rho}/\rho^3,\ldots$, and evaluating the resulting derivatives at $s=0$:

$$\begin{aligned} a_2&=\left(\frac{d^2x}{ds^2}\right)_0=0, & b_2&=\left(\frac{d^2y}{ds^2}\right)_0=\frac{1}{\rho} \\ a_3&=\left(\frac{d^3x}{ds^3}\right)_0=-\frac{1}{\rho^2}, & b_3&=\left(\frac{d^3y}{ds^3}\right)_0=-\frac{\dot{\rho}}{\rho^3} \\ a_4&=\left(\frac{d^4x}{ds^4}\right)_0=\frac{3\dot{\rho}}{\rho^4}, & b_4&=\left(\frac{d^4y}{ds^4}\right)_0=-\frac{\rho^2-3\dot{\rho}^2+\rho\ddot{\rho}}{\rho^5} \\ &\vdots & &\vdots \end{aligned}$$

Substituting these coefficients into the above Maclaurin series yields the **canonical expansion** of a curve tangent to the x-axis at P:

$$\begin{aligned} x&=s-\frac{1}{6\rho^2}s^3+\frac{\dot{\rho}}{8\rho^4}s^4+\cdots \\ y&=\frac{1}{2\rho}s^2-\frac{\dot{\rho}}{6\rho^3}s^3-\frac{\rho^2-3\dot{\rho}^2+\rho\ddot{\rho}}{24\rho^5}s^4+\cdots. \end{aligned} \tag{4}$$

Alternatively, the variable y may also be expressed directly in terms of x by assuming a general expansion of the form $y=Ax^2+Bx^3+Cx^4+\cdots$, substituting from (4), and equating coefficients. The result is

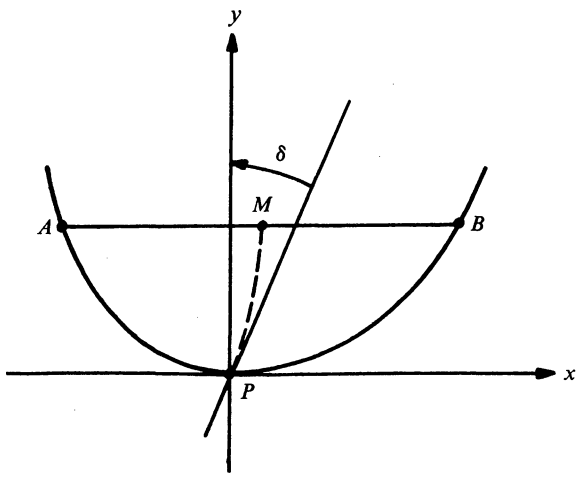
$$y=\frac{1}{2\rho}x^2-\frac{\dot{\rho}}{6\rho^3}x^3+\frac{3\rho^2+3\dot{\rho}^2-\rho\ddot{\rho}}{24\rho^5}x^4+\cdots. \tag{5}$$

These expansions now permit a simple derivation of the aberrancy at P. In FIGURE 1, let the given curve be tangent to the x-axis at the origin P, let AB be a chord parallel to the tangent at P, and let M be the midpoint of AB. As the parallel chord approaches the tangent, the angle between PM and the normal may approach a limiting angle δ. This angle, if it exists, is called the **angle of aberrancy** and the limiting position of PM is called the **axis of aberrancy**. To obtain an expression for δ, let us first find the canonical expansion of the curve for x in terms of y in the neighborhood of P. The expressions (4) and (5) suggest that this expansion has the form $x=\pm ay^{1/2}+by\pm cy^{3/2}+\cdots$, where the plus and minus signs refer to the two branches of the curve as seen from P. Indeed, squaring this expansion, substituting (4), and equating coefficients yields $a=\sqrt{2\rho}$, $b=\dot{\rho}/3\rho$. Thus the abscissas of the points A and B in FIGURE 1 are given by

$$x_A=-\sqrt{2\rho}\,y^{1/2}+\frac{\dot{\rho}}{3\rho}y-\cdots \quad \text{and} \quad x_B=\sqrt{2\rho}\,y^{1/2}+\frac{\dot{\rho}}{3\rho}y+\cdots$$

and, on applying the definition of aberrancy, we obtain

$$\tan\delta=\lim_{y\to 0}\frac{x_A+x_B}{2y}=\frac{\dot{\rho}}{3\rho}. \tag{6}$$



Derivation of aberrancy.

FIGURE 1

The resulting expression is proportional to the first differential invariant of affine transformations in the plane, *viz.*,

$$\frac{1}{\rho}\frac{d\rho}{d\theta}=\frac{d\rho}{ds}.$$

Hence this expression is valid at an arbitrary point of any curve, not necessarily a point at which the curve is tangent to the x-axis. On using this fact and equation (2), the aberrancy at an arbitrary point of any curve may be expressed in Cartesian form as

$$\tan\delta=y'-\frac{1+(y')^2}{3(y'')^2}y'''. \tag{7}$$

This form exhibits the explicit dependence of the aberrancy on the *third* derivative. Both of the formulas (6) and (7) were first derived correctly by Transon [**16**], but in a different manner from that which we have developed here.

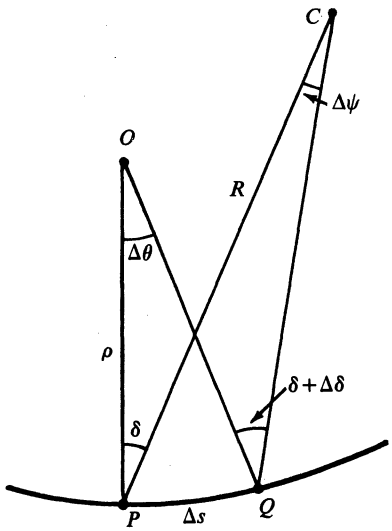
From (6) it is obvious that for a circle ($\rho=$const.) the aberrancy vanishes, i.e., the axis of aberrancy lies along a diameter of the circle. This explains the use of the term "aberrancy" as a measure of the deviation of a curve from its circle of curvature at the point. It also follows immediately from (6) that the only curve whose aberrancy is a non-zero constant at every point, i.e.,

$$\frac{1}{\rho}\frac{d\rho}{d\theta}=\frac{d\rho}{ds}=\text{const.},$$

is $\rho=cs$, namely a logarithmic spiral. (The statement in Boyer [**3**], [**4**] that "... the aberrancy of a conic (is) the same for all points" is incorrect.)

In calculus, the **center of curvature** O of a point P on a curve may be defined as the limiting position of the point of intersection of neighboring normals at P, and the **radius of curvature** ρ at P may then be shown to be the length of the line segment PO. Analogously, the limiting position C of the point of intersection of neighboring axes of aberrancy at P is called the **center of aberrancy** and the length PC is termed the **radius of aberrancy** and denoted by R.

To derive an expression for R we consider two neighboring configurations at P and Q as shown in FIGURE 2. If O and C are the centers of curvature and aberrancy of P, respectively, then the angle OPC between the normal PO and the axis of aberrancy PC is the angle of aberrancy δ. As seen from the neighboring point Q, the angle OQC between the neighboring normal QO and the neighboring axis of aberrancy QC is δ incremented by $\Delta\delta$. The angle POQ between the neighboring normals is also equal to the angle between the two neighboring tangents at P and Q and hence is the increment in the slope angle, $\Delta\theta$. On letting the angle PCQ between the two neighboring axes of aberrancy be



Derivation of radius of aberrancy.

FIGURE 2

$\Delta\psi$, transferring the angles OPC and OQC to the point of intersection of PC and QO, and equating angles at this point, it follows immediately that $\Delta\psi=\Delta\theta-\Delta\delta$. On using the angle $\Delta\theta$ and the radius ρ, the increment in the arclength along the curve between P and Q can be expressed to the first order as $\Delta s=\rho\Delta\theta$. On using the angle $\Delta\psi$ and the radius R, this increment can also be expressed to the first order as $\Delta s=(R\Delta\psi)\sec\delta$. Equating these two expressions for Δs and using the above relation between increments in the angles to eliminate $\Delta\psi$, yields $\rho\Delta\theta=(R\sec\delta)(\Delta\theta-\Delta\delta)$. Dividing this equation through by $\Delta\theta$ and taking the limit as Q approaches P, leads to the following expression for the radius of aberrancy:

$$R=\frac{\rho\cos\delta}{1-\dot{\delta}}. \tag{8}$$

On differentiating (6) and substituting in (8) this radius may be expressed solely in terms of ρ and its derivatives as

$$R=\frac{3\rho^2\sqrt{9\rho^2+\dot{\rho}^2}}{9\rho^2+4\dot{\rho}^2-3\rho\ddot{\rho}}. \tag{9}$$

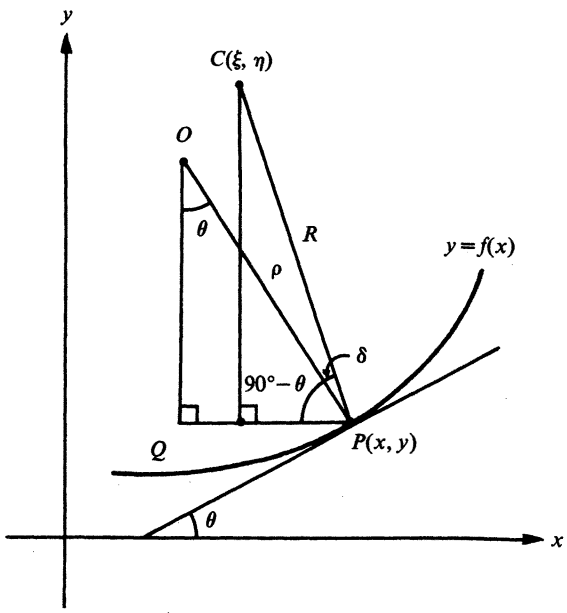
Both of the expressions (8) and (9) appear in Transon [**16**]. Again, on using (1) and (2), the Cartesian form of the radius of aberrancy is found to be

$$R=\frac{3y''\sqrt{(y''')^2+\left[y'y'''-3(y'')^2\right]^2}}{3y''y^{(iv)}-5(y''')^2}. \tag{10}$$

This shows that the radius of aberrancy depends on the *fourth* derivative at P, as well as on the lower order derivatives.

For a circle ($\rho=$const., $\delta=0$), both the expressions (8) and (9) immediately reduce to $R=\rho$. Hence the radius of aberrancy of a circle is the radius of the circle. For a logarithmic spiral ($\delta=$non-zero const.) it follows from (8) that the radius of aberrancy is the projection of the radius of curvature onto the axis of aberrancy.

The coordinates of the center of aberrancy (ξ,η) can now be obtained by an application of trigonometry. In FIGURE 3, let O and C denote the centers of curvature and aberrancy, and let ρ and R be the radii of curvature and aberrancy, respectively, at an arbitrary point P of a curve. Then, since OP makes an angle of $90°-\theta$ with the x-axis and the angle CPO is δ, it follows that CP makes an



Derivation of center of aberrancy.

FIGURE 3

angle of $90° - \theta + \delta$ with the x-axis. Letting Q be the projection of C on the horizontal through P, the following relations then hold for the right triangle CPQ:

$$x - \xi = R \cos (90^0 - \theta + \delta) = R \sin (\theta - \delta)$$
$$\eta - y = R \sin (90^0 - \theta + \delta) = R \cos (\theta - \delta).$$

On using (1), (7), and (10), this yields the Cartesian form of the center of aberrancy:

$$\xi = x - \frac{3y''y'''}{3y''y^{(iv)} - 5(y''')^2}$$
$$\eta = y + \frac{3y''[3(y'')^2 - y'y''']}{3y''y^{(iv)} - 5(y''')^2}. \tag{11}$$

The denominator in formulas (10) and (11) is proportional to the so-called **affine curvature** in differential geometry [1], namely

$$K = -\frac{1}{2}[(y'')^{-2/3}]'' = \frac{1}{9}(y'')^{-8/3}[3y''y^{(iv)} - 5(y''')^2]. \tag{12}$$

For a central conic the affine curvature is a non-zero constant, while for a parabola it is zero. Hence it follows from (11) that a parabola has its center of aberrancy at infinity.

As the point P moves along the given curve, the center of aberrancy will in general change. The locus of centers of aberrancy of a given curve is called its **curve of aberrancy**. It is thus analogous to the **evolute** of a curve which in calculus is defined as the locus of centers of curvature of the given curve. A more detailed study of the aberrancy curve is deferred until later.

Contact and Osculating Conics

We noted above that the first derivative at a point determines a unique tangent line which makes 2-point contact with the curve at that point. Similarly, the second derivative at a point, together with the first derivative, fixes a unique circle which makes 3-point contact with the curve at this point—the

so-called circle of curvature or osculating circle. We now show that in a similar way the third derivative, together with the first and second derivatives, determines a unique parabola which makes 4-point contact with the curve at this point—the so-called osculating parabola. Going one step further, the fourth derivative at a point, together with those of lower order, singles out a unique osculating (5-point) central conic at the point.

To make these ideas more precise we first review some results on the order of contact of curves [11], [9]. Two plane C^{n+1} curves (a curve given by $y=f(x)$ is said to be C^n if and only if its derivatives up to and including order n are continuous) given by $y=f(x)$ and $y=g(x)$ are said to make ***n*th-order contact** at the point $x=a$ if and only if

$$f^{(k)}(a)=g^{(k)}(a) \qquad (k=0, 1, 2,\ldots,n)$$
$$f^{(n+1)}(a)\neq g^{(n+1)}(a).$$

These conditions may be shown to be necessary and sufficient for the curves to have $(n+1)$ coincident points of intersection at $x=a$. Hence we also say that the two curves make **$(n+1)$-point contact** at $x=a$. Now let one of the curves be fixed and given implicitly by $F(x, y)=0$. Let there be given in addition an $(n+1)$-parameter family of C^{n+1} curves of the form $G(x, y; \alpha_1, \alpha_2,\ldots,\alpha_{n+1})=0$. The $(n+1)$ parameters $\alpha_1, \alpha_2,\ldots,\alpha_{n+1}$ can then be determined (from the derivatives of $F(x, y)=0$) so as to single out a unique member of the family, say $G(x, y; \mathring{\alpha}_1, \mathring{\alpha}_2,\ldots,\mathring{\alpha}_{n+1})=0$, which makes nth-order contact [$(n+1)$-point contact] with $F(x, y)=0$ at a point P. The member of the family determined in this manner which makes nth-order contact with the given curve is called an ***n*th-order osculant** of the given curve at P. Clearly an osculant to a given curve at a point is not unique, but depends on the $(n+1)$-parameter family chosen for making contact.

Now consider an arbitrary C^4 curve which, for convenience, is taken to be tangent to the x-axis at the origin so that it can be represented by the canonical expansion (5). In order to find osculants of order two through four to this curve at the origin, it suffices to let the $(n+1)$-parameter family be the family of conics tangent to the x-axis at the origin. Imposing the condition that the family be thus located and oriented already determines two of the five parameters in the equation of the general conic $Ax^2+2Bxy+Cy^2+Dx+Ey+F=0$, namely $F=0$ and $D=0$. Thus the family of *conics* tangent to the x-axis at the origin is the 3-parameter family

$$y=\alpha x^2+2\beta xy+\gamma y^2. \tag{13}$$

This family consists of ellipses, parabolas, or hyperbolas depending on whether the discriminant $\Delta=\alpha\gamma-\beta^2$ is positive, zero, or negative, respectively. In the special case where $\beta=0$, $\alpha=\gamma$, the family reduces to circles. Thus, if we impose the additional condition ($\Delta=0$) that the family be the family of *parabolas* tangent to the x-axis at the origin, then (13) reduces to the 2-parameter family

$$y=\frac{1}{\alpha}(\alpha x+\beta y)^2. \tag{14}$$

Furthermore, for the case of the family of *circles* tangent to the x-axis at the origin, (13) reduces to the 1-parameter family

$$y=\alpha(x^2+y^2). \tag{15}$$

We now use the families of circles (15), parabolas (14), and conics (13) to make 2nd-, 3rd-, and 4th-order contact with the given curve at the origin, respectively. The resulting 2nd-, 3rd-, and 4th-order osculants will be called the osculating-circle, -parabola, and -conic, respectively. In each case, the functional value and the first derivative of the function describing the given curve have already been utilized to ensure that the family in question is tangent to the x-axis at the origin. It remains to determine the parameters α; α and β; and α, β, and γ in each of the three cases from the higher derivatives of the given curve at the origin. These parameters may be obtained by substituting the canonical expansion (5) of the given curve into (15), (14), and (13), respectively, and equating coefficients.

Case 1. Contact of 2nd order: Osculating circle.

The family of circles (15) is used to make 2nd-order contact with the given curve. Substituting (5) into (15) yields

$$\alpha = \frac{1}{2\rho}.$$

This parameter is thus determined in terms of derivatives up to and including order *two*. Substituting α into (15) yields the unique osculating circle

$$x^2+y^2-2\rho y=0$$

to the given curve at the origin. The radius of this osculating circle is the radius of curvature ρ and its center is the center of curvature $(0,\rho)$.

Case 2. Contact of 3rd order: Osculating parabola.

The family of parabolas (14) is used to make 3rd-order contact with the given curve. Substituting (5) into (14) and using (6) yields

$$\alpha = \frac{1}{2\rho}, \qquad \beta = -\frac{\dot{\rho}}{6\rho^2} = -\frac{1}{2\rho}\tan\delta.$$

The parameter β is thus determined in terms of derivatives up to and including order *three*. Substituting α and β into (14) yields the unique osculating parabola

$$(x-y\tan\delta)^2-2\rho y=0 \tag{16}$$

to the given curve at the origin. Its focus is

$$\left(-\frac{\rho}{4}\sin 2\delta, \frac{\rho}{2}\cos^2\delta\right),$$

its directrix is

$$x\tan\delta+y+\frac{\rho}{2}=0,$$

and its axis is

$$x-y\tan\delta+\frac{\rho}{2}\sin 2\delta=0.$$

The axis of the osculating parabola is parallel to the axis of aberrancy at the origin, namely $y=x\cot\delta$. As noted earlier, this fact may be used to give an alternative definition of aberrancy which does not involve a limit argument explicitly.

Case 3. Contact of 4th order: Osculating conic.

The family of conics (13) is used to make 4th-order contact with the given curve. Substituting (5) into (13) yields

$$\alpha = \frac{1}{2\rho}, \qquad \beta = -\frac{\dot{\rho}}{6\rho^2}, \qquad \gamma = \frac{9\rho^2+5\dot{\rho}^2-3\rho\ddot{\rho}}{18\rho^3}. \tag{17}$$

The parameter γ is thus determined in terms of derivatives up to and including order *four*. Substituting α, β, and γ into (13) yields the following unique osculating conic to the given curve at the origin

$$(3\rho x-\dot{\rho}y)^2+(9\rho^2+4\dot{\rho}^2-3\rho\ddot{\rho})y^2-18\rho^3 y=0, \tag{18}$$

or more simply expressed, using (6), (8), and (9),

$$(x-y\tan\delta)^2+(\sec^2\delta)(1-\delta')y^2-2\rho y=0.$$

The discriminant of this conic is

$$\Delta=\alpha\gamma-\beta^2=\frac{1}{36\rho^4}(9\rho^2+4\dot{\rho}^2-3\rho\ddot{\rho})=\frac{\sec^2\delta}{4\rho^2}(1-\dot{\delta}). \tag{19}$$

The conic is an ellipse, parabola, or hyperbola depending on whether Δ is positive, zero, or negative, respectively. Thus, a point on the original curve at which the osculating conic is an ellipse, parabola, or hyperbola is called a point of **elliptic**, **parabolic**, or **hyperbolic curvature**, respectively [1]. The center of the osculating conic (18) is

$$\left(\frac{3\dot{\rho}\rho^2}{9\rho^2+4\dot{\rho}^2-3\rho\ddot{\rho}},\quad \frac{9\rho^3}{9\rho^2+4\dot{\rho}^2-3\rho\ddot{\rho}}\right).$$

This center lies on the axis of aberrancy $y=(3\rho/\dot{\rho})x$ and at a distance

$$\frac{3\rho^2\sqrt{9\rho^2+\dot{\rho}^2}}{9\rho^2+4\dot{\rho}^2-3\rho\ddot{\rho}}=R$$

from the origin. Comparison with (9) thus shows that the center of the osculating conic is the center of aberrancy! This affords an alternative definition and easier method for deriving the radius and center of aberrancy.

Let us now apply the results of Case 3 to an arbitrary conic as the given curve. Then, since it is easily proved that the osculating conic to the given curve at any point is the conic itself, a number of aberrancy properties of an arbitrary conic follow immediately [**14**]:

(a) The center of aberrancy is the center of the conic (in the case of a parabola the center is at infinity).
(b) The axis of aberrancy at any point P is the diameter through P.
(c) The angle of aberrancy at any point P is the angle between the diameter and the normal at P.

Another fact is of interest. From (6), (8), and (9) the discriminant (19) of the osculating conic may be written in the Cartesian form

$$\Delta=\frac{3y''y^{(iv)}-5(y''')^2}{36(y'')^2[1+(y')^2]}. \tag{20}$$

Since any point at which this discriminant vanishes is a point of parabolic curvature, the identical vanishing of this expression yields an intrinsic equation of the parabola. Hence this intrinsic equation may be written $3y''y^{(iv)}-5(y''')^2=0$ or, more compactly, $[(y'')^{-2/3}]''=0$. (For a different derivation of this result, see [**17**]; see also the remark made in the last section concerning the affine curvature (12).)

Penosculating Conics

In the last section we found the equation of the osculating conic (18) to a given C^4 curve by prescribing the three parameters α, β, and γ which appear in the general family of conics tangent to the x-axis at the origin (13). This osculating conic in general makes 4th-order (5-point) contact with the given curve at the origin, for it is evident from the form of the coefficients (17) that derivatives up to and including order *four* are required for their determination. On examining the coefficients in more detail, it appears, however, that only γ actually depends on the fourth derivative, whereas α and β involve only the first three derivatives. Thus, if γ is allowed to be an arbitrary parameter and α and β are prescribed as before, then there results a 1-parameter family of conics tangent to the x-axis at the origin which makes 3rd-order contact with the given curve. This family of conics has been termed [**14**] the family of **4-point conics** or [**19**] the family of **penosculating conics** (since its order of contact falls short of that of the osculating conic by a single unit).

Allowing γ to be an arbitrary parameter and letting α and β defined by (17) and (6), the family of penosculating conics may be written

$$x^2-2xy\tan\delta+2\gamma\rho y^2-2\rho y=0. \tag{21}$$

All of these conics pass through the origin and have the same tangent direction and curvature there as the given curve. In addition, since the third derivative has also been prescribed at the origin, they have the same aberrancy there, i.e., all of their centers

$$\left(\frac{\tan\delta}{4\rho\Delta},\ \frac{1}{4\rho\Delta}\right) \quad \text{where} \quad \Delta=\alpha\gamma-\beta^2=\frac{1}{4\rho^2}(2\gamma\rho-\tan^2\delta)$$

lie on the axis of aberrancy $y=x\cot\delta$. The family consists of ellipses, hyperbolas, and a single parabola depending on whether the discriminant Δ is positive, negative, or zero, respectively. The unique parabola occurs for $\gamma=\beta^2/\alpha=(\tan^2\delta)/2\rho$ and is of course the osculating parabola (16). A degenerate (double tangent line) conic also occurs for $\gamma=\infty$. Among the members of this family of 4-point conics, there is one which makes at least 5-point contact with the given curve. This unique conic is the osculating conic discussed in the last section, and it occurs when γ is prescribed as per (17).

The family of penosculating conics has a number of interesting properties; to study them in detail here would carry us too far afield from our primary objective, which is the study of the aberrancy *per se*. However, as an illustration of the type of results that may be obtained by a further application of aberrancy to the geometry of plane curves, we list here without proof three of the most important properties of the family of penosculating conics. For details in the derivations and proofs the interested reader may consult [**15**].

First of all, it may be shown that as the parameter γ in the family of penosculating conics (21) changes, the center of the corresponding member of the family moves along the axis of aberrancy, while the axes of this conic rotate at the same time. This makes one expect that the family of *axes* of the penosculating conics themselves, say $\mathcal{F}(x, y;\ \gamma)=0$, have an envelope, i.e., a curve to which the axes are all tangent. Indeed this envelope may be found by the usual method of calculus upon eliminating γ between $\mathcal{F}(x, y;\ \gamma)=0$ and $\partial\mathcal{F}/\partial\gamma=0$ and is

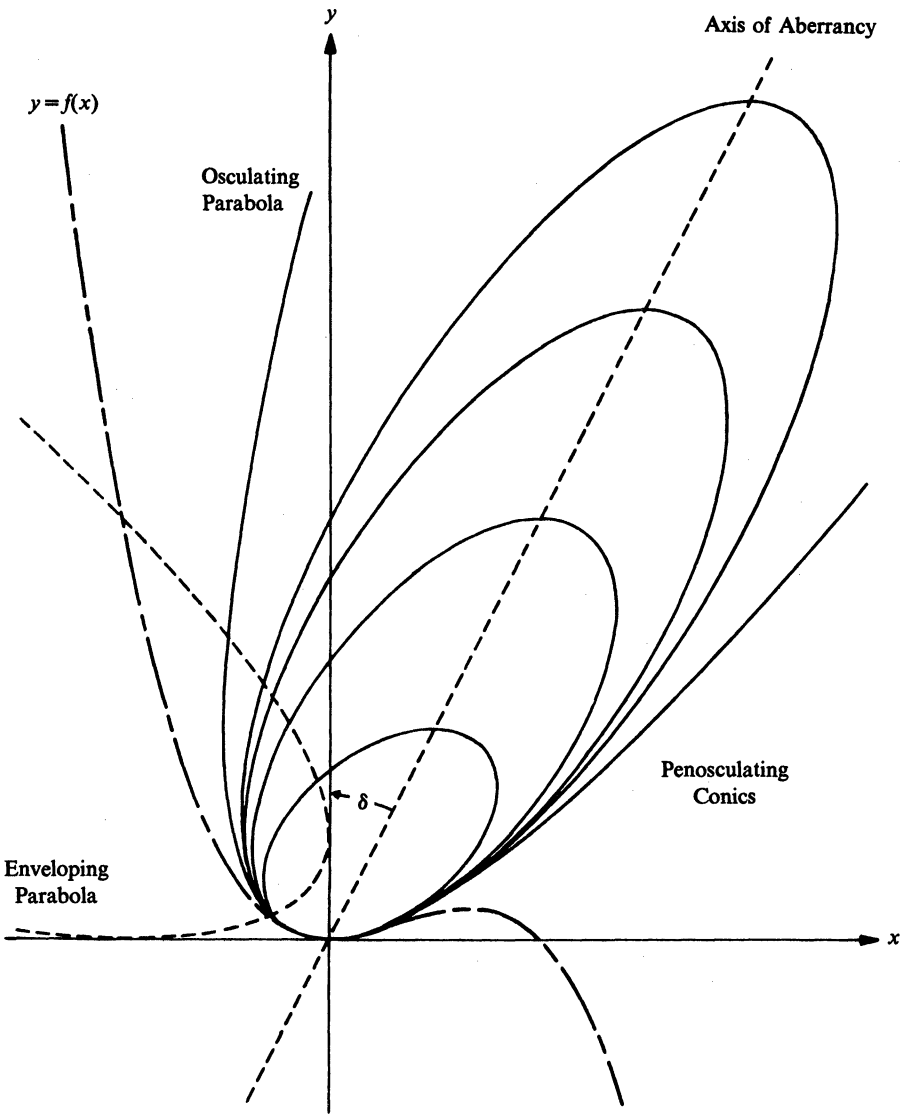
$$(y+x\tan\delta)^2-2\rho(y-x\tan\delta)+\rho^2=0. \tag{22}$$

It is a fixed parabola, called the **enveloping parabola**, whose directrix is the axis of aberrancy, $y=x\cot\delta$, and whose focus is $((-\rho/2)\sin 2\delta,\ \rho\cos^2\delta)$, namely the projection of the center of curvature onto the axis of aberrancy reflected in the normal. Knowledge of the enveloping parabola greatly facilitates the construction of the family of penosculating conics itself. To illustrate its relation to this family, the enveloping parabola and representative penosculating conics on the concave side of a given curve $y=f(x)$ are drawn in FIGURE 4.

Secondly, it may be proven that the locus of foci of the family of penosculating conics bears a remarkably simple relationship to the enveloping parabola. The locus of foci of the family of penosculating conics is in fact the pedal curve with respect to the point of contact of the enveloping parabola. In general, the **pedal curve** of a given curve with respect to a given fixed point (called the **pedal point**) is defined to be the locus of the foot of the perpendicular dropped from the fixed point to a tangent which moves along the given curve [**20**]. In the case at hand, the pedal point is the point at which the family of penosculating conics makes contact with the given curve. The locus of foci of the penosculating conics can thus be found by applying the so-called **pedal transformation** [**20**] to (22) and is

$$(2u-\rho\cos\delta\sin 2\delta)v^2-(\rho\cos\delta\cos 2\delta)uv+(2u+\rho\cos\delta\sin 2\delta)u^2=0, \tag{23}$$

where u and v are new coordinates obtained from the (x, y)−coordinates by a rotation of the coordinate axes through the angle $-\delta$ about the origin, so as to align the new v-axis with the axis of aberrancy. The cubic curve (23) is an *oblique* strophoid [**18**], with its double point at the point of contact. The locus is tangential to the normal and to the tangent of the given curve at this point and



Penosculating conics.

FIGURE 4

touches the enveloping parabola at the point

$$(x, y) = \left(-\frac{\rho}{4} \sin 2\delta, \frac{\rho}{2} \cos^2 \delta\right),$$

i.e., at the focus of the osculating parabola contained in the family (21). In the special case where the aberrancy is unity ($\delta = 45°$), the locus reduces to a *right* strophoid.

Thirdly, we note that if the aberrancy doesn't vanish, the family of penosculating conics (21) contains ellipses, hyperbolas, a parabola, and a degenerate (double tangent line) conic, but no circle. Hence it must contain a minimum eccentricity ellipse and a maximum eccentricity hyperbola. These unique members of the family can be found explicitly and it may be shown that the maximum eccentricity hyperbola is equilateral. The range of eccentricities of the family of penosculating conics (21) is thus confined to an interval $e_m \leqslant e \leqslant e_M$, where e_m is the eccentricity of the minimum eccentricity ellipse and $e_M = \sqrt{2}$ is the eccentricity of the maximum eccentricity hyperbola. For any

attainable eccentricity in this range, except at the endpoints, the family of penosculating conics can be shown to contain precisely *two* members having this eccentricity. (The unique parabola which occurs for $e=1$ may be thought of as being paired with the unique degenerate double tangent line conic which also occurs for $e=1$.) The centers of any two such conics with the same eccentricity are inverse points with respect to a circle whose diameter is the line segment connecting the centers of the minimum eccentricity ellipse and the maximum eccentricity hyperbola.

Curve of Aberrancy

In the previous sections we discussed the geometrical properties of various configurations which are determined by the derivatives at a single fixed point P of a given curve. These configurations will in general change when P moves along the given curve. For example, it is well known from calculus that as P traverses the given curve, the center of curvature O moves along a new curve, called the **evolute** of the given curve. Alternatively, the evolute may also be defined as the envelope of the normals to the given curve. In a similar way, the **curve of aberrancy** of a given curve is defined as the locus of the centers of aberrancy, or equivalently, as the envelope of the axes of aberrancy of the given curve. In affine differential geometry [**1**], [**13**], the curve of aberrancy is known as the **affine evolute** and defined as the envelope of the affine normals.

The center of aberrancy (ξ, η) corresponding to an arbitrary point (x, y) on a given curve can be written in the Cartesian form given in equation (11). This equation may also be derived directly [**5**] from the expression developed above for the center of the osculating conic to the given curve at an arbitrary point. The curve of aberrancy Ω corresponding to a curve C, given explicitly by $y=f(x)$, is now obtained by substituting $y=f(x)$ into the right-hand side of (11) and, if possible, eliminating x from the resulting expressions to obtain $\eta=\Phi(\xi)$. If x cannot be eliminated, then the expressions $\xi=\phi(x)$, $\eta=\psi(x)$ resulting from (11) serve as the parametric equations of Ω. For a curve C given parametrically by $x=f(t)$, $y=g(t)$ the expressions in (11) are easily modified [**5**].

The study of the evolute and its properties is sometimes included in calculus courses as an application of the second derivative. In a similar way, the construction and investigation of the aberrancy curve may serve as an outgrowth of the study of the third and fourth derivatives. Unfortunately, the aberrancy curve does not share some of the remarkable construction properties of the evolute. In particular, since the axes of aberrancy are in general not orthogonal to the given curve, the original curve cannot be obtained from its aberrancy curve by the well-known unwinding procedure which gives the evolute its name (evolute = *ex* + *volvere* = out + to roll). Nevertheless, as in the case of the evolute, the behavior of the aberrancy curve resulting from certain special points on the original curve can be predicted. To do this, a few preliminary definitions and results are needed.

We have defined a point of parabolic curvature as a point at which the discriminant (19) or (20) of the osculating conic, or equivalently

$$\mathcal{P}=3y''y^{(iv)}-5(y''')^2,$$

equals zero. From (12), such a point may also be defined as one at which the affine curvature vanishes. Extending this, a **sextactic point** is defined [**1**] as one at which the affine curvature (12) has a relative extremum, i.e., $[(y'')^{-2/3}]''=0$. On carrying out the differentiation, a sextactic point is thus seen to be a point at which the so-called **Monge expression**,

$$\mathcal{M}=40(y''')^3-45y''y'''y^{(iv)}+9(y'')^2y^{(v)},$$

vanishes. Such a point gets its name (*sex* = six) from the fact that the osculating conic makes at least 6-point contact (also called **hyperosculation**) with the given curve at the point.

The expressions $\mathcal{P}$ and $\mathcal{M}$ appear in the derivatives of equation (11) of the aberrancy curve

$$\frac{d\eta}{d\xi}=\frac{y'y'''-3(y'')^2}{y'''}, \qquad \frac{d^2\eta}{d\xi^2}=\frac{y''\mathcal{P}^3}{(y''')^3\mathcal{M}} \tag{24}$$

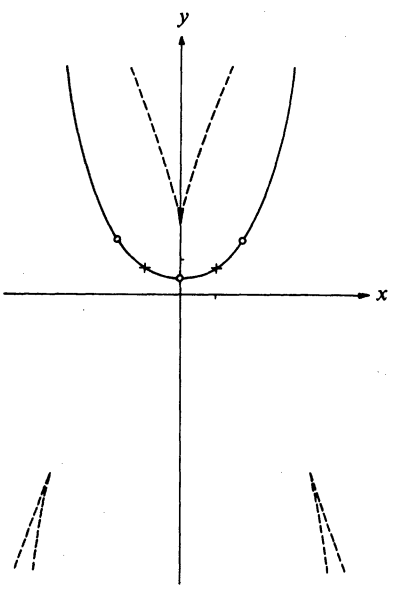
and hence in the expression for the curvature of the aberrancy curve:

$$\mathcal{R} = \frac{\left[1+\left(\frac{d\eta}{d\xi}\right)^2\right]^{3/2}}{\frac{d^2\eta}{d\xi^2}} = \frac{\left\{(y''')^2+\left[y'y'''-3(y'')^2\right]^2\right\}^{3/2}\mathcal{M}}{y''\mathcal{P}^3}. \tag{25}$$

The behavior of the aberrancy curve resulting from certain special points on the original curve can now be deduced from these formulas and is summarized in the following statements. Here the word **cusp** is used in the restricted sense of a point at which the curvature changes sign, going through zero; similarly a **point of inflection** refers to a point at which the second derivative changes sign, going through zero.

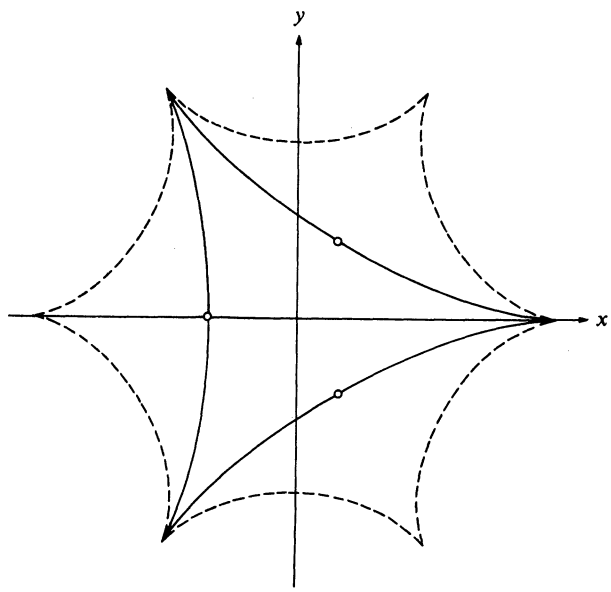
(a) A point of parabolic curvature on the original curve at which the curvature doesn't vanish produces an asymptote on the aberrancy curve (i.e., the aberrancy curve tends to infinity at the corresponding point).
(b) A point of inflection on the original curve which is not also a sextactic point and/or point of parabolic curvature produces a point of inflection on the aberrancy curve at the same point and with the same slope.
(c) A cusp on the original curve which is not also a sextactic point and/or point of parabolic curvature produces a cusp on the aberrancy curve.
(d) A sextactic point on the original curve at which $\mathcal{M}$ changes sign, going through zero, which is not also a point of parabolic and/or zero curvature produces a cusp on the aberrancy curve.

The first property follows from (11), the second property follows from (24), and the last two properties follow from (25). The behavior at higher-order singularities, such as a point which is simultaneously a sextactic point and a point of parabolic curvature will not be examined here, although this type of point does occur in Example 4 below. To illustrate these properties we construct the aberrancy curve for a number of simple plane curves; additional examples appear in [5].



Aberrancy curve for catenary.

FIGURE 5



Aberrancy curve for deltoid.

FIGURE 6

EXAMPLE 1. *Catenary* (FIGURE 5): The aberrancy curve of $y=\cosh x$ is:

$$\xi = x - \frac{3\cosh x \sinh x}{5-2\cosh^2 x}$$

$$\eta = \frac{4\cosh x\,(2+\cosh^2 x)}{5-2\cosh^2 x}.$$

The catenary has two points of parabolic curvature at $x=\pm\ln[(\sqrt{10}\pm\sqrt{6})/2]=\pm 1.0317$ which correspond to the asymptotic points on the aberrancy curve. The points of parabolic curvature divide the points of the catenary into two classes: those for which $|x|<1.0317$ are points of elliptic curvature, while those for which $|x|>1.0317$ are points of hyperbolic curvature. Three sextactic points occur on the catenary at $x=0$, $\ln(\sqrt{10}\pm 3)=\pm 1.8184$. These induce cusps on the aberrancy curve at (0, 4) and $(\pm 3.72, -10.12)$, respectively. The catenary has no points of inflection or cusps.

EXAMPLE 2. *Deltoid* (FIGURE 6): The aberrancy curve of the deltoid parametrized by

$$x = 2\cos t + \cos 2t$$
$$y = 2\sin t - \sin 2t$$

is the 6-cusped hypocycloid:

$$\xi = \frac{1}{2}(5\cos t + \cos 5t)$$

$$\eta = \frac{1}{2}(5\sin t - \sin 5t).$$

All points of the deltoid are of elliptic curvature. The deltoid has sextactic points for $t=\pi/3, \pi, 5\pi/3$ which produce cusps of the aberrancy curve at these points. The deltoid has cusps for $t=0, 2\pi/3, 4\pi/3$, which induce the other three cusps on the aberrancy curve. The deltoid has no points of inflection.

EXAMPLE 3. *Cubic polynomial*: The aberrancy curve of the cubic $y=ax^3+bx^2+cx+d$ is the cubic polynomial:

$$\eta = -\frac{125}{64}a\xi^3 - \frac{125}{64}b\xi^2 + \left(c - \frac{63b^2}{64a}\right)\xi + \left(d - \frac{7b^3}{64a^2}\right).$$

All points of the cubic polynomial are of hyperbolic curvature. Both the original and the aberrancy curve have a point of inflection at $x_0=-b/3a=\xi_0$. Both curves have the slope

$$\left(\frac{dy}{dx}\right)_0 = -\frac{b^2}{3a} + c = \left(\frac{d\eta}{d\xi}\right)_0$$

at this point. The cubic polynomial has no cusps or sextactic points.

EXAMPLE 4. *Special quartic*: The aberrancy curve of the quartic $y=x^4$ is the special quartic:

$$\eta = -1.2005\,\xi^4.$$

All points of the quartic are of hyperbolic curvature, except the origin. The origin is both a point of parabolic curvature and a sextactic point! The aberrancy curve does not tend to infinity or have a cusp due to this point (as the above results would predict if the point were either a point of parabolic curvature or sextactic, but not both). Instead it is tangent to the original curve at this point. The quartic has no cusps or points of inflection.

The behavior of the curve $y=x^4$ at the origin is similar to that of the curve $x^4+y^4=1$ at the points $(0, \pm 1)$ and $(\pm 1, 0)$. Thus we have drawn this more interesting curve as well as its aberrancy curve in FIGURE 7. In addition to the points described, the aberrancy curve also exhibits cusps induced by the sextactic points at the "corners" $(\sqrt[4]{.5}, \pm\sqrt[4]{.5})$ and $(-\sqrt[4]{.5}, \pm\sqrt[4]{.5})$.

Constant Aberrancy and Autoaberrancy

The logarithmic spiral occupies a special place in the theory of aberrancy since it is the *only* curve whose aberrancy is a non-zero constant throughout. On expressing the logarithmic spiral in polar coordinates r, ϕ as

$$r = ae^{\alpha\phi}, \qquad a, \alpha > 0 \tag{26}$$

the following aberrancy properties follow without difficulty from the previous discussion:

(a) The aberrancy at any point of the logarithmic spiral is

$$\tan \delta = \frac{\alpha}{3}.$$

(b) The radius of aberrancy at any point of the logarithmic spiral is

$$R = \rho \cos \delta,$$

i.e., the projection of the radius of curvature ρ on the axis of aberrancy.

(c) The osculating conic at any point of the logarithmic spiral is the minimum eccentricity ellipse. Moreover, the eccentricity of the osculating conic does not change as the curve is traversed and the logarithmic spiral is the only curve which has this property.

(d) All points of the logarithmic spiral are of elliptic curvature.

(e) The center of aberrancy in polar coordinates $(\bar{r}, \bar{\phi})$ of any point (r, ϕ) of the logarithmic spiral is

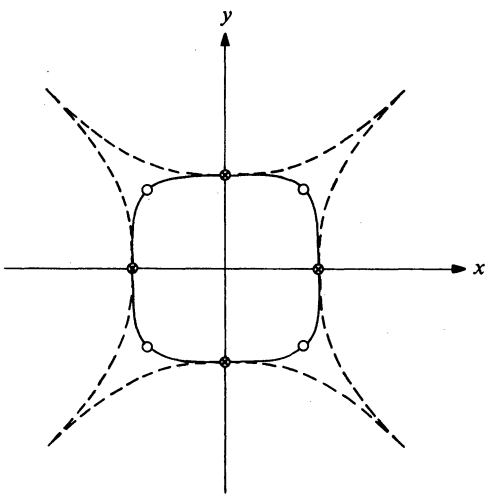
$$\bar{r} = (4 \sin \delta) r, \qquad \bar{\phi} = \phi - \delta + \frac{\pi}{2}.$$

(f) The aberrancy curve of the logarithmic spiral is

$$r = (4a \sin\delta) \exp\left[\alpha\left(\bar{\phi} + \delta - \frac{\pi}{2}\right)\right]. \tag{27}$$

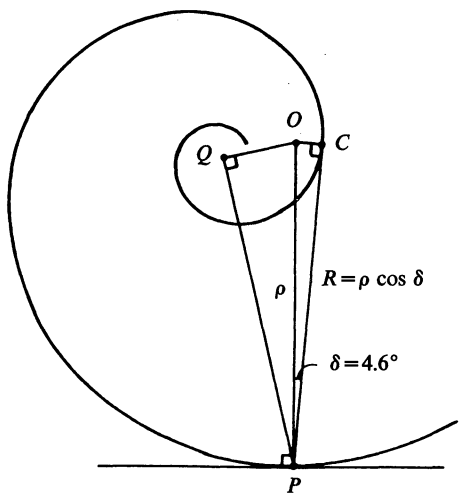
This is another logarithmic spiral *congruent* to the original spiral. The aberrancy curve is at the same time the evolutoid with angle $\delta = \arctan(\alpha/3)$. (The **evolutoid** of angle ω of a curve is the envelope of the lines making a fixed angle ω with the normal to the curve.)

For certain values of α the aberrancy curve can be made to become *identical* with the original spiral. This is analogous to the situation in which, for certain values of α, the logarithmic spiral coincides with its own evolute, i.e., the curve is an **autoevolute**. By analogy we will call a curve which



Aberrancy curve for $x^4 + y^4 = 1$.

FIGURE 7



Autoaberrancy spiral.

FIGURE 8

is identical to its aberrancy curve an **autoaberrancy** curve. We obtain the class of autoaberrancy logarithmic spirals by imposing the condition that the aberrancy curve (27) cut the polar axis at the same point as the original spiral (26), i.e.,

$$(4 \sin \delta)\exp\left[\alpha\left(\delta - \frac{\pi}{2}\right)\right] = \exp[\alpha(2n\pi)]. \tag{28}$$

This transcendental equation has solutions for $n = -1, -2, -3, \ldots$ and the corresponding values of α when substituted in (26) yield the desired class of autoaberrancy logarithmic spirals. For $n = -1$ the solution of (28) is $\alpha = .23917$ or, equivalently $\delta = 4.5581°$. The resulting autoaberrancy spiral is plotted in FIGURE 8.

Jacques Bernoulli (1654-1705) was so impressed with the self-reproducing properties of the "*spira mirabilis*" that he requested that a logarithmic spiral be engraved on his tombstone in the Cathedral of Basel, Switzerland, together with the inscription "*eadem mutata resurgo*," i.e., "though changed, I arise again the same." (It is ironic that the curve actually inscribed on the tombstone is an *Archimedean* spiral!) The above autoaberrancy property adds a new example to this apt characterization of the logarithmic spiral.

This paper is based on an invited address presented at the Va.-D.C.-Md. Section Meeting of the Mathematical Association of America on April 30, 1977, at the University of Maryland, College Park, Md.

References

[1] Wilhelm Blaschke, Vorlesungen über Differentialgeometrie, Vol. II: Affine Differentialgeometrie, 2nd ed., Berlin, Springer, 1923; reprinted N.Y., Chelsea, 1967.
[2] Carl B. Boyer, "Carnot and the Concept of Deviation," Amer. Math. Monthly, 61 (1954) 459–463.
[3] ———, A History of Mathematics, N.Y., Wiley, 1968.
[4] ———, "The Great Carnot," The Mathematics Teacher, 49 (1956) 7–14.
[5] Michael L. Brabanski, "Some Geometric Properties of the Aberrancy Curve," M.A. Thesis Project, American University, Washington, D.C., 1976.
[6] Karl Carda, "Über eine von L. N. M. Carnot berechnete Differentialinvariante," Jahresbericht der deutschen Mathematiker-Vereinigung, 28 (1919) 78–80.
[7] Lazare N. M. Carnot, Géométrie de position, Paris, 1803. (German translation: Geometrie der Stellung, 2 vols., Altona, 1810).
[8] Arthur Cayley, "Find at Any Point of a Plane Curve the Angle Between the Normal and the Line Drawn from the Point to the Centre of the Chord Parallel and Indefinitely Near to the Tangent at the Point," Messenger of Mathematics, 5 (1871) 187–190.
[9] Ernesto Cesáro, Vorlesungen über natürliche Geometrie, 2nd ed. (Translation from the Italian by Gerhard Kowalewski), Leipzig, Teubner, 1926.
[10] Joseph Edwards, An Elementary Treatise on the Differential Calculus, 3rd ed., London, Macmillan, 1900.
[11] Edouard Goursat, A Course in Mathematical Analysis, Vol. I, (Translation from the French by Earl R. Hedrick), Boston, Ginn, 1904; reprinted N.Y., Dover, 1959.
[12] Heinrich W. Guggenheimer, Differential Geometry, N.Y., McGraw-Hill, 1963; reprinted N.Y., Dover, 1977.
[13] Erich Salkowski, Affine Differentialgeometrie, Berlin, de Gruyter, 1934.
[14] George Salmon, Higher Plane Curves, 3rd ed., London, 1879; reprinted N.Y., Chelsea, 1960.
[15] Steven H. Schot, "Geometrical Properties of the Penosculating Conics of a Plane Curve," to appear in Amer. Math. Monthly.
[16] Abel Transon, "Recherches sur la courbure des lignes et des surfaces," Journal de mathématiques pures et appliquées, 6 (1841) 191–208.
[17] A. W. Walker, "The Differential Equation of a Conic and Its Relation to the Aberrancy," Amer. Math. Monthly, 59 (1952) 531–538.
[18] Heinrich Wieleitner, Spezielle ebene Kurven, Leipzig, Teubner, 1908.
[19] Ernest J. Wilczynski, "Some Remarks on the Historical Development and the Future Prospects of the Differential Geometry of Plane Curves," Bull. Amer. Math. Soc., 22 (1915-16) 317–329.
[20] Robert C. Yates, Curves and Their Properties, Washington, D.C., National Council of Teachers of Mathematics, 1952.

Unique Factorization Rings with Zero Divisors

Unique maximal ideals consisting of special elements provide structure theorems for certain common rings.

STEVEN GALOVICH
Carleton College
Northfield, MN 55057

A fundamental theme in any elementary abstract algebra course is unique factorization. Almost invariably, the setting is a commutative integral domain, R, with unity. One asks: Is it true that every nonzero nonunit of R can be expressed in an essentially unique way as a product of irreducible elements of R? In most undergraduate algebra courses and texts, this question is answered affirmatively for $R = Z$, the ring of integers, and $R = k[X]$, the ring of polynomials in one variable over a field k. Of course, millions of other examples of unique factorization domains exist; among the more famous are $Z[\sqrt{n}\,]$ for $n = -1, \pm 2, 3$, $k[X_1, \ldots, X_m]$ (the ring of polynomials in $m \geqslant 1$ variables over a field k), and $R[X]$ where R itself is a unique factorization domain.

In most, if not all algebra texts, the definition of unique factorization is stated for integral domains. However, aside from tradition, there is little reason to impose this requirement. As we shall see, the concept of unique factorization can be formulated for any commutative ring with unity. Thus, it is natural to speculate about unique factorization rings which are not integral domains, i.e., which contain zero divisors. An obvious first question is: What are some examples of such rings? A more ambitious (if somewhat foggy) problem: Can all unique factorization rings with zero divisors be classified according to some simple scheme? In this note we will investigate these questions. In view of the myriad of types of unique factorization domains, it would perhaps be surprising if the last question could be answered affirmatively. However, we shall obtain rather complete structure theorems which in essence place a given unique factorization ring with zero divisors into one of two categories and which provide isomorphism types for each category.

The bulk of the investigation is carried out in Sections 1 and 2. Section 3, while containing the final structure theorems, is largely an exposition of the algebraic results which are necessary to prove these theorems.

1. Definitions and Examples

Let R be a commutative ring with unity. An element u of R is a **unit** if there is $v \in R$ such that $uv = 1$. Let $U(R)$ denote the group of units of R. If $r, s \in R$, then r **divides** s if there exists $t \in R$ such that $rt = s$. Two elements r, s are **associates** if there is $u \in U(R)$ such that $ru = s$. An element $r \neq 0$ is a **zero divisor** if there is $s \neq 0$ in R such that $rs = 0$.

Let $r \neq 0$ be in R; r is **prime** if, whenever r divides ab where $ab \neq 0$, then r divides a or r divides b; r is **reducible** if there exist $a, b \notin U(R)$ such that $r = ab$; r is **irreducible** if r is not reducible. For future reference we observe that if r and s are irreducible, then r divides s if and only if r and s are associates. We can now state our most important definition. A commutative ring R, with unity, is a **unique factorization ring** (UFR) if for each nonzero nonunit $r \in R$,

(1) there exist irreducible elements, $r_1,\ldots,r_n$, such that $r=r_1r_2\cdots r_n$, and

(2) whenever $r=r_1\cdots r_n=s_1\cdots s_m$ where $r_1,\ldots,r_n$, $s_1,\ldots,s_m$, are irreducible, then $n=m$ and each r_i, $1\leqslant i\leqslant n$, is an associate of some s_j, $1\leqslant j\leqslant m$, and each s_k is an associate of some r_i.

Condition (1) says that any nonzero nonunit can be expressed as a product of irreducibles, while condition (2) states that except for unit factors and the order of the irreducibles, any two such factorizations are the same. Notice that no assumption is made concerning the existence of zero divisors. Also notice that we do not insist that 0 have a unique factorization into irreducibles.

Before attempting to study UFR's with zero divisors, let us look to some examples for guidance. Perhaps the most immediate example of a ring which is not an integral domain is Z/mZ where m is a nonprime integer greater than 1.

Consider $m=4$, the first example for which Z/mZ is not an integral domain. Let $\bar{r}$ denote the element $r+(m)$ of Z/mZ. Then $U(Z/4Z)=\{\bar{1},\bar{3}\}$ while the nonunits are $\bar{0}$ and $\bar{2}$. Clearly, $\bar{2}$ is irreducible in $Z/4Z$; hence $Z/4Z$ is a UFR. (Before reading on, the reader may wish to generalize this example.)

Next let $m=6$; then $U(Z/6Z)=\{\bar{1},\bar{5}\}$ and $R-U(R)=\{\bar{0},\bar{2},\bar{3},\bar{4}\}$. (In general $\bar{r}$ is a unit in Z/mZ if and only if r is relatively prime to m.) However, notice that $\bar{4}=\bar{2}\cdot\bar{2}$, $\bar{3}=\bar{3}\cdot\bar{3}$, and $\bar{2}=\bar{2}\cdot\bar{4}$. Therefore, $Z/6Z$ contains no irreducible elements, and hence fails to be a UFR! (The reader may also wish to generalize this example.)

These examples create a wealth of problems: For which m does Z/mZ have no irreducible elements? More generally, for an arbitrary m, find all irreducible elements of Z/mZ. Finally, for which m is Z/mZ a UFR? The last question is answered by Billis [2]: Z/mZ is a UFR if and only if m is a power of a prime. This result motivated the current inquiry.

2. Structure of UFR's with Zero Divisors

Our aim is to say as much as possible about the structure of a UFR R which has zero divisors. Hence for the rest of this section we assume that R is a UFR which contains zero divisors. In our investigation of the algebraic properties of R, we will concentrate on the irreducible elements of R. Because unique factorization exists in R, observations concerning the irreducible elements of R may prove useful in our analysis of R. For example, we will show in Lemma 3 that every irreducible is nilpotent. It follows that every nonunit is nilpotent. This fact is an important ingredient in the main result of this section, Theorem 7.

Our first result, although very elementary, will be used frequently in this section.

LEMMA 1. *Every irreducible element is prime.*

Proof. Let r be irreducible in R and suppose that r divides ab where a, $b\in R$ and $ab\neq 0$. Then $ry=ab$ for some $y\in R$. Factoring y, a, and b into irreducibles, we have

$$ry_i\cdots y_k=a_1\cdots a_nb_1\cdots b_m$$

where the y_i, a_j, b_h are all irreducible. Since R is a UFR, r is an associate of one of the a_j or one of the b_h; i.e., r divides a or r divides b, so r is prime.

Next, because R has zero divisors, it is an easy exercise to show that R contains an irreducible element which is a zero divisor. It is natural to ask: Is every irreducible a zero divisor? A quick answer is provided by the following lemma.

LEMMA 2. *Every irreducible in R is a zero divisor.*

Proof. Suppose s is an irreducible which is not a zero divisor. Let r be an irreducible which is a zero divisor, so that $rx=0$ for some $0\neq x\in R$.

Let $z=r+s$. Since s does not divide r (otherwise s would be a zero divisor), s does not divide z. Now

$$zx=rx+sx=sx\neq 0. \tag{1}$$

Factoring x and z into irreducibles and substituting into (1), we have

$$z_1 \cdots z_m \cdot x_1 \cdots x_n = s x_1 \cdots x_n.$$

By unique factorization, $m=1$ and s divides $z_1 = z$. This contradiction implies that no such element s exists in R.

Our third lemma is suggested by a property of the UFR Z/p^nZ, $n>1$. In this ring every nonunit is divisible by $\bar{p}$ and hence has the form $\bar{p}x$ for some $x \in Z/p^nZ$. Note that $(\bar{p}x)^n = 0$. In an arbitrary ring an element r such that $r^n = 0$ for some integer $n>1$ is called **nilpotent**. Thus, in Z/p^nZ, every nonunit is nilpotent. The following lemma and its first corollary generalize this observation to all UFR's with zero divisors.

LEMMA 3. *Every irreducible element of R is nilpotent.*

Proof. By Lemma 2, we know that $rx=0$ for some $x \neq 0$. Letting $r=r_1$ and factoring x into irreducibles, $x = r_2 \cdots r_n$, we have $r_1 r_2 \cdots r_n = 0$. We can rewrite this relation in the form $r_1^{a_1} \cdot r_2^{a_2} \cdots r_m^{a_m} = 0$ where $r=r_1$, r_i and r_j are nonassociate irreducibles if $i \neq j$, and the a_i are positive integers.

If $m=1$, then r is nilpotent. If $m>1$, then let $z = r_1^{a_1} + (r_2^{a_2} \cdots r_m^{a_m})$. Note that since $x \neq 0$, $r_2^{a_2} \cdots r_m^{a_m} \neq 0$. If r_1 divides z, then r_1 divides $r_2^{a_2} \cdots r_m^{a_m}$ and hence, by Lemma 1, r_1 divides r_i for some $i>1$. Thus r_1 does not divide z. But $r_1^{a_1} z = r_1^{2a_1} + r_1^{a_1} r_2^{a_2} \cdots r_m^{a_m} = r_1^{2a_1}$ which violates unique factorization unless $r_1^{2a_1} = 0$.

COROLLARY 4. *Every nonunit in R is nilpotent.*

COROLLARY 5. *Z/mZ is a UFR if and only if m is a power of a prime.*

Proof. Suppose m is not a power of a prime. Let p be a prime dividing m. Then $\bar{p}$ is not a nilpotent element of Z/mZ, for if so, then $\bar{p}^k = 0$ for some $k>1$, and therefore, m divides p^k which is impossible. Hence Z/mZ is not a UFR. We leave the converse as an exercise for the reader.

The result of Corollary 4 suggests that we consider the set of nonunits more closely. The following proposition indicates the importance of this set.

PROPOSITION 6. *Let M be the set of nonunits of R. Then M is the unique maximal ideal of the ring R.*

Proof. Let $x, y \in M$. By Corollary 4, there exist integers m, n such that $x^m = 0 = y^n$. By the binomial theorem $(x+y)^{m+n} = 0$. Moreover, $(rx)^m = 0$ for all $r \in R$. Therefore, $x+y$ and rx are elements of M, and M is an ideal of R. Finally, it is easy to check that M is a maximal ideal and that any ideal of R is contained in M.

A ring with exactly one maximal ideal is called a **local ring** (or, by some authors, a quasi-local ring). So every UFR with zero divisors is a local ring. Another well-known example of a local ring is the ring Z_p of all rational numbers which can be represented in the form r/s where r and s are relatively prime integers and s is not divisible by the prime p. The maximal ideal of Z_p consists of all elements r/s which are in lowest terms and for which p divides r. Local rings have played a major role in the modern development of algebra, algebraic number theory, and algebraic geometry (e.g., see [3] or [5]). Therefore, it is not surprising that local rings have been studied extensively. In the next section we shall apply some important results concerning local rings to our present situation.

For now let us record some further observations on UFR's with zero divisors. Again, we use the ring Z/p^nZ, $n>1$, as a model for our investigation. It is easy to see that any irreducible element in Z/p^nZ is an associate of $\bar{p}$. Does an analogous statement hold for every UFR with zero divisors? If R is a UFR with zero divisors, then is it true that every irreducible is an associate of some fixed irreducible? Equivalently, we can ask: If R is a UFR with zero divisors, then is the maximal ideal M of R a principal ideal, i.e., does $M=(r)$ for some $r \in R$? We offer a succinct answer to these questions: Almost.

Let us suppose that there are irreducibles r, s of R which are not associates. What conclusions about R can we draw? First choose the least integers m, n such that $r^m = 0 = s^n$. Now suppose $rs \neq 0$.

Then $0 \neq rs = r(r^{m-1}+s)$. By Lemma 1, we conclude that s divides $r^{m-1}+s$ and thus s divides r which is a contradiction. Therefore, if r and s are nonassociate irreducibles, then $rs=0$.

With the same assumptions about r and s, suppose that $m<n$. Then, since $rs=0$, $(r+s)^m = r^m + s^m = s^m \neq 0$. By Lemma 1 s divides $r+s$ and hence s divides r. If $n<m$, then a similar contradiction is reached. If $m=n>2$, then $r(r+s)=r^2 \neq 0$ which again leads to the absurd conclusion that r and s are associates. We collect our results in the following theorem.

THEOREM 7. *Any UFR with zero divisors is a local ring whose maximal ideal M is the set of all nonunits. In such a case, either M is principal, or $rs=0$ for all irreducibles r and s (not necessarily distinct).*

Note that if M is not principal, $xy=0$ for all $x, y \in M$. In a sense this can be regarded as the "trivial" case where the product of any two nonunits is zero and (hence) every nonzero nonunit is irreducible. Clearly under such conditions unique factorization must exist. How could it fail?

Observe that all the possibilities in Theorem 7 can be realized: For a prime p and for $n \geqslant 3$, Z/p^nZ has a principal maximal ideal with non-trivial products; $k[X, Y]/(X^2, XY, Y^2)$ where k is a field has a non-principal maximal ideal with trivial products; and Z/p^2Z is an example with both a principal maximal ideal and trivial products.

3. Final Structure Theorems

Although Theorem 7 gives us a comprehensive description of UFR's with zero divisors, it is an internal description. In this section we provide an external description (Theorems 8 and 9) which uses rather more advanced algebraic machinery. Although we do introduce all the conceptual framework for stating these results, the development of the theory of complete local rings is drawn from such standard references as [4], [9], [10]. Essentially, our structure theorems assert that any UFR with zero divisors is isomorphic to the ring $A[S]/J$ where A is either a field or a complete local domain whose maximal ideal is generated by a prime, S is a set of indeterminates which is determined by M (e.g., if M is principal, then S contains exactly one element), $A[S]$ is the ring of polynomials whose coefficients are in A and whose variables are in S, and J is a certain ideal in the ring $A[S]$. This characterization follows directly from some important results on local rings which are due to I. S. Cohen. In order to state Cohen's theorems, we pause to introduce several new concepts and to become acquainted with some interesting rings.

Let R be an arbitrary local ring with maximal ideal M. Let M^n denote the ideal consisting of all finite sums of n-fold products of elements of M: $M^n = \{\sum_{i=1}^{k} x_i | x_i = r_{i_1} \cdots r_{i_n}$ where $r_{i_j} \in M\}$. A sequence of elements of R, $\{a_j | j \geqslant 1\}$, is called a **Cauchy sequence** if for each integer $m \geqslant 1$, there is an integer m_0 such that $a_k - a_h \in M^m$ for $k, h \geqslant m_0$. It is perhaps strange that the term "Cauchy sequence" makes an appearance in this algebraic setting. Usually the term refers to a sequence of elements in a metric space. In order to justify the present use of this terminology, we indicate briefly how metric and topological notions arise naturally in the study of local rings.

For an element x of the local ring R, let $v(x)$ be the largest integer n such that $x \in M^n$, if such an integer exists; if $x \in \cap_{n=0}^{\infty} M^n$, then let $v(x) = +\infty$. The function v is called the **order function** on R. We record the following properties of v ([10], p. 249):

(1) $v(xy) \geqslant v(x) + v(y)$ for all $x, y \in R$,

(2) $v(x+y) \geqslant \min\{v(x), v(y)\}$ for all $x, y \in R$.

Let c be a fixed real number greater than 1. For $x, y \in R$, define the **distance** between x and y, $d(x, y)$, to be $d(x, y) = c^{-v(x-y)}$. Thus two elements $x, y \in R$ are "close" if $x-y$ lies in M^n for large n. We observe that

(i) $d(x, y) = d(y, x)$ for all $x, y \in R$,

(ii) $d(x, z) \leqslant \max\{d(x, y), d(y, z)\}$ for all $x, y, z \in R$ (strong triangle inequality),

(iii) $d(x, y) \geqslant 0$ and $d(x, y) = 0$ if and only if $x - y \in \cap_{n=1}^{\infty} M^n$.

The truth of (ii) follows from property (2) of v. We note that d is not necessarily a metric on R. In

fact, d is a metric on R if and only if $d(x, y)=0$ implies $x=y$, or equivalently, if and only if $\cap_{n=1}^{\infty} M^n=(0)$. Henceforth, we only consider local rings for which $\cap_{n=1}^{\infty} M^n=(0)$, i.e., local rings in which d is a metric. In such a local ring the sets $\{x+M^n | x \in R, n \geqslant 1\}$ form an open base for the topology induced by the metric d (see [**10**, pp. 251–54]).

It is easily checked that $\{a_n | n \geqslant 1\}$ is a Cauchy sequence in R if and only if $d(a_n, a_m) \to 0$ as $n, m \to \infty$. In other words, $\{a_n\}$ is a Cauchy sequence as defined here if and only if $\{a_n\}$ is a Cauchy sequence in the metric space R with metric d.

A Cauchy sequence $\{a_n\}$ in a local ring R **converges** to a in R if $d(a_n, a) \to 0$ as $n \to \infty$. (Note that $d(a_n, a) \to 0$ as $n \to \infty$ if and only if for each positive integer m, there is a positive integer m_0 such that $a_n - a \in M^m$ for $n \geqslant m_0$.) A local ring R is **complete** if every Cauchy sequence of R converges to an element of R. Thus to say that R is a complete local ring is merely to say that when regarded as a metric space, R is a complete metric space. We now list a few important complete local rings. These examples play a major role in Theorems 8 and 9.

EXAMPLE 1. For a prime p and an integer $n \geqslant 1$, Z/p^nZ is a complete local ring. In fact, any Cauchy sequence is eventually constant: If $\{a_n\}$ is a Cauchy sequence in Z/p^nZ, then there is an integer m_0 such that $a_k - a_h \in (\bar{p})^n=(0)$ for $k, h \geqslant m_0$. (Recall that $\bar{r}$ is the element $r+(p^n)$ of Z/p^nZ.) Thus, for $k, h \geqslant m_0$, $a_k = a_h = a$ and $\{a_n\}$ converges to a.

EXAMPLE 2. In general, if R is a local ring for which $M^n=(0)$ for some $n \geqslant 2$, then R is a complete local ring. As in the case of Z/p^nZ, any Cauchy sequence in R is eventually constant and hence is convergent.

Now Theorem 7 can be interpreted as saying that if R is a UFR with zero divisors, then $M^n=(0)$ for some $n \geqslant 2$. (If R has a principal maximal ideal, n can be chosen to be the least integer k such that $r^k=0$ for any irreducible r; otherwise we can take $n=2$.) *Therefore, any UFR with zero divisors is a complete local ring*, and, of course, any structure theorem concerning complete local rings applies to UFR's with zero divisors.

EXAMPLE 3. Let k be a field and let $R=k[[X]]$ be the ring of formal power series in one variable over k. Specifically, $R=\{\sum_{m=0}^{\infty} a_m X^m | a_m \in k\}$ where addition and multiplication are defined as follows:

$$\text{(i)} \qquad \left(\sum_{m=0}^{\infty} a_m X^m\right)+\left(\sum_{m=0}^{\infty} b_m X^m\right)=\sum_{m=0}^{\infty}(a_m+b_m)X^m$$

$$\text{(ii)} \qquad \left(\sum_{m=0}^{\infty} a_m X^m\right)\left(\sum_{m=0}^{\infty} b_m X^m\right)=\sum_{m=0}^{\infty} c_m X^m \quad \text{where} \quad c_m=\sum_{i=0}^{m} a_i b_{m-i}.$$

Note that R is an integral domain. Using the method of undetermined coefficients, one can show that any power series $f=\sum_{m=0}^{\infty} a_m X^m$ with $a_0 \neq 0$ has an inverse, i.e., there is $g \in R$ such that $fg=1$. It follows that R is a local ring with maximal ideal consisting of all multiples of X. Two elements f, $g \in R$ are "close" with respect to the metric d if and only if the power series $f-g$ is divisible by a high power of X. An example of a Cauchy sequence in R is $\{\sum_{i=0}^{n} X^i | n \geqslant 0\}$; observe that this sequence converges to $1+X+X^2+\cdots+X^n+\cdots$. More generally, the reader is invited to check that R is a complete local ring.

In general, the ring of formal power series in n variables over k (where addition and multiplication are defined in the analogous ways) is a complete local domain. (See [**10**, Ch. 8, §1].)

EXAMPLE 4. Let $S=k[[X]]$ and let $R=S/I$ where I is the ideal consisting of all multiples of X^n. Then as in Example 2, S is a complete ring. For future reference we observe that S is isomorphic to the ring $k[X]/(X^n)$.

EXAMPLE 5. Let p be a prime number. Let S_p be the collection of all sequences of integers $\{x_n | n \geqslant 0\}$ such that $x_n = x_{n-1} (\mathrm{mod}\ p^n)$ for all $n \geqslant 1$. Two such sequences $X = \{x_n\}$, $X' = \{x_n'\}$ are **equivalent** (written $X \sim X'$) if $x_n \equiv x_n' (\mathrm{mod}\ p^{n+1})$ for $n \geqslant 0$. Clearly, $\sim$ is an equivalence relation on S_p. The set of equivalence classes, denoted by $\hat{Z}_p$, is called the set of ***p*-adic integers**. If $X = \{x_n\}$ and $Y = \{y_n\}$ are in S_p, then we define $X + Y = \{x_n + y_n\}$ and $XY = \{x_n y_n\}$. One can check that $X + Y$ and XY are elements of S_p, and that if $X \sim X'$ and $Y \sim Y'$, then $X + Y \sim X' + Y'$ and $XY \sim X'Y'$. Under the operations induced on equivalence classes, $\hat{Z}_p$ is a commutative integral domain. Moreover, the ring Z of ordinary integers can be embedded in $\hat{Z}_p$ by sending $m \in Z$ to the equivalence class of $\{x_n | x_n = m\}$ which class will also be denoted by m. One can show that a p-adic integer $X = \{x_n\}$ is a unit in $\hat{Z}_p$ if and only if $x_0 \not\equiv 0 (\mathrm{mod}\ p)$. (See [**1**, p. 274] or [**3**, p. 22].) It follows that $X = \{x_n\}$ is a nonunit if and only if $x_0 \equiv 0 (\mathrm{mod}\ p)$; in view of the condition $x_n \equiv x_{n-1} (\mathrm{mod}\ p^n)$, $x_n \equiv 0 (\mathrm{mod}\ p)$ for all n. Thus any nonunit of $\hat{Z}_p$ is a multiple of the p-adic integer $p = \{x_n | x_n = p\}$. Therefore, $\hat{Z}_p$ is a local ring whose maximal ideal is generated by the element p. Let X be a p-adic integer. It is easy to see that X can be represented by a sequence of the form $\{a_0, a_0 + a_1 p, a_0 + a_1 p + a_2 p^2, \ldots, a_0 + a_1 p + \cdots + a_n p^n, \ldots\}$ where $0 \leqslant a_i \leqslant p - 1$. Thus, we can think of X as an infinite series $\Sigma_{n=0}^{\infty} a_n p^n$. (We hasten to add that $\hat{Z}_p$ is not isomorphic to the ring of formal power series $k[[X]]$ where $k = Z/pZ$ (reason?).) Finally, one can also show that $\hat{Z}_p$ contains a subring isomorphic to Z_p. In fact, Z_p (as well as Z) is a dense subring of $\hat{Z}_p$ with respect to the metric topology on $\hat{Z}_p$. The restriction of the metric to Z_p (or Z) endows Z_p (or Z) with a metric space structure. One can show that $\hat{Z}_p$ is the completion of Z_p (or Z) as a metric space. Thus $\hat{Z}_p$ is a complete local ring. Full descriptions of the ring $\hat{Z}_p$ are found in the excellent books by Agnew [**1**, Unit 8] and Borevich-Shafarevich [**3**, pp. 18–40].

We now turn to the theorems of I. S. Cohen on the structure of complete local rings. We remark that in his original work [**4**], Cohen analyzes complete local rings in which the maximal ideal has a finite set of generators: There exist $x_1, \ldots, x_n \in M$ such that any $x \in M$ can be written as $x = r_1 x_1 + \cdots + r_n x_n$ for suitable $r_1, \ldots, r_n \in R$. However, Cohen's theorems have been extended by Nagata to complete local rings without finitely generated maximal ideals. Aside from Cohen's paper, the reader will find it useful to consult [**10**, Ch. VIII], and [**9**, section 31].

To state Cohen's First Theorem, we require two additional concepts. For any local ring R with maximal ideal M, the quotient field $P = R/M$ is called the **residue field** of R. Letting $\mathrm{char}(A)$ denote the characteristic of the ring A, we have four possible cases.

Case 1. $\mathrm{char}(R) = 0 = \mathrm{char}(P)$. For example, this possibility occurs if $R = k[[X]]$ where k is a field of characteristic 0.

Case 2. $\mathrm{char}(R) = p = \mathrm{char}(P)$. Example: $k[[X]]$ where k is a field of characteristic p.

Case 3. $\mathrm{char}(R) = 0$ and $\mathrm{char}(P) = p$. Example: $R = Z_p$ or $R = \hat{Z}_p$, for $R/M \cong \hat{Z}_p/(p\hat{Z}_p) \cong Z_p/(pZ_p) \cong Z/pZ$.

Case 4. $\mathrm{char}(R) = p^e$, $e > 1$, and $\mathrm{char}(P) = p$. Example: $R = Z/p^e Z$, $e > 1$.

For Cohen's First Theorem, we assume that either Case 1 or Case 2 holds, the so-called **equicharacteristic** cases. We claim that in both cases R contains a field. If $\mathrm{char}(R) = 0 = \mathrm{char}(P)$, then for each positive integer, n the n-fold sum $1 + \cdots + 1$ is not in M. Thus, every integer is a unit of R, and hence R contains a field isomorphic to the field of rational numbers. If $\mathrm{char}(R) = p = \mathrm{char}(P)$, then R clearly contains a field isomorphic to Z/pZ.

Suppose now that R contains a field K. Then under the projection $R \to R/M = P$, K is mapped isomorphically onto a subfield of P. If K is mapped onto the entire field P under this projection, then K is called a **coefficient field** or **field of representatives** in R. Using this vocabulary, we can now state Cohen's First Theorem: *If R is a complete local ring such that $\mathrm{char}(R) = \mathrm{char}(P)$, then R contains a coefficient field. Moreover, suppose the maximal ideal M of R has a minimum generating set B. Let S be a set of indeterminates which is in one-to-one correspondence with B. Then R is a homomorphic image of the ring $P[[S]]$ of formal power series with coefficients from P and variables from S.* For a proof, see [**4**, pp. 72–77], [**9**, pp. 106–09], or [**10**, pp. 279–80, pp. 304–06]. We remark that under the homomorphism from $P[[S]]$ onto R, S is mapped bijectively onto B.

THEOREM 8. *Let R be a UFR with zero divisors. Suppose that char(R) = char(P).*

(1) *If M is principal, then $R \cong P[X]/(X^m)$.*

(2) *If M is not principal, but has a minimum generating set B, then $R \cong P[S]/I$ where S is a set of indeterminates which is in one-to-one correspondence with B, and I is the ideal of $P[S]$ generated by all products xy where $x, y \in S$.*

Proof. By Cohen's Theorem, R is a homomorphic image of $P[[X]]$ with S being mapped bijectively onto B. In Case (1), S contains only one element, $S = \{X\}$. If m is the least integer such that $r^m = 0$ for any irreducible r of R, then

$$R \cong P[[X]]/(X^m) \cong P[X]/(X^m).$$

The proof of (2) is equally immediate.

Next we turn to cases (3) and (4), the nonequicharacteristic cases. We claim that if R is a UFR with zero divisors, then case (3), char$(R) = 0$ and char$(P) = p$, cannot occur. Suppose char$(R) = 0$. If $n \in M$ where n is a positive integer, then, by Corollary 4, $n^k = 0$ for some integer k. Thus char(R) divides n^k which is a contradiction. As a result, $n \in U(R)$, and $U(R)$ contains a field isomorphic to the field of rational numbers. The image of this field in P is a field of characteristic 0, and therefore char$(P) = 0$.

Finally, we consider case (4): R is a complete local ring such that char$(R) = p^e$, $e > 1$, and char$(P) = p$. A complete local domain whose maximal ideal is generated by a prime number p is called a **v-ring**. For example, $\hat{Z}_p$ is a v-ring. Using this definition we can state Cohen's Second Theorem: *Let R be a complete local ring with residue field P of characteristic p. Suppose that the maximal ideal M has a minimum generating set B. Let S be a set which is in one-to-one correspondence with B. Then R is a homomorphic image of the power series ring $V[[S]]$ where V is a v-ring whose residue field is P. If $p \notin M^2$, then S may be replaced by the set $T = S - \{x\}$ for some element $x \in S$.* For a proof, see [**4**, Theorem 12] or [**9**, pp. 106–09].

THEOREM 9. *Let R be a UFR with zero divisors and maximal ideal M. Suppose that char$(R) = p^e$, $e > 1$.*

(1) *If M is principal, then $R \cong V[X]/(f(X), X^n)$ where V is a v-ring $f(X)$ is an irreducible polynomial in $V[X]$ of degree k where $(pR) = M^k$, and n is the least integer such that $M^n = (0)$.*

(2) *If M is not principal and has a minimum generating set B, then char$(R) = p^2$ and $R \cong V[S]/I^2$ where S is a set of indeterminates in one-to-one correspondence with $B - \{x\}$ for some $x \in B$ and I is the ideal in $V[X]$ generated by $\{p\} \cup S$.*

Proof. The first conclusion is a consequence of Cohen's Second Theorem. A complete argument can be found in [**8**, Theorem 3.2]. Suppose then that M is not principal. Since char$(R) = p^e$, $e > 1$, p is a nonzero nonunit. By Theorem 7, p is irreducible and $p^2 = 0$. Therefore, char$(R) = p^2$. By Cohen's Second Theorem R is a homomorphic image of $V[[T]]$ where V is a v-ring and T is a set in one-to-one correspondence with $B - \{x\}$ for some $x \in B$. Under this homomorphism p in V is sent to p in R and the subset of R consisting of p and the image of T generates M. The theorem now follows.

Note that in case (1), if p does not divide k, the degree of f, then $f(X)$ has the form $f(X) = X^p - pu$ where $u \in U(V)$. (See [**8**, Theorem 3.5].) In general, $f(X)$ is an Eisenstein polynomial in $V[X]$: $f(X) = X^k + a_1 X^{k-1} + \cdots + a_k$ where $a_i \in (pV)$ for all i and $a_k \notin (p^2 V)$. If p does divide k, then apparently no explicit formulas for $f(X)$ are known.

4. Concluding Remarks

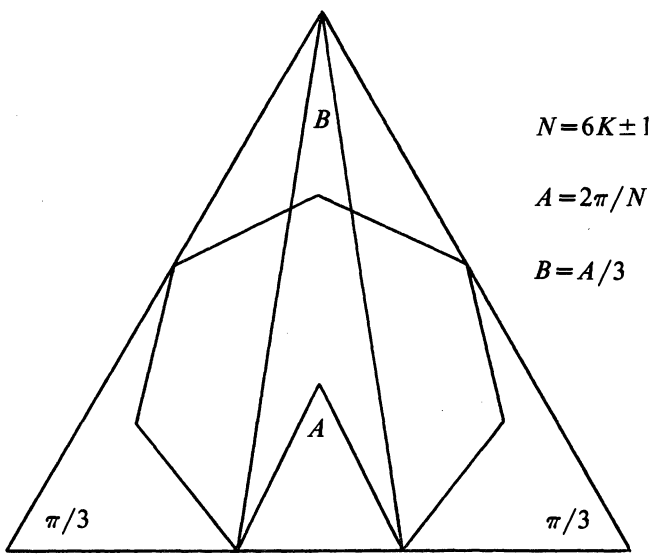
In this note we have studied a natural generalization of the notion of unique factorization domain. As a result of this investigation, we have obtained an almost complete description of unique factorization rings with zero divisors. Different extensions of the unique factorization concept are possible. In fact (as I learned after submitting this note), one such generalization was made several

years ago by Fletcher ([6], [7]). Without giving details, suffice it to say that Fletcher's definitions of irreducible element and unique factorization ring are different from those used in this paper. Not surprisingly, his theorems are also different. Roughly, his definitions are less restrictive than ours; however, he does assume that 0 has a unique decomposition. The upshot is that a UFR in our sense whose maximal ideal is (resp., is not) principal is (resp., is not) a UFR in Fletcher's sense. On the other hand, rings such as Z/nZ (or any principal ideal ring) are UFR's according to Fletcher's definition. The reader may wish to propose and to investigate other generalizations of the idea of unique factorization domain.

References

[1] J. Agnew, Explorations in Number Theory, Brooks-Cole, Belmont, 1972.
[2] M. Billis, Unique factorization in the integers modulo m, Amer. Math. Monthly, 75 (1968) 527.
[3] Z. I. Borevich and I. R. Shafarevich, Number Theory, Academic Press, New York, 1966.
[4] I. S. Cohen, On the structure and ideal theory of complete local rings, Trans. Amer. Math. Soc., 59 (1946) 54–106.
[5] J. D. Emerson, Simple points of an affine algebraic variety, Amer. Math. Monthly, 82 (1975) 132–47.
[6] C. R. Fletcher, Unique factorization rings, Proc. Cambridge Philos. Soc., 65 (1969) 579–583.
[7] ——, The structure of unique factorization rings, Proc. Cambridge Philos. Soc., 67 (1970) 535–540.
[8] K. R. McLean, Commutative Artinian principal ideal rings, Proc. London Math. Soc. (3) 26 (1973) 249–72.
[9] M. Nagata, Local Rings, Wiley, New York, 1962.
[10] O. Zariski and P. Samuel, Commutative Algebra, Vol. II, Van Nostrand, Princeton, 1960.

Proof Without Words: Trisecting the Angles of a Regular N-gon



MICHAEL GOLDBERG
5823 Potomac Ave., NW
Washington, D.C. 20016

NOTES

Adding Totitives

MIRIAM HAUSMAN
Baruch College (CUNY)
New York, NY 10010

HAROLD N. SHAPIRO
New York University
New York, NY 10012

The Euler function $\phi(n)$ is defined as equal to the number of positive integers less than n which are relatively prime to n. Each of these $\phi(n)$ integers is called a **totitive** of n. Santos [1] considered the problem of finding the largest integer with the property that, when added to each of its totitives, yields a prime number. He proved that a largest such integer does exist, but only conjectured that 12 was actually this largest integer.

In this note the more general question is considered of finding the largest integer n such that for some value of k, (k depending on n), kn plus each totitive of n yields a prime. It is shown that for $k=1$ the Santos conjecture $n=12$ is correct, and that for the more general problem described above, $n=18$ (with $k=892$) is the largest such integer.

The corresponding questions as to whether for some k there is a largest integer n with the property that each of the prime (or alternatively composite) totitives of n added to kn always results in a prime (or always results in a composite) are also answered. Although we do not find specific values for the largest such integers, we do decide in each case whether or not such largest integers actually exist.

A central point of the method which is used here is to focus on the quantity p^* (depending on n), which is the smallest prime which does not divide n. We will require an upper bound for p^*. In this regard, it is amusing that such a bound is an immediate consequence of an old result [2] which provides that 30 is the largest integer, all of whose totitives (other than 1) are primes. Then if $n>30$, there is a totitive $s>1$, of n, which is composite. Thus s has a prime divisor less than or equal $\sqrt{s} \leqslant \sqrt{n}$ which does not divide n. Hence we conclude that *for* $n>30$, $p^*<\sqrt{n}$.

Our first objective is the proof of the following:

THEOREM 1. *Only for* $n=1,2,4,6,10,12,18$ *does there exist a value of* k *for which* kn *plus each totitive of* n *is prime. For* $n=1,2,4,6,10,12$, *the smallest such* k *is* $k=1$, *and for* $n=18$, *the smallest such* k *is* $k=892$.

The idea of the proof of this theorem is to show that for each residue class, $i^* \pmod{p^*}$, $0 \leqslant i^* < p^*$, there exists a totitive a of n such that $a \equiv i^* \pmod{p^*}$. Once this is established, it follows that for any positive integer k, kn plus the totitives of n cannot all be prime. For, given k, we have a totitive $a \equiv -kn \pmod{p^*}$, so that p^* divides $kn+a$.

The above proposal is carried out by direct counting. This in turn is facilitated by the Möbius function $\mu(n)$, ($\mu(n)=0$ if n has a square factor bigger than 1, $\mu(1)=1$, and for n square free $\mu(n)=(-1)^t$ where $t=$the number of distinct prime factors of n). More specifically the critical property is that $\Sigma_{d|n}\mu(d)$ equals 1 if $n=1$ and 0 otherwise. Thus if (a,n) denotes the greatest common divisor of n and a,

$$\sum_{d|(a,n)} \mu(d) = \begin{cases} 1 & \text{if } (a,n)=1 \\ 0 & \text{if } (a,n)>1. \end{cases} \tag{1}$$

Proof of Theorem 1. For a fixed i^*, $0 \leqslant i^* < p^*$, let $N(i^*)$ be the number of positive integers $a \leqslant n$ such that $(a,n)=1$ and $a \equiv i^* \pmod{p^*}$. We wish to find those integers n for which $N(i^*)>0$. Using (1) we have

$$N(i^*) = \sum_{\substack{a \leqslant n \\ a \equiv i^* \pmod{p^*}}} \sum_{d|(a,n)} \mu(d) = \sum_{d|n} \mu(d) \sum_{\substack{a \leqslant n \\ a \equiv i^* \pmod{p^*} \\ a \equiv 0 \pmod{d}}} \tag{2}$$

Let $0 \leqslant a(d) < p^*d$ be the unique solution to the congruences $a \equiv i^* \pmod{p^*}$, $a \equiv 0 \pmod{d}$. Since $(d,p^*)=1$, $a(d)$ exists by the Chinese Remainder Theorem; and the inner sum on the right of (2) is the number of $t \geqslant 0$ such that $a(d)+tp^*d \leqslant n$. Thus we have

$$\frac{n}{p^*d} - 1 < \sum_{\substack{a \leqslant n \\ a \equiv i^* \pmod{p^*} \\ a \equiv 0 \pmod{d}}} 1 < \frac{n}{p^*d} + 1. \tag{3}$$

Utilizing (3) in (2) gives

$$N(i^*) > \sum_{d|n} \mu(d) \frac{n}{p^*d} - \sum_{d|n} |\mu(d)|$$

or

$$N(i^*) > \frac{n}{p^*} \cdot \frac{\phi(n)}{n} - 2^r \tag{4}$$

where r is the number of distinct prime factors of n. Thus from (4) we see that our desired inequality $N(i^*)>0$ is satisfied if

$$\phi(n) \geqslant 2^r p^*. \tag{5}$$

For $n = \prod_{i=1}^{r} p_i^{\alpha_i}$,

$$\phi(n) = \prod_{i=1}^{r} p_i^{\alpha_i - 1}(p_i - 1) \geqslant \prod_{i=1}^{r} (p_i - 1)$$

so that (5) is implied by

$$(p_1-1)(p_2-1)\cdots(p_r-1) \geqslant 2^r p^*. \tag{6}$$

Thus it suffices to seek the largest integer which fails to satisfy (6). This is achieved by substituting specific values for p^* and analyzing in each case the values of n which fail to satisfy either (5) or (6). In the case of specific integers n, elimination can also be achieved by noting directly that the totitives of n include a complete residue system modulo p^*.

Noting that the left side of (6) is minimized by taking the p_i as small as possible, we proceed as follows:

Case 1: $p^*=2$. As 2 does not divide n, only $n=3$ fails to satisfy (5). But the totitives of 3 are 1 and 2, and the integers $3k+1$ and $3k+2$ cannot both be prime, as one of them must be even.

Case 2: $p^*=3$. Since 2 does and 3 doesn't divide n, $(2-1)(p-1) \geqslant 2^2 \cdot 3$ for $p \geqslant 13$, and $(2-1)(5-1)(7-1) \geqslant 2^3 \cdot 3$, the values of n which do not satisfy (6) are of the form $n = 2^\alpha 5^\beta$, $2^\alpha 7^\beta$, $2^\alpha 11^\beta$, $\alpha > 0$. As (5) is satisfied by $n = 2^\alpha 5^\beta$ if $\alpha > 3$ or $\beta > 1$, and by $n = 2^\alpha 7^\beta$ and $2^\alpha 11^\beta$ if $\alpha = 1$ and $\beta > 1$, we need only consider $n = 2, 4, 8, 2\cdot 5, 2^2\cdot 5, 2^3\cdot 5, 2\cdot 7, 2\cdot 11$. For $n = 2^3 \cdot 5$, (5) is satisfied. The totitives of $n = 2^2 \cdot 5 = 20$ are $1, 3, 7, 9, 11, 13, 17, 19$ and include a complete residue system modulo 3, thereby eliminating this value of n. The same argument eliminates $n = 8$, 14, and 22. The remaining integers, $n = 2, 4, 10$ all have the property that when added to their totitives they yield only primes.

Case 3: $p^*=5$. Since 2 and 3 divide n, and 5 doesn't, $(2-1)(3-1)(p-1) \geqslant 2^3 \cdot 5$ for $p \geqslant 21$; also, $(2-1)(3-1)(7-1)(11-1) \geqslant 2^4 \cdot 5$. These imply that the values of n which do not satisfy (6) are of the

form $n=2^\alpha 3^\beta q^\gamma$, $\alpha>0$, $\beta>0$, with $q=7,11,13,17,19$. For each of these q, the totitives of $n=2\cdot 3\cdot q$ contain a complete residue system modulo 5, (e.g., $1,5,23,29,37$). If $n=2^\alpha 3^\beta$, (5) is satisfied except in the following cases: $(\alpha,\beta)=(1,1)$, $(2,1)$, $(3,1)$, $(4,1)$, $(1,2)$, $(1,3)$, $(2,2)$, which must be checked individually. For $n=2\cdot 3$ and $n=2^2\cdot 3$ adding the totitives yields primes. For $n=2\cdot 3^2$ the smallest positive integer k such that $18k$ plus the totitives of n are all primes is $k=892$. The remaining values $n=24,36,48,54$ are all eliminated since their totitives include a complete residue system modulo 5.

Case 4: $p^*=7$. Since $2,3,5$ divide n and 7 doesn't, $(2-1)(3-1)(5-1)(p-1)\geqslant 2^4\cdot 7$ for $p\geqslant 15$, and $(2-1)(3-1)(5-1)(11-1)(13-1)\geqslant 2^5\cdot 7$, imply that the values of n which fail to satisfy (6) are of the form $n=2^\alpha\cdot 3^\beta\cdot 5^\gamma\cdot q^\delta$, where $q=11,13$, and $\alpha>0$, $\beta>0$, $\gamma>0$. Since the totitives of $n=2\cdot 3\cdot 5$ include a complete residue system modulo 7, all of these cases with $\delta=0$ are eliminated. For $\delta>0$, if $q=11$, one loses the residue 4 modulo 7, but this is replaced by 53; and if $q=13$, one loses 6 modulo 7 which is replaced by 41.

Case 5: $p^*=11$ and $p^*=13$ may be treated simultaneously. For since $(2-1)(3-1)(5-1)(7-1)(13-1)\geqslant 2^5\cdot 11$, and $(2-1)(3-1)(5-1)(7-1)(11-1)\geqslant 2^5\cdot 13$, the values of $n=2^\alpha 3^\beta\cdot 5^\gamma\cdot 7^\delta$, $\alpha\beta\gamma\delta\neq 0$, which fail to satisfy (5) are the same in both cases. Since the totitives of $n=2\cdot 3\cdot 5\cdot 7$ include a complete residue system both modulo 11 and 13, these integers are eliminated.

We now note that since $(2-1)(3-1)(5-1)(7-1)(11-1)(13-1)=90\cdot 2^6$, the next value of p^* which would have to be considered is $p^*=97$. However, the proof may now be completed by showing that all remaining n satisfy (5). Since for $n>30$, $p^*<\sqrt{n}$, (5) is implied by

$$\phi(n)\geqslant \sqrt{n}\cdot 2^r. \tag{7}$$

Writing $n=\prod_{i=1}^r p_i^{\alpha_i}$, we have

$$\frac{\phi(n)}{\sqrt{n}\cdot 2^r}=\prod_{i=1}^{r} p_i^{((\alpha_i/2)-1)}\left(\frac{p_i-1}{2}\right)\geqslant \prod_{i=1}^{n}\left[\frac{\sqrt{p_i}-\dfrac{1}{\sqrt{p_i}}}{2}\right]$$

so that (7) will be satisfied if

$$\prod_{i=1}^{r}\left[\frac{\sqrt{p_i}-\dfrac{1}{\sqrt{p_i}}}{2}\right]\geqslant 1. \tag{8}$$

Since $\sqrt{p_i}-1/\sqrt{p_i}\geqslant 2$ if $p_i\geqslant 7$, any n all of whose prime factors are $\geqslant 7$ satisfies (8). Also, as the left hand side of (8) is minimized by taking the p_i as small as possible, it remains to consider all $n=\prod_{i=1}^r p_i$ where the p_i are the first r primes in ascending order. Direct computation gives that if $r\geqslant 8$, n satisfies (8). However, the values of n with $r\leqslant 7$ are clearly among the set of n with $p^*\leqslant 89$, which have previously been considered. Thus the proof of the theorem is completed.

In [1] the question is raised of adding *prime* totitives to n either to yield only primes or only composites, again with a view to a largest such n. These problems are resolved in Theorems 2 and 3. First, quite trivially, we have:

THEOREM 2. *There exist infinitely many integers n to which each of its prime totitives can be added to yield a composite.*

Proof. This follows from the fact that there exist infinitely many odd integers n such that $n+2$ is composite, for adding any odd prime to such an n the resulting integer is always even.

Thus far, all the arguments have been rather self-contained, and quite elementary. For the theorems which follow, though the structure of the proofs is equally simple, they depend on various results concerning prime number. These in turn are stated and referenced, but lie deeper than the previous material.

THEOREM 3. *There is a largest integer n with the property that for some positive integer k each of the prime totitives of n added to kn always is prime.*

Proof. By our familiar argument, if the prime totitives of n include a complete residue system modulo p^*, for every $k \geqslant 1$ the prime totitives of n plus kn cannot all be prime (one such sum would be divisible by p^*). For a fixed i^* let $N'(i^*)$ = the number of primes $p \leqslant n$ such that $(p,n)=1$ and $p \equiv i^* \pmod{p^*}$, so that the theorem will follow if we can show that $N'(i^*)>0$ for all n sufficiently large.

To prove this we note first that $\Pi_{p<p^*} p \leqslant n$ implying $\Sigma_{p<p^*} \log p \leqslant \log n$. Since it is well known [3] that $\Sigma_{p<p^*} \log p \geqslant c_1 p^*$, $(c_1>0, p^*>2)$ it follows that $p^* \leqslant c_2 \log n$. Hence, although p^* depends on n, because it is limited in size in this way, a result called the Siegel-Walfisz Theorem [3] implies that for fixed i^*, $0<i^*<p^*$, and all large n, $\Pi(i^*,p^*,n)$ = the number of primes $p \leqslant n$ such that $p \equiv i^* \pmod{p^*}$ satisfies the inequalities

$$\frac{c_3}{\phi(p^*)} \frac{n}{\log n} \geqslant \Pi(i^*,p^*,n) \geqslant \frac{c_4}{\phi(p^*)} \frac{n}{\log n} \tag{9}$$

(where c_3, c_4, are positive constants not depending on i^*, p^*, or n.)

Since for $n>2$, p^* is a totitive of n, $N'(0)>0$; and we consider $0<i^*<p^*$. Then using the lower bound in (9) and subtracting $\nu(n)$ = the number of primes dividing n yields

$$N'(i^*) \geqslant \frac{c_4}{\phi(p^*)} \frac{n}{\log n} - \nu(n).$$

But $\phi(p^*)<p^*<c_2 \log n$ and $\nu(n) \leqslant \log n / \log 2$ so that this implies

$$N'(i^*) \geqslant \frac{c_5 n}{\log^2 n} - \frac{\log n}{\log 2}$$

which is positive for n large. This completes the proof of Theorem 3.

In [1] the question is also raised of adding *composite* totitives to n either to yield only primes or only composites. These problems are resolved in Theorems 4 and 5.

THEOREM 4. *There is a largest integer n with the property that for some positive integer k each of its composite totitives added to kn always is a prime.*

Proof. By our familiar argument it suffices to show that for all large n the composite totitives include a complete residue system modulo p^*. For each i^*, $0 \leqslant i^* < p^*$, let $\bar{N}(i^*)$ = the number of composite totitives of n which are congruent to i^* modulo p^*, so that our objective is to show that $\bar{N}(i^*)>0$. From (4) we have a lower bound for all totitives in this progression, and using the upper bound in (9) to compensate for those which are prime, we obtain

$$\bar{N}(i^*) \geqslant \frac{\phi(n)}{p^*} - 2^r - \frac{c_3}{\phi(p^*)} \frac{n}{\log n}. \tag{10}$$

Since $\phi(p^*)<p^* \leqslant c_2 \log n$, and from a well known theorem of Mertens [4], $\phi(n)>c_6 n/\log\log n$, (10) yields

$$\bar{N}(i^*) \geqslant \frac{1}{p^*}\left(\phi(n) - c_3 \frac{n}{\log n} \cdot \frac{p^*}{\phi(p^*)}\right) - 2^r \geqslant \frac{n}{c_2 \log n}\left(\frac{c_6}{\log\log n} - \frac{c_3 \log\log n}{c_6 \log n}\right) - 2^r,$$

and since $2^r < c_7 n^{1/2}$ (see [4]), the right side of the above is clearly positive for all large n. This completes the proof of Theorem 4.

THEOREM 5. *There exists a largest integer n to which each of its composite totitives can be added to obtain a composite. However, in the general case, there exist infinitely many integers n for which there exists a positive integer k such that kn added to each of the composite totitives of n is composite.*

Proof. The last part of the theorem follows if it can be shown that for each n, there exists some k such that the integers $kn+1, kn+2, \ldots, kn+n-1$ are all composite. Let $\tilde{p}$ be a prime such that $\tilde{p}>n!$,

then choose t large so that $\tilde{p}$ divides $tn!+1$ and note that $tn!+1>\tilde{p}$. Then if $k=t(n-1)!$, $\tilde{p}$ divides $kn+1$, and the integer $kn+i=tn!+i$, $i=2,\ldots,n-1$, is divisible by i and is thus composite.

To prove that there exists a largest integer n to which each of its composite totitives can be added to obtain a composite, note first that for n large there exist primes p such that $n+p^*<p<2n$ [4]. Moreover, since $p^*\leqslant c_2\log n$, the Siegel-Walfisz Theorem [3] implies that for any $\varepsilon>0$, the inequalities (9) hold with $c_3=1+\varepsilon$, $c_4=1-\varepsilon$, for all sufficiently large n. This in turn provides the existence of a prime p in the interval $[n+p^*,2n]$ such that $p-n\equiv 0 \pmod{p^*}$. Since $n<p<2n$, $c=p-n$ is a positive composite integer such that $c\leqslant n$ and $(c,n)=1$. Thus c is a composite totitive of n satisfying $n+c=p$ a prime. The theorem follows since this can be achieved for all large n.

References

[1] B. R. Santos, Twelve and its Totitives, this MAGAZINE, 49, (1976) 239–240.
[2] L. E. Dickson, History of the Theory of Numbers, Vol. 1, Chelsea, New York, 1952, p. 132.
[3] K. Prachar, Primzahlverteilung, Springer-Verlag, Berlin, 1957, p. 144.
[4] G. H. Hardy, and E. M. Wright, An Introduction to the Theory of Numbers, Oxford University Press, 1968.

Hoover's Problem

KUN-YUAN CHEN
The Upjohn Co.
Kalamazoo, MI 49001

THOMAS L. SAATY
University of Pennsylvania
Philadelphia, PA 19174

J. Edgar Hoover, the former FBI Chief, is reported to have had a great aversion to left turns, arising from the time his chauffeured car was hit while attempting a left-hand turn. Thereafter, particularly when the old lawman visited a new city, the driver had to work out the route in advance to avoid making left turns.

He actually faced two problems. The first was to avoid making Hoover dizzy from too many right turns, for he might then ban them too. That would have made it practically impossible to go anywhere by car and the driver would have been out of a job. So he always selected a route with what appeared to him to be the minimum number of right turns. The second problem facing the driver was that, on occasions which depended on the mood of the FBI director and in particular on whether he was late for an appointment, the driver had to opt for the shortest route without left turns even if such a route required more than the minimum number of right turns. When the director had more time, the driver sacrificed distance for the sake of minimizing right turns. The object of this paper is to assist Hoover's chauffeur and any large trailer-truck driver making deliveries in their effort to avoid left turns. In particular, we will determine means of finding routes from any point of origin to any destination that yields the minimum number of right turns or the minimum distance subject to the constraint that neither left turns nor U-turns are allowed.

Let us first model the problem by representing the planar road map of a city (which may include one-way streets) by a graph M. (We assume, of course, that the number of roads and intersections in the network is finite.) Denote the vertices of M (representing the intersections of the road map) by

$v_1,\ldots,v_n$ and the connecting arcs (representing the road links) by a_{ij} whenever there is a road directed from v_i to v_j. Note that a two-way street is represented by an arc a_{ij} from v_i to v_j and by another arc a_{ji} from v_j to v_i. (For simplicity we assume that all direct links from v_i to v_j are represented by the single arc a_{ij} and that the network has no loops.) We note that making a left or right turn when approaching an intersection along an arc is not just a property of the graph M, but rather of the way it is embedded in the plane. Different embeddings of the same graph yield different answers for the two Hoover problems.

From the point of view of avoiding left turns, travel in M from any origin that is an intermediate point of an arc is equivalent to travel from the terminal vertex of that arc having proceeded to it from the initial vertex of the same arc. (In other words, there are no left turns in mid-block.) Similarly, travel to a destination that is an intermediate point of an arc is equivalent to travel to the terminal vertex of that arc, having proceeded to it from the initial vertex of the same arc. Hence, to solve Hoover's problems in general, we need only solve them for trips from one vertex to another.

To begin, we "split" the vertices of M to obtain a new graph G in the following manner: A vertex v_j^i is in G iff a_{ij} is in M, and an edge (v_j^i, v_k^j) is in G iff the "walk" $v_i, a_{ij}, v_j, a_{jk}, v_k$ in M corresponds to a directed path in M from v_i to v_k through v_j that contains no left-hand or U-turns. We shall call such arcs a_{jk} (in M) the **admissible** arcs of v_j^i, and in this case we shall speak of travel from arc a_{ij} to arc a_{jk} as being **admissible**. Note that an arc out of v_j may be admissible for v_j^i, but not for $v_j^r, r \neq i$. (FIGURE 1 illustrates the relation between the two graphs M and G. In this illustration, arc a_{35} in M is admissible for vertex v_3^4 in G, but inadmissible for v_3^2.)

A well-known problem in graph theory which at first seems related is to find a way of orienting city streets (some of which may be one way) so that the resulting directed graph is **strongly connected**, i.e., so that any pair of vertices can be joined by a path so that one can travel from any point of the city to any other. The necessary and sufficient condition for the existence of such an orientation is that each edge belongs to at least one circuit. Hoover's problem, however, is more restrictive: the orientation is given and certain arcs may be admissible for some paths but not for others. Suppose the origin is an intermediate point on arc a_{pr} and the destination is an intermediate point on arc a_{qs}. Then a path π_1 from v_r^p to v_s^q in G corresponds to a path in M along admissible arcs from the origin to the destination. Thus, it is clear that Hoover's problem in M has a solution for all choices of v_i and v_j if and only if G is strongly connected. We shall call the graph M (as embedded) **Hooverian** if G is strongly connected.

We need algebraic methods to test whether a given graph is Hooverian, as well as to find the required minimum distances. To do this, we create the vertex matrix U of G; if G has N vertices, then U is an $N \times N$ matrix indexed by the vertices of G. Because of the way G was defined, U will have a unit entry corresponding to (v_j^i, v_k^j) when (and only when) a_{jk} is admissible for v_j^i; all other entries in U are zero. (The matrix $(I+U)^m$ is called by some the **reachability** matrix, for it describes which vertices may be reached from others in at most m arcs.) We shall show that M is Hooverian if and only if there is a positive integer $m < N$ (the size of U) such that $(I+U)^m > 0$, i.e., all entries are >0. This can be seen by considering the expression on the right hand side of

$$(I+U)^m = \sum_{k=0}^{m} \binom{m}{k} U^k.$$

The entry in the (i,j) position $(i \neq j)$ of $(I+U)^m$ is positive if and only if for at least one of the terms of the sequence of non-negative matrices $\binom{m}{1}U, \binom{m}{2}U^2, \ldots, U^m$, the entry in the i,j position is positive. The graph G is strongly connected if and only if for each pair of vertices v_r^p and v_s^q there is a simple path from v_r^p to v_s^q. To such a path must correspond a nonzero entry in U^l for $1 \leqslant l < N$ (since a simple path does not repeat vertices and can at most be of length $N-1$). Let m be the maximum of all such l. Then $m < N$, as desired.

Since each vertex of G corresponds to exactly one arc in M, it is easy to see that a Hamiltonian path in G, the path which traverses every vertex of G exactly once, corresponds to a Eulerian path in

the Hooverian graph, the path which traverses every arc of M exactly once without violating local admissibility. Thus a Hooverian graph M is Eulerian if and only if G is Hamiltonian. (This information tells us when we may tour without repetition all the streets of a town where no left turns are allowed.)

Now we take a Hooverian graph and show how to travel between any two points making a minimum number of right turns. In this case we use a modified vertex matrix V of G whose v_j^i, v_k^jth entry is indicated by one of the following:

∞: if travel in M from arc a_{ij} to arc a_{jk} is inadmissible.
0: if travel in M from arc a_{ij} to arc a_{jk} is admissible, but no right turn is needed.
1: if travel in M from arc a_{ij} to arc a_{jk} is admissible, and exactly one right turn is needed.

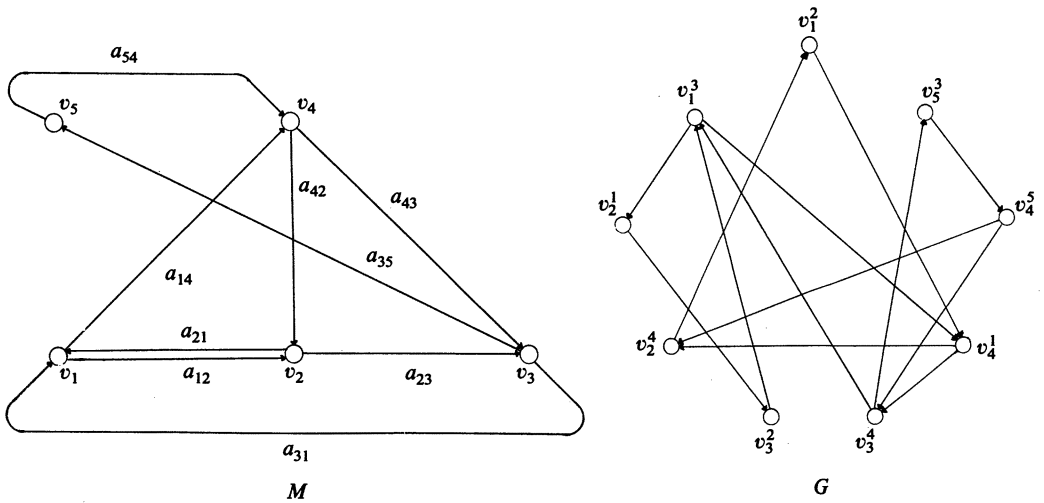
All other entries in V, namely those corresponding to v_j^i, v_k^l where $j \neq l$, are ∞.

We define the special matrix operation $C \equiv A * B$, by $c_{ij} = \min_k (a_{ik} + b_{kj})$ where $A = (a_{ij})$, $B = (b_{ij})$, $C = (c_{ij})$, and $a_{ij}, b_{ij} \geqslant 0$. It is clear that applying this operation m times to V, which we denote by V_*^m, will give the minimum number of right turns to go from any given initial vertex to a given terminal vertex in at most m steps. (Note that V has a zero at each diagonal entry.) Now all one has to check is walks of at most N arcs in G to go from any arc to any other arc. V_*^N gives the minimum number of right turns from any arc to any other arc in G. As the powers of V increase, the minimum operation assures us of the minimum number of right turns indicated by the number in each entry. To minimize labor it is sufficient to stop at a power m for which $V_*^m = V_*^{m+1}$.

To determine the minimum number of right turns from a vertex v_i to another vertex v_j in M we consider all arcs directed out of v_i and all arcs directed towards v_j and select that path with the minimum number of right turns. This can be computed as follows.

Let S_i be an N-tuple whose entries are zero in positions corresponding to the vertices of G which represent arcs directed out of v_i, and ∞ otherwise; let E_j be a corresponding N-tuple for all those arcs directed towards v_j. Then the minimum number of right turns from v_i to v_j is given by $S_i * V_*^m * E_j'$ (where E_j' is the transpose of E_j.) Let S be the matrix whose ith row is S_i and let E be the matrix whose jth row is E_j. Then the minimal number of right turns from vertex i to vertex j is given by the i,jth entry of the matrix $S^*V_*^m * E'$.

A similar procedure to that of finding the minimal number of right turns in a Hooverian graph is also available for finding the minimal length path between any two vertices in such a graph. Let L be



An (imbedded) graph M and the related graph G of arcs of M whose edges represent left-turn free routes in M.

FIGURE 1

the matrix of arc lengths corresponding to G. Its entries are zero on the main diagonal, ∞ for inadmissible arcs and the lengths of the admissible arcs otherwise. Use the operation $*$ and let m be the smallest integer such that $L_*^m = L_*^{m+1}$. Finally, define $\bar{S}_i$ analogously to S_i, but replace the 0 entries in S_i by the length of the arc directed out of v_i. Then the length of a minimal path between v_i and v_j is given by the i,jth entry of $\bar{S}_i * L_*^m * E_j'$. The minimum lengths are obtained from the matrix $\bar{S} * L_*^m * \bar{E}'$ where the rows of $\bar{S}$ are $\bar{S}_i$.

In either case, to determine the route with a minimum number of right turns and the shortest route, one must keep track of the matrices as they are raised to powers and trace backwards to determine the route giving rise to them.

As an illustration let us find the minimum number of right turns in the graph M of FIGURE 1. If we construct the vertex matrix V of G and compute powers m of V using the matrix operation $*$, it turns out that $m=6$. We then have

$$V = \begin{array}{c} \\ v_1^2 \\ v_1^3 \\ v_2^1 \\ v_2^4 \\ v_3^2 \\ v_3^4 \\ v_4^1 \\ v_4^5 \\ v_5^3 \end{array} \begin{array}{c} \begin{array}{ccccccccc} v_1^2 & v_1^3 & v_2^1 & v_2^4 & v_3^2 & v_3^4 & v_4^1 & v_4^5 & v_5^3 \end{array} \\ \begin{bmatrix} 0 & \infty & \infty & \infty & \infty & \infty & 1 & \infty & \infty \\ \infty & 0 & 1 & \infty & \infty & \infty & 0 & \infty & \infty \\ \infty & \infty & 0 & \infty & 0 & \infty & \infty & \infty & \infty \\ 1 & \infty & \infty & 0 & \infty & \infty & \infty & \infty & \infty \\ \infty & 1 & \infty & \infty & 0 & \infty & \infty & \infty & \infty \\ \infty & 0 & \infty & \infty & \infty & 0 & \infty & \infty & 1 \\ \infty & \infty & \infty & 1 & \infty & 1 & 0 & \infty & \infty \\ \infty & \infty & \infty & 1 & \infty & 0 & \infty & 0 & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & 0 & 0 \end{bmatrix} \end{array}$$

$$S = \begin{bmatrix} \infty & \infty & 0 & \infty & \infty & \infty & 0 & \infty & \infty \\ 0 & \infty & \infty & \infty & 0 & \infty & \infty & \infty & \infty \\ \infty & 0 & \infty & \infty & \infty & \infty & \infty & \infty & 0 \\ \infty & \infty & \infty & 0 & \infty & 0 & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & 0 & \infty \end{bmatrix},$$

$$E = \begin{bmatrix} 0 & 0 & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & 0 & 0 & \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & 0 & 0 & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & 0 & 0 & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & 0 \end{bmatrix},$$

and

$$(S * V_*^6) * E' = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 2 & 0 & 1 & 3 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

Thus, for example, the minimum number of right turns from v_2 to v_5 is 3, from v_3 to v_4 is 0, etc. Since no ∞ appears in this matrix, our graph is Hooverian. We note that if both the origin and the destination lie on arcs instead of on vertices, the minimum number of right turns is simply indicated by the number in the corresponding entry of matrix V_*^m.

Now assume that arc lengths in the graphs of FIGURE 1 from v_i to v_j are as follows:

$$\begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{array} \begin{array}{c} \begin{array}{ccccc} v_1 & v_2 & v_3 & v_4 & v_5 \end{array} \\ \begin{bmatrix} 0 & 1 & \infty & 2 & \infty \\ 1 & 0 & 1 & \infty & \infty \\ 2 & \infty & 0 & \infty & 3 \\ \infty & 1 & 2 & 0 & \infty \\ \infty & \infty & \infty & 1 & 0 \end{bmatrix} \end{array}$$

Then we have

$$L = \begin{array}{c} \\ v_1^2 \\ v_1^3 \\ v_2^1 \\ v_2^4 \\ v_3^2 \\ v_3^4 \\ v_4^1 \\ v_4^5 \\ v_5^3 \end{array} \begin{array}{c} \begin{array}{ccccccccc} v_1^2 & v_1^3 & v_2^1 & v_2^4 & v_3^2 & v_3^4 & v_4^1 & v_4^5 & v_5^3 \end{array} \\ \begin{bmatrix} 0 & \infty & \infty & \infty & \infty & \infty & 2 & \infty & \infty \\ \infty & 0 & 1 & \infty & \infty & \infty & 2 & \infty & \infty \\ \infty & \infty & 0 & \infty & 1 & \infty & \infty & \infty & \infty \\ 1 & \infty & \infty & 0 & \infty & \infty & \infty & \infty & \infty \\ \infty & 2 & \infty & \infty & 0 & \infty & \infty & \infty & \infty \\ \infty & 2 & \infty & \infty & \infty & 0 & \infty & \infty & 3 \\ \infty & \infty & \infty & 1 & \infty & 2 & 0 & \infty & \infty \\ \infty & \infty & \infty & 1 & \infty & 2 & \infty & 0 & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & 1 & 0 \end{bmatrix} \end{array}$$

$$\bar{S} = \begin{bmatrix} \infty & \infty & 1 & \infty & \infty & \infty & 2 & \infty & \infty \\ 1 & \infty & \infty & \infty & 1 & \infty & \infty & \infty & \infty \\ \infty & 2 & \infty & \infty & \infty & \infty & \infty & \infty & 3 \\ \infty & \infty & \infty & 1 & \infty & 2 & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & 1 & \infty \end{bmatrix},$$

and, consequently,

$$(\bar{S} * L_*^6) * E' = \begin{bmatrix} 4 & 1 & 2 & 2 & 7 \\ 1 & 4 & 1 & 3 & 8 \\ 2 & 3 & 4 & 4 & 3 \\ 2 & 1 & 2 & 4 & 5 \\ 3 & 2 & 3 & 1 & 6 \end{bmatrix}.$$

For example, the path of minimum length from v_2 to v_5 is 8.

Although our solution will not do Hoover much good, maybe there are some eccentrics who will find our left-turn-free procedure useful in traffic or perhaps to solve some other problem.

References

[1] D. R. Fulkerson, editor, Studies in Graph Theory, Parts I and II, Math. Assoc. of America, 1975.
[2] Frank Harary, Graph Theory, Addison-Wesley, 1969.
[3] Clifford Marshall, Applications of Graph Theory, Wiley, 1972.

Solving Linear Congruences

Donald R. Byrkit
University of West Florida
Pensacola, FL 32504

Most textbooks on number theory tend to present linear congruences in a more or less theoretical fashion, paying little attention to the mechanical problems inherent in the solution of a congruence with a large modulus such as $73{,}124x \equiv 1613 \pmod{81{,}763}$. Some semi-systematic approaches are given, but even the best of such approaches are largely trial and error. The Euclidean algorithm may be used, as in the introductory approach to Diophantine equations found in most books, but this overlooks the use of congruence notation entirely and totally undermines the inherent superiority of the congruence approach to Diophantine equations. The purpose of this note is to present a relatively

simple algorithm for solving linear congruences which uses the congruence notation only, takes advantage of the short cuts available with congruences, and will always yield a solution, if there is one, or indicate if there is not.

Consider the congruence $ax \equiv b \pmod{m}$. It is well known that this congruence has a solution if and only if d, the greatest common factor of a and m, is a factor of b. If there is a solution, we may find it by adding multiples of m to b until we obtain a multiple of a; that is, we wish to determine a value of y so that $b+my$ is divisible by a. This can be written as $b+my \equiv 0 \pmod{a}$, or as $my \equiv -b \pmod{a}$. Since d is a factor of b, this congruence has a solution, say y_0. Now this solution of the latter congruence provides a solution to the original congruence since $b+my_0$ will be divisible by a, and thus $x \equiv (b+my_0)/a \pmod{m/d}$ will be the desired solution.

This simple relationship between the two congruences forms the basis for a systematic procedure. It is apparent that if $m > a$, the congruence $my \equiv -b \pmod{a}$ can be replaced by $m'y \equiv -b \pmod{a}$ where $m' < a$. Thus a sequence of such congruences, with decreasing moduli and decreasing coefficients, can be employed as needed to obtain increasingly simpler congruences. Since the moduli form a decreasing sequence of positive integers, the sequence eventually terminates. From a practical viewpoint, if least absolute remainders are used and reductions made where obvious, the sequence will usually yield an easily solvable congruence rather quickly. If d (the *g.c.f.* of a and m) is not 1, this will become apparent at some point and d can be removed from the prior congruences as well. If no solution exists, this will also become apparent during the reduction. Some examples will illustrate the procedure.

To solve the congruence $863x \equiv 1586 \pmod{2151}$, we first examine the congruence $2151y \equiv -1586 \pmod{863}$. This reduces to $425y \equiv 140 \pmod{863}$ and can be reduced further to $85y \equiv 28 \pmod{863}$. The next step in the chain is to look at the congruence $863z \equiv -28 \pmod{85}$. This reduces to $13z \equiv -28 \pmod{85}$. The congruence $85w \equiv 28 \pmod{13}$ reduces to $7w \equiv 28 \pmod{13}$ which is easily solved for $w \equiv 4 \pmod{13}$. Substituting back into the prior congruences, we successively obtain $z \equiv [-28+(85)(4)]/13 \equiv 24 \pmod{85}$; $y \equiv [28+(863)(24)]/85 \equiv 244 \pmod{863}$; $x \equiv [1586+(2151)(244)]/863 \equiv 610 \pmod{2151}$.

Use of a hand calculator is helpful in these problems. Determining successive remainders, say when 2151 is divided by 863, is accomplished by dividing 2151 by 863, subtracting the integer part of the quotient, then multiplying again by 863, obtaining 425 (or perhaps 424.9997, if the calculator is quite inexpensive).

If a and m are not relatively prime, the same procedure still will yield the solution or show that there is none. The greatest common factor, d, of a and m, will remain the greatest common factor of the modulus and the coefficient in each successive congruence, and will become apparent at some point. An example of this type is $418x \equiv 649 \pmod{737}$. Beginning the algorithm, we set $737y \equiv -649 \pmod{418}$ which reduces to $-99y \equiv -231 \pmod{418}$ or $99y \equiv 231 \pmod{418}$. Then $418z \equiv -231 \pmod{99}$ or $22z \equiv -33 \pmod{99}$. At this point it becomes apparent that $d = 11$, and $2z \equiv -3 \equiv 6 \pmod{9}$ or $z \equiv 3 \pmod{9}$. If a hand calculator is being used, it may just be simpler to substitute back in directly, obtaining $y \equiv [231+(418)(3)]/99 \equiv 15 \pmod{38}$ and $x \equiv [649+(737)(15)]418 \equiv 28 \pmod{67}$. If calculations are a factor, d can be factored out of each congruence *before* substitution; that is, $99y \equiv 231 \pmod{418}$ can be replaced by $9y \equiv 21 \pmod{38}$ and solved for $y \equiv [21+(38)(3)]/9 \equiv 15 \pmod{38}$. In either case, the solution is obtained if we remember to divide the original value of m by d.

As a final example, we solve the congruence which was posed at the beginning of this note. It shows the unique advantages of the procedure, combining the reduction possibilities inherent in congruence notation with the inexorable infallibility of an algorithm. We first reduce $73{,}124x \equiv 1613 \pmod{81{,}763}$ to $-8639x \equiv 1613$ or $8639x \equiv -1613 \pmod{81{,}763}$. Then, applying our procedure, we have $81{,}763y \equiv 1613 \pmod{8639}$ which reduces to $4012y \equiv 1613 \equiv 10{,}252 \pmod{8639}$, or $1003y \equiv 2563 \pmod{8639}$. Then $8639z \equiv -2563 \pmod{1003}$, $-388z \equiv 446 \pmod{1003}$, $194z \equiv -223 \equiv 780 \pmod{1003}$, $97z \equiv 390 \pmod{1003}$. Finally $1003w \equiv -390 \pmod{97}$, reducing to $33w \equiv -2 \equiv -99 \pmod{97}$, so that $w \equiv -3 \pmod{97}$. Substituting, $z \equiv [390+(1003)(-3)]/97 \equiv -27 \pmod{1003}$; $y \equiv [2563+$

$(8639)(-27)]/1003 \equiv -230 \pmod{8639}$; and $x \equiv [-1613+(81{,}763)(-230)]/8639 \equiv -2177 \equiv 79{,}586 \pmod{81{,}763}$.

The reader is invited to compare this procedure with trial and error (trial and error first, please!) by solving some of the more difficult linear congruences given in any number theory text. Here are some interesting problems: $63{,}131x \equiv 5743 \pmod{81{,}214}$, $7033x \equiv 8131 \pmod{11{,}303}$, and $271{,}828x \equiv 637{,}421 \pmod{89{,}375}$.

References

[1] D. M. Burton, Elementary Number Theory, Allyn and Bacon, Boston, 1976.
[2] I. Niven and H. S. Zuckerman, An Introduction to the Theory of Numbers, Wiley, New York, 1966.
[3] O. Ore, Number Theory and Its History, McGraw-Hill, New York, 1948.
[4] A. J. Pettofrezzo and D. R. Byrkit, Elements of Number Theory, Prentice-Hall, Englewood Cliffs, N.J., 1970.
[5] B. M. Stewart, Theory of Numbers, Macmillan, Boston, 1964.

A Euclidean Construction?

LEON HENKIN
University of California, Berkeley
Berkeley, CA 94720

WILLIAM A. LEONARD
California State University, Fullerton
Fullerton, CA 92634

It has been said that there is no such thing as an uninteresting problem in mathematics... there are only uninteresting solutions. But suppose one had to make a choice: Would it be better to have an interesting solution, or a correct one? Or would we trade them both for a more provocative problem?

Mathematicians certainly recognize that "interest", as well as "correctness", is an important attribute of a mathematical proposition or method, even though we do not write in much detail about the former because we lack objective criteria for evaluating the degree of interest in any particular case. There are other non-objective aspects of a mathematical result which affect its value for mathematicians, such as esthetic considerations—of which simplicity may, perhaps, be considered one. A somewhat controversial criterion that is sometimes applied in evaluating a proof is "purity of method"; for instance, a theorem considered to belong to geometry ought to be proved, if possible, by "purely geometric methods," according to some mathematicians—even though no one can explain exactly what this means.

Several such non-objective questions of value arise when we attempt to perform Euclidean constructions, limited to straight-edge and compass, upon conics other than circles and straight lines. We begin with a simple problem, whose intuitive solution soon raises unexpected foundational questions.

The original problem. Given a portion of a parabola, as in FIGURE 1, can we find its vertex using only a compass and straightedge? (The precise origin of this problem is not known, but it seems to have passed around in meetings of the California Mathematics Council.)

A seeming solution. Draw any chord of the parabola, RS, and then construct a second chord, $R'S'$, parallel to the first. Letting M and M' be the respective midpoints of these chords, construct the line m which joins them. We *claim* that m must be parallel to (or coincide with) the axis of the parabola.

To justify this claim, imagine a system of Cartesian coordinate axes set up in such a way that the origin is at the vertex of the parabola, and the positive portion of the y-axis coincides with the parabola's axis. In this case the parabola will have an equation of the form $y=ax^2$ for some $a>0$. Letting b, b', c and c' be the respective abscissas of the points R, R', S and S', we see that the ordinates of these points are respectively ab^2, ab'^2, ac^2 and ac'^2, while for the abscissas of M and M' we obtain $\frac{c+b}{2}$ and $\frac{c'+b'}{2}$ respectively. But because RS and $R'S'$ are parallel, their slopes are equal. So

$$\frac{ac^2-ab^2}{c-b}=\frac{ac'^2-ab'^2}{c'-b'}$$

from which we easily get $c+b=c'+b'$. Thus M and M' have the same abscissa so that the line joining them, m, must be parallel to (or coincide with) the y-axis, i.e., the axis of the parabola, as claimed.

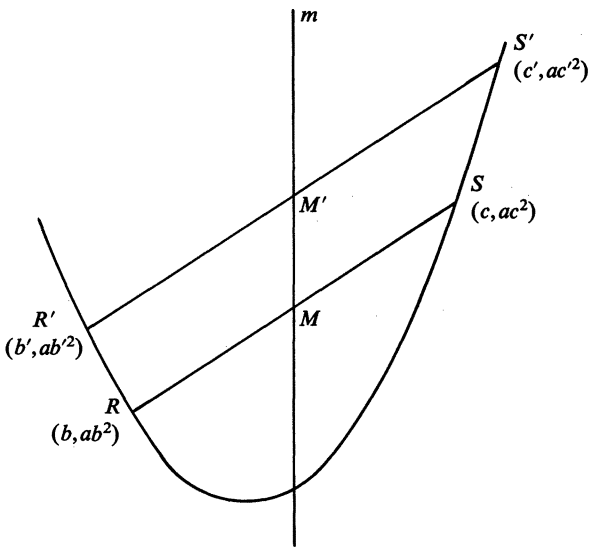


FIGURE 1

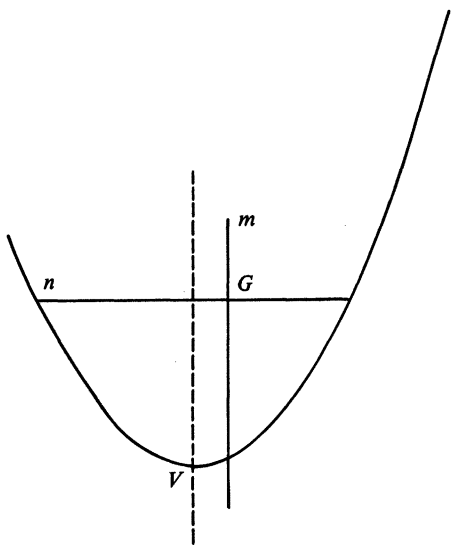


FIGURE 2

Having established our claim, we can proceed with our construction as follows. At a suitable point G on m, construct a chord n of the given parabolic arc, with n perpendicular to m. Then the perpendicular bisector of n is obviously the axis of the parabola, and so its point of intersection V with the parabola is the desired vertex (see FIGURE 2).

Is this a geometric solution to our geometric problem? Although we resorted to some algebra to justify a claim about our construction, the construction itself involved only classical geometric components: constructing a line parallel to a given one, finding the mid-points of given segments, erecting a perpendicular to a given line at a given point, and forming the perpendicular bisector of a given segment. Thus there is no doubt that we have obtained a geometric solution to our original problem.

No doubt whatever? Well, it is true that if we are feeling argumentative, we might raise a little doubt. For instance, in our formulation of the original problem we assumed that a portion of a parabola was given "as pictured". But what if the given portion of the parabola did not contain the vertex? Then the line n would not meet the parabola at two points, and our solution could not be carried out.

We look at the picture again. It certainly *seems* as if the given portion of the parabola contains the vertex! So the doubt raised above does not apply to our original problem *with picture*, but only indicates that a further problem, corresponding to a different picture, remains unsolved.

Are we satisfied? Let's doubt a little harder. Even if the vertex is present in the given portion of the parabola, recall that we constructed the line n perpendicular to m at a certain point G. But just how did we choose G? According to our directions, G was any point on m such that the perpendicular through it would intersect the given portion of the parabola at two points. Are we supposed to find G "by eye"? If we try one point and find that the perpendicular to m through it only meets the given parabolic arc in a single point, can we use this information to choose a second point G' which will "work" better? Is it possible that *no* point on m will "work"? Doubts, doubts, doubts! Let's see if we can find another solution.

The problem restated and resolved. We are given *any* parabolic arc and we seek to construct its vertex by straight-edge and compass. Starting from points R and S on the arc, we proceed as before to construct a line, m, through the midpoint, M, of RS, which is parallel to (or coincides with) the axis of the parabola. This time, however, we produce m until it interests the given arc at a point, say O, and there we construct a perpendicular to m, say q (see FIGURE 3).

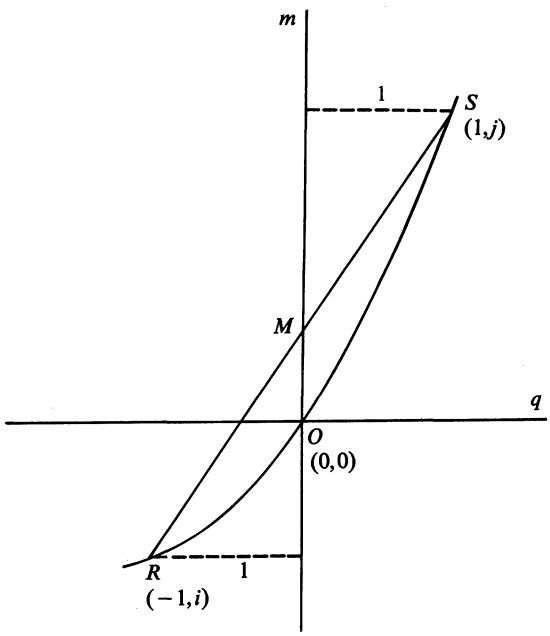


FIGURE 3

Now let us consider a Cartesian coordinate system in which q and m are the axes of abscissas and ordinates respectively, while the unit distance is the distance of S (and hence of R) from m. If S has coordinates $(1,j)$ for some j, then R will have coordinates $(-1,i)$ for some i, and of course O will have coordinates $(0,0)$. By dropping perpendiculars from S and from R upon q, we can, of course, obtain segments of length $|j|$ and $|i|$.

Now if (h,k) are the coordinates of the elusive vertex, our parabola will have an equation of the form $y-k=a(x-h)^2$, for some number, a. Since R, O and S lie upon the parabola, their coordinates must satisfy this equation, thus giving us the relations

$$i-k=a(-1-h)^2$$
$$-k=ah^2$$
$$j-k=a(1-h)^2$$

and from these we compute, by elementary algebra,

$$h=\frac{i-j}{2(i+j)} \quad \text{and} \quad k=\frac{-(i-j)^2}{8(i+j)}.$$

These formulas enable us to proceed from our earlier segments of lengths $|i|$ and $|j|$, and to construct new segments of lengths $|h|$ and $|k|$; we can do this by straight-edge and compass by using repeatedly the well-known methods of constructing, from any two given segments, new segments representing the sum, difference, product and quotient of the given ones.

Finally, by means of the segments of length $|h|$ and $|k|$, we can find the point with coordinate h on the line q, and the point with coordinate k on the line m. Then, by constructing the lines through h and k that are respectively parallel to m and q, and obtaining their intersection, we finally locate the vertex of our given parabola.

The solutions compared. While our second solution avoids some of the doubts and inadequacies of the first one, a persistent doubter could still raise questions about it. To begin at the beginning, in order to obtain m, we first construct a chord RS of the given parabolic arc. Do the rules of Euclidean construction allow us to pick points R and S from a given figure in an arbitrary way? Even if we allow ourselves this liberty, and then pick a third point R' on the arc—what if the line through R' parallel to RS fails to intersect the given arc? Well, we just try again, making sure this time to pick R' on the part of the arc that is between R and S. But are we allowed to do this "by eye", or do we need a construction?

It seems that we are not going to be able to rid ourselves of *all* doubts unless we face the fundamental question of just what it *means* to construct a point by straight-edge and compass from a given parabolic arc. In the classical Euclidean constructions, we are always given a finite set of points (or a figure formed by straight lines and circles determined by a finite point set), and we construct new points by intersecting lines and circles that are determined by given (or previously constructed) points. But our present problem is not of this type, so the meaning we are seeking will have to be invented, rather than found. As this quest would take us beyond the scope of the present paper, let us look at our two solutions from another perspective. Putting aside questions of adequacy and correctness, which is more *interesting*?

Hardly the second one! Instead of seeking the vertex of our parabola in a truly geometric spirit, we have switched to analytic geometry, solved a system of equations, and then ground away with straight-edge and compass to find a segment of length $(i-j)/2(i+j)$. What could be more geometrically opaque? By contrast, each of the elements of our first construction has a clear, intuitive meaning. What a shame it's inadequate! Could there possibly be a solution to our problem which is both interesting and adequate?

We shall see below that there is, indeed, a really geometric solution to our little problem. But before we look at it, let us consider a bonus that comes to us through the use of analytic methods. Namely, we can see immediately that the same method will solve some related problems. For example, we can proceed at once to construct the focus or the directrix of our given parabola. But what about the other conics—the ellipses and hyperbolas? Given an arbitrary arc of one of them, can we obtain various special points and lines associated with it? If so, how much less than a continuous arc would it suffice to be given?

When we seek to answer these questions by methods of analytic geometry, we come out with formulas involving not only the "rational operations" of addition, subtraction, multiplication and division, but square roots as well. However, it is well known how, when a segment is given, we can use straight-edge-and-compass constructions to obtain a second segment whose length is the square root of the first, providing a unit segment is agreed upon. For this reason, we *can* get straight-edge-and-compass solutions to various construction problems about hyperbolas and ellipses. But are there *interesting* solutions—"truly geometric" ones? Let us return to our parabola to get a feeling for what is wanted.

The geometric location of the parabola's vertex. As in our first two solutions, we draw a chord RS on our given parabolic arc, and then construct a line m through its midpoint which is parallel to the parabola's axis. Extending m to the point O where it intersects the arc, we now construct a line p through O which is parallel to RS. Look at the diagram: Does p seem to be tangent to the parabola at O? It is. We can see this by checking the computations in our first solution and finding that O has abscissa $(c+b)/2$, while RS (and hence p) have slope $a(c+b)$, for the coordinate system in which the parabola has equation $y=ax^2$. Since the derivative in general is $2ax$, its value at the point O is $2a(\frac{a+b}{2})=a(c+b)$, which is the slope of p. Thus p is a tangent, as claimed (see FIGURE 4).

At this point we recall the principle of the reflecting telescope: Rays parallel to the axis of a parabolic mirror are reflected through the focus. Treating m as such a ray, and using p to determine the angle of incidence of m, we construct a line r through O in such a way that angle of reflection equals angle of incidence. Then we can be sure that r passes through the focus, F, of our parabola. By shifting from the original chord RS to a nearby $R'S$ and repeating our construction, we obtain a second line r' through F. Finally, the intersection of r and r' gives us F itself. The line s through F that is parallel to m will therefore be the axis of our parabola, and if s happens to intersect our given

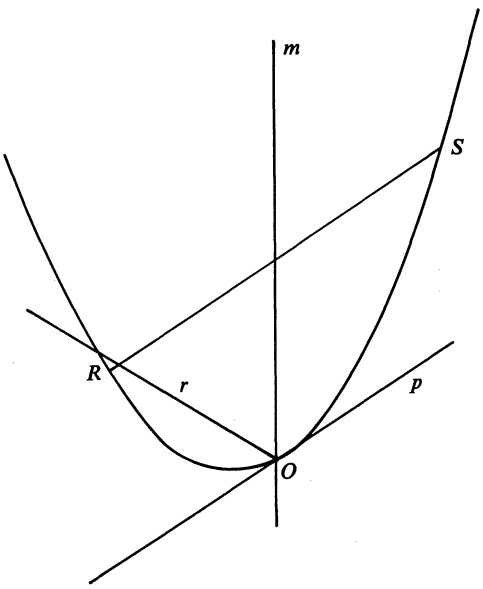


FIGURE 4

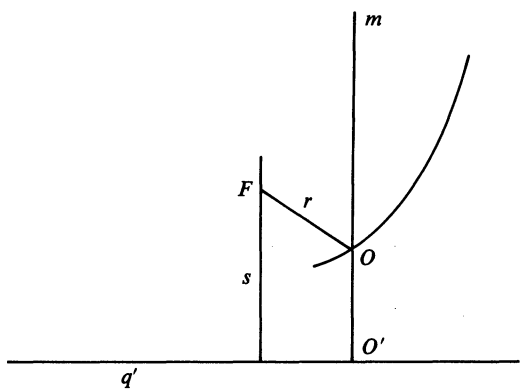


FIGURE 5

arc, then we have the sought-for vertex. In the contrary case we extend m to a point O' so that the distance OO' equals OF; this guarantees that the line q' through O' that is perpendicular to m is the directrix of the parabola, as a parabola is the locus of points equidistant from a given point (the focus) and given line (the directrix). Then, by dropping the perpendicular s from F to q' and finding the midpoint of the resulting segment, we arrive at the vertex that we have been seeking (see FIGURE 5).

It is hard to deny that this solution (which we owe to the referee of an earlier version of this paper) is "truly geometric". Whether solutions of this kind can be found for the ellipse and hyperbola, where we only have solutions based on analytic formulas, is open. Beyond these problems lies the foundational question that we by-passed: What do we mean, in general, by saying that a given point is constructable by straightedge and compass from a given conic arc?

Henkin's work is partially supported by the National Science Foundation, Grant No. MCS74-23878.

Orthomodular Lattices and Quantum Physics

ROBERT PIZIAK
Centre College of Kentucky
Danville, KY 40422

One of the famous problems David Hilbert presented to the International Congress of Mathematicians in Paris in 1900 was his 6th problem on the mathematical treatment of the axioms of physics. As with the other famous problems set forth that day, much interesting mathematics has resulted from the attempt to apply the axiomatic method to physical science. This kind of research continues to the present day as a significant group of mathematicians, physicists and even philosophers seek out what the mathematical foundations of quantum physics (or any empirical science for that matter) should be.

Most current research on the conceptual foundations of quantum physics is based on the collection of "yes-no propositions" of a physical system, the so-called "logic" of the system. These are the empirically verifiable propositions about the system which allow only two outcomes, "yes" or "no". But, what structure is it reasonable to impose on the propositional calculus of quantum mechanics? It is the purpose of this paper to describe the currently most accepted model, that of an orthomodular lattice. Once this model is accepted, the concepts of states, observables, symmetries, etc. can be developed in a more or less natural way.

The word "logic" and the phrase "propositional calculus" bring to mind the work of Boole. Boole gave us an algebraic model of Aristotle's classical propositional calculus, a structure we now call a Boolean algebra. During the early part of the twentieth century, the Boolean algebra model for the logic of propositions came under criticism by logicians and mathematicians. The result was that several non-Boolean models for the logic of propositions were created. The reader may be familiar with the Brouwerian model or the many-valued logics. While logicians were trying to replace the Boolean model for logic, physicists were being forced to supplant the classical, deterministic Newtonian model of the physical world by the models of Schrodinger and Heisenberg of the microworld of quantum phenomena. It can be argued that the logic of empirically testable propositions of Newtonian physics fits the Boolean model of logic. [5, p78ff] However, the novelties of quantum mechanics show this model to be no longer tenable. It would be pleasing to report that logicians and mathematicians had anticipated the need and created a new logical model of the sort quantum physics required. However, this was not the case. It was a famous paper of Birkhoff and von Neumann [2] in 1936 which launched the still active investigation of the structure of the logic of a quantum system.

The structure of a Boolean algebra is no doubt familiar to anyone who has studied set theory, or classical propositional logic or even the theory of switching circuits. An orthomodular lattice can be viewed as a generalization of a Boolean algebra. Indeed, their theories are closely intertwined. Perhaps the most intuitive approach to the concept of an orthomodular lattice is through the language of a "logic of propositions". Let us now build up the concept of an orthomodular lattice in this language.

Let L denote a set of objects which we shall call "propositions". Suppose $\wedge$ and $\vee$ are two binary operations on L. We shall call them "meet" and "join" respectively. They model the logical connectives "and" and "or". Also let $\leqslant$ be a partial order relation on L related to $\wedge$ and $\vee$ by $a \leqslant b$ if and only if $a = a \wedge b$ if and only if $b = a \vee b$. Let $' : L \mapsto L$ be a unary operation on L modeling logical negation. Finally designate two elements of L, 0 and 1 to model the contradictory proposition and tautological proposition respectively. Then the structure $(L, \wedge, \vee, \leqslant, 0, 1, ')$ is called an **orthocomplemented lattice** if it satisfies the following requirements:

AXIOM I. The commutative law holds for meet and join; i.e., $a \vee b = b \vee a$ and $a \wedge b = b \wedge a$ for all a, b in L.

AXIOM II. The associative law holds for meet and join; i.e., $a \vee (b \vee c) = (a \vee b) \vee c$ and $a \wedge (b \wedge c) = (a \wedge b) \wedge c$ for all a, b, c in L.

AXIOM III. The absorption laws hold between meet and join; i.e., $a \vee (a \wedge b) = a$ and $a \wedge (a \vee b) = a$ for all a, b in L.

AXIOM IV. The element 0 in L is such that $0 \vee a = a$ for all a in L and the element 1 in L is such that $1 \wedge a = a$ for all a in L.

AXIOM V. The unary operation $'$ satisfies

(i)	$a = a''$	(double negation)
(ii)	if $a \leqslant b$, then $b' \leqslant a'$	(contraposition)
(iii)	$a \wedge a' = 0$	(noncontradiction)
(iv)	$a \vee a' = 1$	(excluded middle)

The reader should consult Birkhoff [1] for a more complete introduction to the elements of lattice theory.

The partial order relation, which allows an interpretation as an implication relation between propositions, also allows us to draw pictures of these logical structures called **Hasse diagrams**. If $a \leqslant b$ holds between a and b, a Hasse diagram denotes this pictorially by placing the point b above a and connecting these points by a line segment. The following simple example is a system which satisfies all of the axioms imposed thus far:

Example 1. Let L be the set of four objects $\{0, a, b, 1\}$, and let the operations $\vee, \wedge, '$ be defined by:

$\vee$	0	a	b	1
0	0	a	b	1
a	a	a	1	1
b	b	1	b	1
1	1	1	1	1

$\wedge$	0	a	b	1
0	0	0	0	0
a	0	a	0	a
b	0	0	b	b
1	0	a	b	1

$0' = 1$
$a' = b$
$b' = a$
$1' = 0$

The Hasse diagram for L is given in FIGURE 1.

So far the axioms we have imposed on L seem to reflect the familiar logical relationships between propositions. These relationships can also be supported physically. Now comes the delicate matter of what other structure is allowed (perhaps demanded?) by quantum physics. To get a Boolean algebra, we need only add the distributive law:

DISTRIBUTIVE LAW. $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$ for all a, b, c in L.

Thus a **Boolean algebra** is just an orthocomplemented distributive lattice.

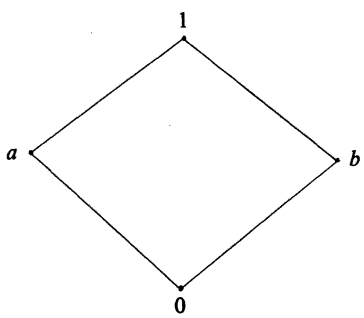


FIGURE 1

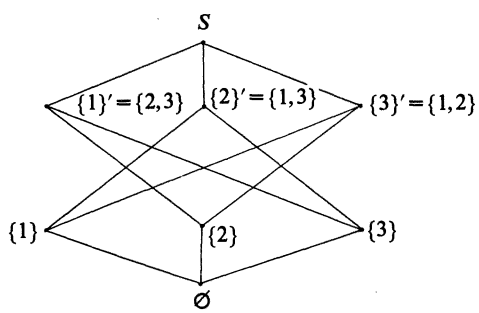


FIGURE 2

Example 2. Let S be any set and let $\mathcal{P}(S)$ denote the set of all subsets of S. Let $L=\mathcal{P}(S)$. Then L is a Boolean algebra if meet is taken to be ordinary set intersection, join is taken to be set union, the empty set $\varnothing$ is taken as 0, the set S is taken as 1 and negation corresponds to set complementation. The induced partial order relation is just set inclusion. Then $(\mathcal{P}(S), \cap, \cup, \subseteq, \varnothing, S, ')$ is a Boolean algebra. The reader may wish to draw some Venn diagrams to help visualize the validity of the distributive law $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ in this example. As an explicit example to which we shall later refer, consider $S=\{1,2,3\}$. The corresponding Hasse diagram of $\mathcal{P}(S)$ is given in FIGURE 2. This lattice is sometimes referred to as 2^3.

It turned out that the objection to Boolean algebras as appropriate for the logic of quantum mechanics came from an unexpected direction. We quote from Birkhoff and von Neumann [2]:

> "... whereas logicians have usually assumed that the properties of negation were the ones least able to withstand a critical analysis, the study of (quantum) mechanics points to the *distributive identities* as the weakest link in the algebra of logic."

The point is that in quantum physics not all propositions about a quantum mechanical system can be simultaneously measured to arbitrary accuracy (e.g. statements about the position and momentum of an electron). Gone is the determinism of Newtonian physics. The quantum physics is an intrinsically statistical theory with essential uncertainties built in. To demand that the logic of quantum mechanics be Boolean would demand a compatibility between all propositions that is not present.

Birkhoff and von Neumann suggested that the distributive law be replaced by something a bit weaker, the modular law:

MODULAR LAW. If $a \leqslant c$, then $(a \vee b) \wedge c = a \vee (b \wedge c)$ for all a,b,c in L.

It is not difficult to show that if a lattice satisfies the distributive law, it must also satisfy the modular law, but not conversely. The suggestion then was made that the appropriate model to study the logic of quantum mechanics is an orthocomplemented modular lattice. At first, this was particularly appealing because of the close connection between these structures and the structures of projective geometry. The suggested projective geometries also possessed dimension functions which allowed an *a priori* probability notion into the theory. Again, let us consider some examples.

Example 3. Let V be the vector space of real n-tuples (or complex n-tuples). Let $L=\mathrm{Lat}(V)$ be the set of all linear subspaces of V. If M and N are in L, take $M \vee N = M+N$ and $M \wedge N = M \cap N$. As the negation, consider the passage from a subspace to its orthogonal complement, $M \mapsto M^{\perp}$ with respect to the usual inner product. Again, the induced partial order is inclusion. Then $(\mathrm{Lat}(V), \cap, +, \subseteq, (0), V, ^{\perp})$ is an orthocomplemented modular lattice which is not Boolean unless the dimension of V is two or less.

Example 4. A visual example of an orthocomplemented modular lattice which is not Boolean is given in FIGURE 3. Note that $a \wedge (b \vee d) = a \wedge c' = a$, although $(a \wedge b) \vee (a \wedge d) = 0 \vee 0 = 0$. Thus the distributive law fails, making this example not Boolean.

Subsequent research has shown that even the modular law is not tenable in the logic of quantum mechanics. The argument against orthocomplemented modular lattices is somewhat difficult and will not be attempted here (see [5]). The law that has gained the widest current acceptance is the orthomodular law, still a further weakening of the distributive law:

ORTHOMODULAR LAW. If $a \leqslant b$, then $b = a \vee (b \wedge a')$ for all a,b in L.

Thus, an **orthomodular lattice** is an orthocomplemented lattice that satisfies the orthomodular law. It is not difficult to see that any Boolean algebra or any orthocomplemented modular lattice is an orthomodular lattice. Here is an example of an orthomodular lattice which is neither:

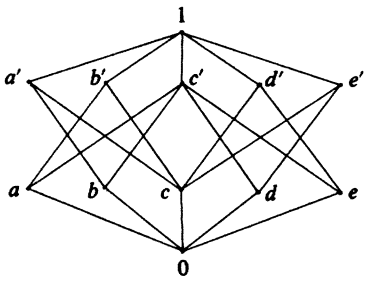


FIGURE 3

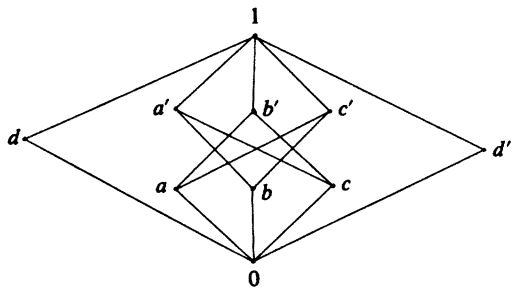


FIGURE 4

Example 5. In the lattice of FIGURE 4 we have $a \leq b'$ but $(a \vee d) \wedge b' = 1 \wedge b' = b'$ while $a \vee (d \wedge b') = a \vee 0 = a$. Thus the modular law and hence also the distributive law fail. However, this is an orthomodular lattice.

Although the global distributive law fails in the general orthomodular lattice, there are many Boolean subalgebras in any orthomodular lattice and hence there is a great deal of "local" distributivity. This is usually studied through the relation of **compatibility**: aCb if and only if $a = (a \wedge b) \vee (a \wedge b')$. Compatible propositions have the interpretation of being "simultaneously testable." In a Boolean algebra, all pairs of propositions are compatible. It is the existence of incompatible pairs of propositions in quantum mechanics that led to the rejection of Boolean algebras as appropriate models for the logic of quantum mechanics. One can consider the set of all elements in an orthomodular lattice which are compatible with every element. This is called the **center** of the orthomodular lattice and it is always a Boolean subalgebra. It measures in some sense how much the original lattice may be broken down into simpler pieces. Physically, the center seems to have something to do with superselection rules. If we refer back to Example 4, it can be seen that the center of this orthomodular lattice is the lattice given in FIGURE 5, which is just an isomorphic copy of Example 1. Can you find the center of Example 5?

So, as we said at the beginning, an orthomodular lattice can be seen as a generalization of a Boolean algebra. Indeed, Boolean algebras are crucial to the understanding of orthomodular lattices. One of the great unanswered questions in this theory is just how the maximal Boolean subalgebras in an orthomodular lattice fit together to determine its structure. Look back at the Hasse diagrams of Example 2 and Example 4 to see how two copies of the lattice 2^3 of Example 2 are "pasted" together to yield the lattice of Example 4.

Further insight into orthomodular lattices and quantum mechanics requires knowledge of Hilbert spaces and their linear operators. Example 3 gives a very brief hint of the relevant structure since that

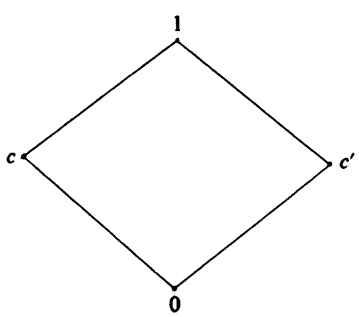


FIGURE 5

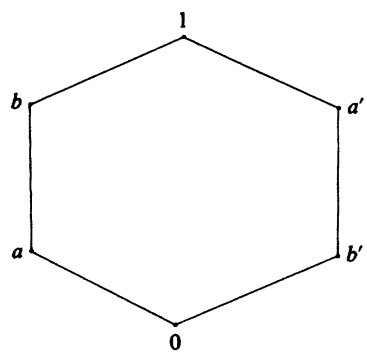


FIGURE 6

example deals with finite dimensional Hilbert spaces. The knowledgeable reader may appreciate at this point that the projection operators on an infinite dimensional Hilbert space form an orthomodular lattice which is not modular.

We end with two remarks. First, orthomodular lattices are of interest independent of their usefulness in the study of the logical foundations of quantum physics; see, for example, the excellent article of Holland [4]. Second, every lattice depicted so far in this paper is an orthomodular lattice. Lest the reader think these structures too ubiquitous, we invite consideration of the simple orthocomplemented lattice given in FIGURE 6.

This paper is based on a talk presented by the author to the 1975 Kentucky section meeting of the Mathematical Association of America. The revised version of this paper was written while the author was supported by a Summer Grant from Centre College of Kentucky. The author would like to thank Kathy Boorman for preparing the figures for this paper.

References

[1] G. Birkhoff, Lattice theory, Amer. Math. Soc. Colloq. Publ., XXV.
[2] G. Birkhoff, and J. von Neumann, The logic of quantum mechanics, Ann. of Math., 37 (1936) 823–843.
[3] P. Halmos, Algebraic Logic, Chelsea Pub. Co., New York, 1962.
[4] S. S. Holland, Jr., The current interest in orthomodular lattices, Trends in Lattice Theory, J. C. Abbot ed., Princeton, N.J. (1970), pp. 41–126.
[5] J. M. Jauch, Foundations of Quantum Mechanics, Addison-Wesley, Reading, Mass. (1968).
[6] T. F. Jordan, Linear Operators For Quantum Mechanics, J. Wiley & Sons, New York (1969).
[7] G. W. Mackey, Mathematical Foundations of Quantum Mechanics, W. A. Benjamin Inc., New York (1963).
[8] C. H. Randall, and D. J. Foulis, An approach to empirical logic, Amer. Math. Monthly, 77 (1970) 363–374.
[9] P. Suppes, The probabilistic argument for a nonclassical logic of quantum mechanics, Philo. Sci., 33 (1966) 14–21.
[10] J. von Neumann, Mathematical Foundations of Quantum Mechanics, Princeton University Press, Princeton, N.J. (1955).

Uncorrelated Dependent Random Variables

JAVAD BEHBOODIAN
Pahlavi University
Shiraz, Iran

It may not be easy for a student (or even an instructor) of probability and statistics to construct off-hand a pair of uncorrelated dependent random variables. Feller [2, p. 236] gives a beautiful example by taking $X=U+V$ and $Y=U-V$, where U and V have the same distribution. This example can be generalized further by considering two arbitrary non-degenerate random variables U and V, discrete or continuous, with finite equal means and variances. By choosing appropriate U and V we can construct many pairs of uncorrelated dependent random variables X and Y. However, showing dependency of X and Y, especially when U and V are continuous, requires some effort. The purpose of this note is to present a simple general method for constructing such pairs by using symmetric random variables.

A random variable X is **symmetric** about a constant c if and only if $X-c$ and $c-X$ have the same distribution, and the expectation of X, if it exists, is equal to c. When $c=0$, we call X simply a symmetric random variable. The key to our method is the following useful theorem about symmetric random variables. Our proof depends on a simple lemma.

THEOREM. *Let $g(x)$ be an odd and $h(x)$ an even real-valued (measurable) function. If X is a symmetric random variable, then the random variables $Y=g(X)$ and $Z=h(X)$ are uncorrelated, provided that Y and Z are non-degenerate and all the first and second moments of Y and Z exist.*

LEMMA. *Let X be a symmetric random variable, and let $y=g(x)$ be an odd real-valued function. Then, the random variable $Y=g(X)$ is also symmetric.*

Proof. Since X and $-X$ have the same distribution by symmetry, it is easy to show that $g(X)$ and $g(-X)$ have also the same distribution. Using the fact that $g(x)$ is an odd function, we have $g(-X)=-g(X)$. Hence $g(X)$ and $-g(X)$ have the same distribution, i.e., $g(X)$ is symmetric.

Proof of the Theorem. It is clear that $g(x)h(x)$ is an odd function. Thus, by the above lemma and the fact that all the first and second moments of Y and Z exist, Y and YZ are both symmetric with zero means. Hence, we have $\mathrm{Cov}(Y,Z)=0$, i.e., Y and Z are uncorrelated.

To construct uncorrelated random variables that are dependent, we need an additional result: *if the odd function $g(x)$ is one-to-one and the even function $h(x)$ is two-to-one, then Y and Z are dependent*. To show dependency of Y and Z, we need only construct two Borel sets B and C such that $P(Y\in B \text{ and } Z\in C)$ does not equal $P(Y\in B)P(Z\in C)$ (see [1, p. 49], for example).

Since Y and Z are non-degenerate, we cannot have $P(X=0)=1$ or $P(X=\pm x)=\frac{1}{2}$ for some $x>0$. Hence, by the fact that X is symmetric, there exist two disjoint symmetric intervals $A=(0,a)$ and $A'=(-a,0)$ for some $a>0$ such that $0<P(X\in A)=P(X\in A')<\frac{1}{2}$. Let $B=g(A)$ and $C=h(A)$ be the images of A under $g(x)$ and $h(x)$. Since $g(x)$ is one-to-one and the even function $h(x)$ is two-to-one, the inverse images of B and C are $g^{-1}(B)=A$ and $h^{-1}(C)=A\cup A'$. Hence

$$\begin{aligned} P(Y\in B \text{ and } Z\in C) &= P(g(X)\in B \text{ and } h(X)\in C) \\ &= P(X\in g^{-1}(B) \text{ and } X\in h^{-1}(C)) \\ &= P(X\in A \text{ and } X\in A\cup A') \\ &= P(X\in A) \end{aligned}$$

while

$$\begin{aligned} P(Y\in B)P(Z\in C) &= P(X\in A)P(X\in A\cup A') \\ &= 2P^2(X\in A). \end{aligned}$$

Since $P(X\in A)$ is neither 0 nor $\frac{1}{2}$, these two probabilities cannot be equal. Hence Y and Z cannot be independent.

Here are two simple examples of uncorrelated dependent random variables constructed by our method.

EXAMPLE 1. Let the symmetric random variable X take -1, 0, 1 with probabilities $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$ respectively. The random variables $Y=X^3$ and $Z=X^2$ are uncorrelated but dependent.

EXAMPLE 2. Let the symmetric random variable X have standard normal distribution. The random variables $Y=X$ and $Z=|X|$ are uncorrelated but dependent.

The author wishes to thank the referee for his useful suggestions.

References

[1] K. L. Chung, A course in Probability Theory, 2nd. ed., Academic Press, 1974.
[2] W. Feller, An Introduction to Probability and its Applications, Vol. 1, 3rd. ed., Wiley, 1968.

PROBLEMS

Dan Eustice, Editor
Leroy F. Meyers, Associate Editor
The Ohio State University

Proposals

To be considered for publication, solutions should be mailed before June 1, 1979.

1054. a*. Show how to construct triangle ABC by straightedge and compass, given side a, the median m_a to side a, and the angle bisector t_a to side a.

b*. Show how to construct triangle ABC by straightedge and compass, given angle A, m_a, and t_a. [*Jerome C. Cherry, Santa Maria, California.*]

1055. Find the limit as $n\to\infty$ of

$$a_n(p)=\sum_{k=0}^{n}\binom{2n+1}{k}p^k(1-p)^{2n+1-k},\quad 0<p<1.$$

[*Andreas N. Philippou, University of Patras.*]

1056. "Oh, drat!" exclaimed the meteorologist stormily. "I've just anchored my new rain gauge onto a cement post, and it seems to be crooked."

"What does your rain gauge look like?" asked his friend, the math student.

"It's in the shape of a circular cylinder 8 centimeters in diameter with height-markings all around its sides. Its axis is only 3 degrees off-vertical, but this will affect the amount of rain entering the top, and besides, which height-marking should I use? The water-level will look tilted. I'm very discouraged about this whole business."

"Are you interested in measuring extremely light rains?" asked his friend.

"Not really. Anything less than a half-centimeter is too hard to measure accurately anyway, so I just record it as being a 'trace of precipitation'."

"I think I can help you," said the math student.

Tell the meteorologist how to correct the readings on his crooked rain gauge. [*Daniel A. Moran, Michigan State University.*]

1057. Dissect a regular pentagon into six pieces and reassemble the pieces to form three regular pentagons whose sides are in the ratio 2:2:1. [*D. M. Collison, Anaheim, California.*]

Solutions

A Simple Closed Curve **January 1977**

1006. A simple closed curve in the plane encloses a region R of area A. There is a point P in the interior of R such that every line through P intersects R in a line segment of length d. Find the greatest lower and least upper bounds for A. Are there curves where these bounds are attained? [*G. A. Heuer, Concordia College, Moorhead, Minnesota.*]

Editor's Comment. All solutions had some deficiency in discussing the least upper bound. These deficiencies ranged from asserting the correct value for the least upper bound, but only showing the value was an upper bound, to asserting that the least upper bound was not attained, but not supplying a convincing proof.

Solution determining the greatest lower bound and the least upper bound: Taking P as origin in polar coordinates and denoting the radius vector $r(\theta)$ to the boundary of R by r, and assuming the boundary is piecewise smooth, we have

$$A=\frac{1}{2}\int_0^{\pi} r^2 d\theta+\frac{1}{2}\int_0^{\pi}(d-r)^2 d\theta=\frac{1}{2}\int_0^{\pi}[d^2-2r(d-r)]d\theta$$

$$=\frac{1}{2}\int_0^{\pi}\left[2\left(r-\frac{d}{2}\right)^2+\frac{d^2}{2}\right]d\theta, \quad 0<r<d. \tag{1}$$

For a minimum A we must have $r=d/2$. The minimum $A=\pi d^2/4$ is attained by the circle with center P and radius $d/2$.

From the second integral above, we see that $A<\int_0^{\pi} d^2 d\theta/2=\pi d^2/2$. The upper bound is the *least* upper bound, as can be seen by the following.

Example. Let n be a large odd integer and let ϵ be a small positive number. Define

$$r(\theta)=\begin{cases} d-\epsilon, & \frac{\pi}{2n}<\theta<\pi-\frac{\pi}{2n} \\ \epsilon, & \pi+\frac{\pi}{2n}<\theta<2\pi-\frac{\pi}{2n} \\ \frac{d}{2}+\frac{d-2\epsilon}{2}\sin n\theta, & \text{otherwise, on } [0,2\pi]. \end{cases}$$

HOWARD EVES
University of Maine at Machias

Solution showing that the least upper bound is not attained (excerpted by the editor): Suppose there is a curve which attains the least upper bound. Then (1) shows that for such a curve, $r(\theta)$ would be 0 or d on the interval $[0,\pi]$. But this is inconsistent with the conditions that $r(\theta)+r(\theta+\pi)=d$ and that P, the pole of our polar coordinate system, is an interior point of R.

J. M. STARK
Lamar University

Other partial solutions by S. Floyd Barger, David & Santo Diano, Alan D. Frank, Donald C. Fuller, Marguerite Gerstell, Michael Goldberg, Eli L. Isaacson, Larry O. Olson, P. J. Pedler, Joseph Silverman, Richard Troxel, Samuel Weinberg, and the proposer.

A True Result **March 1977**

1009*. Let x, y, and n be positive integers and define $f(x)=x^2-x+41$ and $g(y)=y^2-y+68501$. Prove or disprove that n divides $g(y)$ for some y if and only if n divides $f(x)$ for some x. [*Sidney Kravitz, Dover, New Jersey.*]

Solution: The statement is true. The following identities are useful in establishing the proof: $f(x)=f(1-x)$ and $41^2f(x)=g(41x-20)$. Using the second identity it is obvious that when n divides $f(x)$ for some x, n will also divide $g(y)$ for $y=41x-20$.

Now suppose that n divides $g(y)$ for some y. Two cases can be considered: case I when $(n,41)=1$ and case II when $n=41^ik$ where $(k,41)=1$ and $i>0$.

Case I. Since $(n,41)=1$, there exists an $x>0$ such that $41x-20\equiv y \pmod n$. Then $g(y)\equiv g(41x-20)\equiv 41^2f(x) \pmod n$. Therefore if n divides $g(y)$, it also divides $f(x)$.

Case II. Since $n=41^ik$ where $i>0$, the following statement shows that n divides $g(y)$ only if $y=41s-20$ for some integer s:

$$g(y)=y^2-y+68501\equiv y^2+40y+400\equiv(y+20)^2 \pmod{41}.$$

Now $g(y)=g(41s-20)=41^2f(s)$ and $(k,41)=1$; so if n divides $g(y)$, then k divides $f(s)$. Since $f(s)=f(1-s)$ and either $(s,41)=1$ or $(1-s,41)=1$, the proof will be complete when it is shown that if k divides $f(r)$ with $(r,41)=1$, then 41^ik divides $f(x)$ for some $x>0$ with $(x,41)=1$. This is done by induction on i. If $i=1$, by the Chinese Remainder Theorem, there exists a positive integer x with $x\equiv r \pmod k$ and $x\equiv 1 \pmod{41}$. Then k divides $f(x)$ since $f(x)\equiv f(r) \pmod k$, 41 divides $f(x)$, and $(x,41)=1$. Now assume that when k divides $f(r)$ with $(r,41)=1$ there exists a positive integer x with $(x,41)=1$ such that 41^jk divides $f(x)$. Since k divides $f(x)=x^2-x+41$ and $(k,41)=1$, it follows that $(x,k)=1$. Therefore there exists an integer $a<0$ such that $ax\equiv 1 \pmod{41^{j+1}k}$. It then follows that

$$f(1-41a)=41(41a^2-a+1)\equiv 41(41a^2-a^2x+a^2x^2)\equiv 41a^2(41-x+x^2) \pmod{41^{j+1}k}.$$

Therefore $41^{j+1}k$ divides $f(1-41a)$ and $(1-41a,41)=1$.

KAY DUNDAS
Hutchinson Community Junior College

Also solved by Roderick Caldwell, Clayton W. Dodge, Robert S. Fisk, Allan Wm. Johnson Jr., and James Lee Murphy.

Roots in Progression **March 1977**

1010. Prove that if the roots of a fourth degree polynomial are in arithmetic progression, then the roots of its derivative are also in arithmetic progression. [*Marius Solomon, student, University of Pennsylvania.*]

Solution: Let the roots be $a-3d/2, a-d/2, a+d/2, a+3d/2$. Using the relationships between the roots and the coefficients, we find that the equation, expressed as a monic polynomial, is

$$x^4-(4a)x^3+(6a^2-5d^2/2)x^2-(4a^3-5ad^2)x+(a^4-5a^2d^2/2+9d^4/16)=0.$$

Thus the roots of the derivative are the roots of the equation

$$x^3-(3a)x^2+(3a^2-5d^2/4)x-(a^3-5ad^2/4)=0.$$

The roots of this equation are $a-\sqrt{5}\,d/2, a, a+\sqrt{5}\,d/2$ as can be verified by using the relationships between the coefficients and the roots. Thus the roots of the derivative are also in arithmetic progression.

DANIEL S. FREED, student
New Trier West High School
Northfield, Illinois

Also solved by Peter Addor (Switzerland), Mangho Ahuja, John M. Atkins, George Berzsenyi, J. C. Binz (Switzerland), Clyde A. Bridger, John Brunn, D. Burton & J. Henderson, Robert H. Cornell, James Avery Davis &

William Yslas Velez, Jesse Deutsch, Santo M. Diano, Clayton W. Dodge, Ragnar Dybvik (Norway), Thomas E. Elsner, Gail Eisner, Win Emmons, Howard Eves, David Farnsworth, Alex G. Ferrer (Mexico), Robert S. Fisk, Marjorie A. Fitting, Charles D. Friesen, Donald C. Fuller, Howard Fulmer, Ralph Garfield, Marguerite Gerstell, Richard A. Gibbs, Michael Goldberg, C. Ann Goodsell, William E. Gould, M. G. Greening (Australia), Richard A. Groeneveld, David Hammer, Paul M. Harms, Kent Harris, George C, Harrison, J. D. Hiscocks (Canada), Dinh Thê' Hūng, Eli L. Isaacson, Steve Kahn, Geoffrey C. Kandall, Lew Kowarski, Sidney Kravitz, R. Laumen (Belgium), Jordan I. Levy, Graham Lord (Canada), John S. Maginnis, Jerry Metzger, James L. Murphy, Roger B. Nelsen, Leonard L. Palmer, F. D. Parker, P. J. Pedler, Bob Prielipp, Charles S. Rees, James J. Reynolds, Lawrence A. Ringenberg, Sally Ringland, J. Rue, Harold W. Schneider, Benjamin L. Schwartz. Roberta Silver, Roland Smith, J. M. Stark, David R. Stone, Philip Straffin, Steven Szabo, Gregory Testini (Thailand), Zalman Usiskin, H. R. van der Vaart, Edward T. H. Wang (Canada), August J. Weidner, S. Weinberger, Kenneth M. Wilke, Paul Y. H. Yiu, Ken Yocum, Qazi Zameeruddin (India), Gene Zirkel, and the proposer. Gibbs, Testini, and van der Vaart observed that the result is true if the roots of the polynomial are only symmetric about their mean.

An Old Dice Problem **March 1977**

1011. Is it possible to load a pair of dice so that the probability of rolling each possible sum is $1/11$? [*Richard A. Gibbs, Fort Lewis College.*]

Editor's Note: This is an old problem that we believe might be interesting to new readers.

Solution I: Label the dice A and B and let X_i be the probability of rolling i with die X. Suppose that each possible sum is equally likely, then denoting the probability of rolling sum a by $P(a)$, we have

$$P(2)=1/11=A_1B_1, \qquad P(12)=1/11=A_6B_6,$$

and

$$P(7)=1/11=A_1B_6+A_2B_5+A_3B_4+A_4B_3+A_5B_2+A_6B_1.$$

These equations imply $A_1B_1>A_1B_6$, and $A_6\,B_6>A_6B_1$. Hence the supposition is false; there can be no such dice.

P. MERKEY, student
Michigan Technological University

Solution II: It is impossible. Let a_i $(i=1,2,\ldots,6)$ be the probability that die A yields value i and b_i $(i=1,2,\ldots,6)$ be the probability that die B yields value i. Using probability generating functions yields

$$\left(\Sigma_{i=1}^{6}a_ix^{i-1}\right)\left(\Sigma_{i=1}^{6}b_ix^{i-1}\right)=\Sigma_{j=2}^{12}c_{j-2}x^{j-2}$$

where c_{j-2} is the probability that the sum on the two dice is j. For equality all c_{j-2} must equal $1/11$. But then

$$11\cdot\Sigma_{j=2}^{12}c_{j-2}x^{j-2}=p(x)=(1-x^{11})/(1-x).$$

Since $1-x^{11}$ has one real zero and ten non-real zeros, the polynomial $p(x)$ has ten non-real zeros. Hence $p(x)$ cannot be factored into the product of two polynomials of degree five with real coefficients since each such polynomial has at least one real zero.

MARTIN BERMAN
Bronx Community College

Also solved by Mangho Ahuja, J. Binz (Switzerland), David Farnsworth, Robert S. Fisk & Gary D. Peterson, Paul K. Garlick, L. V. Glickman (England), Michael Goldberg, Arnold Hammel, J. D. Hiscocks (Canada), Eli L. Isaacson, Jordan I. Levy, James C. McKim, Joseph Mercer, James L. Murphy, P. J. Pedler (Australia), James T. Sandefur, Philip D. Straffin, Paul H. Yiu (Hong Kong) and the proposer. Several solvers provided references to this problem and its generalizations. See problem E925, Amer. Math. Monthly, 57 (1951), 191–192 *and "Equally Likely Dice Sums do not Exist" by E. J. Dudewicz and R. E. Dann, American Statistician,* 26 (1972), 41–42.

REVIEWS

PAUL J. CAMPBELL, Editor
Beloit College

PIERRE MALRAISON, Editor
Control Data Corp.

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.

Roark, Anne C., *The mathematician who may be this century's greatest,* Chronicle of Higher Education 17:1 (5 Sept. 1978) 1, 10-11.

Biographical sketch of Charles Fefferman, 29-year-old mathematical prodigy, first winner of the $150,000 Waterman Award and most recently recipient of a Fields Medal at the Helsinki International Congress of Mathematicians. His most important achievements have been in Fourier analysis in three dimensions.

Edwards, Harold M., *Fermat's Last Theorem,* Scientific American 239:4 (October 1978) 104-122, 188.

An excellent survey of the attempts to resolve the famous assertion by Fermat that $x^n + y^n = z^n$ is impossible in integers for $n > 2$. The most recent result, by Sam Wagstaff (Illinois), is that the assertion is true for all $n \leq 125{,}000$. This leads the author to conclude "...in a certain sense Fermat's Last Theorem is empirically true. If there is a solution...then the numbers in it are so inconceivably large that human beings will never be able to deal with them."

Mostow, George D., *Fields Medals (I): Relating the continuous and the discrete,* Science 202 (20 Oct. 1978) 297-298.

An account of the work of Gregory A. Margoulis on lattice subgroups of Lie groups.

Bass, Hyman, *The Fields Medals (II): Solving geometry problems with algebra,* Science 202 (3 Nov. 1978) 505-506.

Account of the work of Daniel Quillen in topology and algebraic K-theory.

Moser, Jürgen, *The Fields Medals (III): A broad attack on analysis problems,* Science 202 (10 Nov. 1978) 612-613.

An account of the work of Charles Fefferman in harmonic analysis and several complex variables.

Mumford, David and Tate, John, *The Fields Medals (IV): An instinct for the key idea,* Science 202 (17 Nov. 1978) 737-739.

An account of the work of Pierre Deligne in algebraic geometry.

Brams, Steven J. and Muzzio, Douglas, *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.

Tarjan, Robert Endre, *Complexity of combinatorial algorithms*, SIAM Review 20:3 (June 1978) 457-491.

A gem of an expository article, examining recent work, highlighting the mathematical tools used, and featuring the important results achieved. A large bibliography is included.

Gibbons, Jean D., Olkin, Ingram and Sobel, Milton, *Baseball competitions--are enough games played?* American Statistician 32:3 (August 1978) 89-95.

The current number of games played is found to be highly inadequate for the World Series, and only barely sufficient for the pennant race, in order to have a reasonable level of confidence that the best team wins.

Schaaf, William L., A Bibliography of Recreational Mathematics, V. 4, NCTM, 1978; xii + 172 pp, $9 (P).

Update of *Volume 3* (TR, May 1974), following the same organization, including primarily references between 1972 and 1977. Includes a complete chronological list (Dec. '56-Aug. '77) of Martin Gardner's column in *Scientific American*, and a supplementary glossary of recreational terms.

Gardner, Martin, *Mathematical games: Puzzling over a problem-solving matrix, cubes of many colors and three-dimensional dominoes*, Scientific American 239:3 (Sept. 1978) 20-30, 242.

Whether or not you found the "Instant Insanity" cubes boring, this article on the work of P.A. MacMahon will keep your interest. Working on generalized domino problems, he came to investigate the set of 30 different same-size cubes whose faces each bear a single color, with each one exhibiting the same six colors. The classic problem: given any one cube, find eight among the remaining 29 from which to build a 2x2x2 model whose 2x2 outer faces match the faces of the original cube and whose inside faces match (like dominoes).

Gardner, Martin, *Mathematical games: Puzzles and number-theory problems arising from the curious functions of ancient Egypt*, Scientific American 239:4 (October 1978) 23-30, 188.

The ancient Egyptians represented all proper fractions except 2/3 as sums of distinct unit fractions, and the Rhind papyrus exhibits the arithmetic they developed to deal with such a form of representation. Gardner notes the main results and open problems on various algorithms to generate representations. The splitting method he mentions was first proved to work for any fraction by R.L. Graham (Bell Labs) and R. Jewett (Hewlett-Packard) in 1964 (unpublished).

Smith, John Maynard, *The evolution of behavior*, Scientific American 239:3 (Sept. 1978) 176-194, 242.

Applies probability and game-theoretic models to account for the evolution of behavior. The goal is to develop a model in which the given behavior increases the fitness of the individual, so that there is no need to appeal to "the good of the species." Example behaviors are altruism and rituals in combat. Assuming altruism is genetically transmitted, a simple probabilistic model shows that even if the altruistic individual dies, the frequency of altruism genes increases.

Golubitsky, Martin, *An introduction to catastrophe theory and its applications*, SIAM Review 20:2 (April 1978) 352-387.

Basic theory, heuristic explanations, and applications to optics, buckling of beams, and convex conservation laws.

NEWS & LETTERS

MINIMAL MATHEMATICS FOR COLLEGE GRADUATES

The Committee on the Undergraduate Program in Mathematics (CUPM) of The Mathematical Association of America has established a Panel to consider the question: "What should every graduate of an American college or university know of mathematics?" It is hoped that the ultimate recommendations of this Panel will provide welcome guidance to colleges, universities, and such bodies as state boards of education, many of which are already actively considering such questions in the current wave of renewed interest in core curricula and general education.

The Panel, relying extensively on surveys of informed opinion, wishes to arrive at a list of minimum mathematical competencies for all college graduates, where "mathematical" is meant to include statistics, computing, etc., as well as mathematics in the narrow sense. Its report should contain, besides this list, a reasoned statement about why every college graduate should have acquired some understanding of mathematical thought; suggestions about courses in which the minimal competencies might be acquired; and general observations about such matters as interinstitutional coordination in furthering mathematical literacy.

The Panel has begun to collect information and opinions on the problem, and will welcome contributions from readers of this announcement. Facts about other local, regional, or national efforts (past or present) in the area of the Panel's charge, personal views about the general issue or specific aspects, and copies of or references to pertinent documents are among the things the Panel would be glad to receive. They may be sent to the Panel in care of its chairman, D. Bushaw, Department of Pure and Applied Mathematics, Washington State University, Pullman, WA 99164.

THE MISSING NEGATIVE

On p. 239 of the paper "A Property of 70" by Paul Erdös (this *Magazine*, September 1978, pp. 238-240), the key word "not" has been left out of the second sentence of the text. The sentence should read "If it is one of the a_k's, then...is not satisfied for $n > 289$."

Richard Friedlander
University of Missouri
St. Louis
Missouri 63121

FALSE COINS AGAIN

"61. The Counterfeit Coin. This problem came to light, apparently for the first time, in 1945, when it was contributed to the *Dial* in October of that year by Dwight A. Stewart, RCA, Camden." --C.A. Graham, *Ingenious Mathematical Problems and Methods*, Dover, New York, 1959, p. 37. The author apparently refers to *The Graham Dial*, a publication of Graham Transmissions, Inc., which includes a problem section to which Murray Klamkin, Dick and Josephine Andree, L.R. Ford (Sr.), V. Thébault, Norman Anning, Viktors Linis, and others have contributed.

Roy Meyers
Ohio State University
Columbus
Ohio 43210

EQUILATERAL PENTAGONAL TILINGS

D.C. Hunt and I have recently shown that Doris Schattschneider's list (in "Tiling the Plane with Congruent Pentagons," this *Magazine*, January 1978, pp. 29-44) of *equilateral* convex pentagons which tile the plane is indeed complete.

That is to say, putting it as succinctly as I can, an equilateral convex pentagon tiles the plane if and only if it has two angles adding to 180°, or it is the unique equilateral convex pentagon with angles A, B, C, D, E satisfying $A + 2B = C + 2E = A + C + 2D = 360°$ ($A \cong 70.88°$, $B \cong 144.56°$, $C \cong 89.26°$, $D \cong 99.93°$, $E \cong 135.37°$).

It should be remarked that in obtaining this result we made no assumptions concerning periodicity of any tiling, yet it is a fact that every equilateral convex pentagon which tiles yields a periodic tiling.

We intend to submit our work for publication in the not-too-distant future. Anyone wanting copies of *Parabola*, V. 13, No's 1 & 2 (reference [16] in Schattschneider's paper) need only write to me.

M.D. Hirschhorn
University of New South Wales
P.O. Box 1, Kensington
NSW 2033, Australia

1978 MATHEMATICAL OLYMPIAD SOLUTIONS

THE 1978 U.S.A. MATHEMATICAL OLYMPIAD

The seventh U.S.A. Mathematical Olympiad took place on May 2, 1978, and the problems were published that same month in this column. The examination was prepared by a committee consisting of M.S. Klamkin (Chairman), C.C. Rousseau, T. Griffith.

The following sketches of solutions were adapted by Loren Larson from Samuel Greitzer's pamphlet "Olympiads for 1978." 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. Given that a, b, c, d, e are real numbers such that

$$a + b + c + d + e = 8,$$
$$a^2 + b^2 + c^2 + d^2 + e^2 = 16 .$$

Determine the maximum value of e.

Sol. From the Cauchy inequality

$$(a+b+c+d)^2 \leq (1+1+1+1)(a^2+b^2+c^2+d^2) ,$$

it follows on substitution that

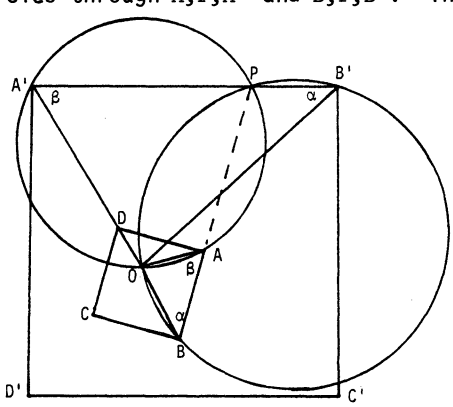
$$(8-e)^2 \leq 4(16-e^2) .$$

From this we find that $0 \leq e \leq 16/5$, and this upper bound is attained when $a = b = c = d = 6/5$.

2. $ABCD$ and $A'B'C'D'$ are square maps of the same region of a country but drawn to different uniform scales and are superimposed as shown below. Prove that there is only one point O on the small map which lies directly over a point O' of the large map such that O and O' each represent the same place of the country. Also, give a Euclidean construction (straight edge and compass) for O.

Sol. There is one fixed point by the Brouwer Fixed-Point Theorem. For uniqueness, construct even smaller similar squares, each within the previous set of squares, and apply the Bolzano-Weierstrass Theorem to the resulting set of nested regions.

To locate the fixed point O, let AB intersect $A'B'$ at point P, and construct circles through A,P,A' and B,P,B'. The



circles intersect at the desired point O. To see this, first note that the angles labeled α in the figure are equal, as are the angles labeled β (both supplementary to $\angle PAO$). Thus $\Delta OA'B' \sim \Delta OAB$. Therefore a rotation about O through the angle AOA' places the square $ABCD$ with its sides parallel to $A'B'C'D'$ and a dilation with ratio $A'B'/AB$ makes them coincide.

3. An integer n will be called *good* if we can write $n = a_1 + a_2 + \ldots + a_k$ where $a_1, a_2, \ldots, a_k$ are positive integers (not necessarily distinct) satisfying $1/a_1 + 1/a_2 + \ldots + 1/a_k = 1$. Given that it is known that the integers 33 through 73 are good, prove that every integer ≥ 33 is good.

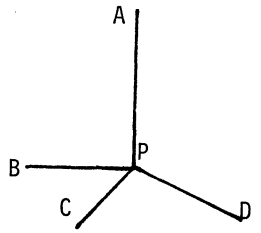
Sol. If $(a_1, a_2, \ldots, a_k)$ is a partition of n that is good, then $2n + 8$ and $2n + 9$ are good, as given in the partitions $(4,4,2a_1,2a_2,\ldots,2a_k)$ and $(3,6,2a_1,2a_2,\ldots,2a_k)$. Let P_n denote the statement: all the integers $n, n+1, \ldots, 2n+7$ are good. We are given P_{33} is true, and the previous sentence shows that P_n implies P_{n+1}. The result follows by induction.

4. (a) Prove that if the six dihedral angles (i.e., angles between pairs of faces) of a given tetrahedron are congruent, then the tetrahedron must be regular. (b) Must the tetrahedron be regular if five dihedral angles are congruent?

Sol. (a) Construct a sphere with its center at one vertex--say vertex A. The three planes forming the trihedral angle at A will intersect the sphere in a spherical triangle which is equiangular and therefore equilateral. It follows that the three face angles at A are equal; label this angle α. In a similar way, we find the three face angles at vertex B, vertex C, and vertex D are equal--label these angles β, γ and δ respectively.

The sum of all face angles is 4 x 180° or 720°. This means that $3(\alpha+\beta+\alpha+\delta) = 720°$, or $\alpha + \beta + \gamma + \delta = 240°$. Since $\alpha + \beta + \gamma = 180°$, it follows that $\delta = 60°$. Similarly $\alpha = \beta = \gamma = 60°$, and therefore all the faces are equilateral triangles. Hence the tetrahedron is regular.

(b) Consider four concurrent vectors equally inclined to each other in space.



Raise P along PA to decrease angles BPC, CPD, DPB by the same small amount. Now tilt AP around P symmetrically with respect to PB and PC so as to decrease angles APB and APC to the new angle BPC. Then $\angle APD > \angle BPC$. Now construct planes perpendicular to the segments PA, PB, PC, PD and their endpoints, to obtain a non-regular tetrahedron with five congruent dihedral angles.

5. Nine mathematicians meet at an international conference and discover that among any three of them, at least two speak a common language. If each of the mathematicians can speak at most three languages, prove that there are at least three of the mathematicians who can speak the same language.

Sol. Assume that at most two mathematicians speak a common language. Then each mathematician can speak to at most three others, one for each language s/he knows. Suppose A can speak only with B, C and D. Consider any E different from A, B, C, D. E can speak with at most three others (these may be among B, C, D). Since there are nine mathematicians, there remains another, say F, who cannot speak with A or E. This is a contradiction, since among any three, at least two speak a common language.

THE 1978 INTERNATIONAL MATHEMATICAL OLYMPIAD

The September issue of Mathematics Magazine contained the problems from the 20th International Mathematical Olympiad, which took place in Bucharest in July 1978. Here are sketches of solutions to these olympian problems for readers who wish aid or confirmation. These too were adapted by Loren Larson from "Olympiads for 1978."

1. m and n are natural numbers with $n > m \geq 1$. In their decimal representations, the last three digits of 1978^m are equal, respectively, to the last three digits of 1978^n. Find m and n such that $m + n$ has its least value. *(Cuba)*

Sol. From what is given we know that $1978^m(1978^{n-m}-1) = 2^3 \cdot 5^3 N$. It follows that 2^3 divides 1978^m. Since $1978 = 2 \cdot 989$, we must have $m \geq 3$. Also, 5^3 must divide $1978^{n-m}-1$, which is to say, $1978^{n-m} \equiv 1 \pmod{125}$. 1978 is relatively prime to 125, so we know that $(1978)^{\phi(125)} \equiv 1 \pmod{125}$. Since $\phi(125) = 100$, it follows that $n - m$ divides 100.

Also, $1 = 1978^{n-m} \equiv (-2)^{n-m} \pmod 5$. But $(-2)^2 \equiv 4 \equiv -1 \pmod 5$, and thus $(-2)^4 \equiv 1 \pmod 5$. Hence $n - m$ is a multiple of 4. It follows that either $n - m = 4$, $n - m = 20$, or $n - m = 100$. It is straightforward to check that $1978^4 \equiv 6 \pmod{125}$ and $1978^{20} \equiv 26 \pmod{125}$. Hence $n - m = 100$, $m = 3$, and $n + m = 106$.

2. P is a given point inside a given sphere and A, B, C are any three points on the sphere such that PA, PB and PC are mutually perpendicular. Let Q be the vertex diagonally opposite to P in the parallelopiped determined by PA, PB and PC. Find the locus of Q. *(USA)*

Sol. Place the sphere S of radius R with its center at the origin of a cartesian coordinate system whose axes are parallel to PA, PB, PC. Let the coordinates of P be (x,y,z) and let $PA = a$, $PB = b$, $PC = c$. Then A, B, C have coordinates $(x+a,y,z)$, $(x,y+b,c)$, $(x,y,z+c)$. Since A, B, C lie on S, we have:

$$(x+a)^2 + y^2 + z^2 = R^2$$
$$(y+b)^2 + x^2 + z^2 = R^2$$
$$(z+c)^2 + x^2 + y^2 = R^2 .$$

Adding,

$$(x+a)^2+(y+b)^2+(z+c)^2+2(x^2+y^2+z^2) = 3R^2.$$

However,

$$(x+a)^2 + (y+b)^2 + (z+c)^2 = (OQ)^2 ,$$

so $OQ^2 = 3R^2 - 2OP^2$. Therefore OQ^2 is independent of a, b, c and Q lies on a sphere Σ whose radius is $\sqrt{3R^2-2OP^2}$ and center is O. This proves necessity.

For the converse, we will need the following lemma: If $ABCD$ is a rectangle, with diagonals AC and BD, and O is any point, then $OA^2 + OC^2 = OB^2 + OD^2$. This can be established using vectors or as a consequence of Stewart's Theorem.

Let Q lie on sphere Σ. Then $OQ^2 = 3R^2 - 2OP^2$. The sphere with diameter PQ intersects sphere S ($OP<R<OQ$) in a circle. Let C be a point on this circle, and consider the rectangle $PCQX$, where X and C are endpoints of one diagonal and P and Q endpoints of the other. From the lemma, $OP^2 + OQ^2 = OC^2 + OX^2$. Hence $OX^2 = OP^2 + (3R^2-2OP^2) - R^2 = 2R^2 - OP^2$, so X lies outside sphere S.

Construct plane α perpendicular to PC and containing P. The point X lies in α, while the plane α intersects S in a circle K. P lies within K and X lies outside K. Therefore the circle with diameter PX lying in α intersects K. Let B be a point of intersection. It lies on S and PB and PC are perpendicular.

Finally, consider the rectangle $PBXA$, where A and B are endpoints of a diagonal. From the lemma, $OP^2 + OX^2 = OA^2 + OB^2$, or $OA^2 = OP^2 + (2R^2-OP^2) - R^2 = R^2$. That is, A also lies on sphere S. We have now shown that for every point on the sphere Σ, there are three mutually perpendicular lines PA, PB, PC with A, B, C lying on sphere S.

3. The set of all positive integers is the union of two disjoint subsets

$$\{f(1),f(2),\ldots f(n),\ldots\}$$
$$\{g(1),g(2),\ldots g(n),\ldots\} ,$$

where

$$f(1) < f(2) < \ldots < f(n) < \ldots ,$$
$$g(1) < g(2) < \ldots < g(n) < \ldots ,$$

and

$$g(n) = f(f(n)) + 1 \text{ for all } n \geq 1 .$$

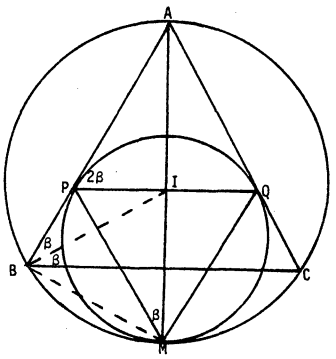
Determine $f(240)$. *(Gt. Britain)*

Sol. Let $A = \{f(k)\}$, $B = \{g(k)\}$, $k = 1,2,\ldots$. Since $f(1)$ is less than every other $f(k)$, $k \neq 1$, and also less than every $g(k)$, and since every positive integer is in either A or B, it follows that $f(1) = 1$.

The n-th integer not in A is $g(n) = f(f(n)) + 1$. Since $f(f(n))$ is in A, there are $n - 1$ elements not in A and less than $f(f(n))$, and therefore $f(f(n)) = f(n) + n - 1$. Using this recursion formula we can compute $f(n)$ for various n, thus: $f(2) = 3$, $f(3) = 4$, $f(4) = 6$, $f(6) = 9$, $f(9) = 14$, $f(14) = 22$, $f(22) = 35$, $f(35) = 56$, $f(56) = 90$, $g(35) = 91$, $f(57) = 92$, $f(92) = 148$, $f(148) = 239$, $f(239) = 386$, $g(148) = 387$, $f(240) = 388$.

4. In triangle ABC, $AB = AC$. A circle is tangent internally to the circumcircle of triangle ABC and also to sides AB, AC at P, Q respectively. Prove that the midpoint of segment PQ is the center of the incircle of triangle ABC. *(USA)*

Sol. Consider the figure below.



AM is a diameter which bisects $\angle A$, $\angle PMQ$, and also line segment PQ. Let $\angle APQ = 2\beta$. Then $\angle ABC = 2\beta$. $\angle APQ$ and $\angle PMQ$ are equal, since both equal half of arc PQ. Hence $\angle PMQ = 2\beta$ and $\angle PMI = \beta$. Since $\angle ABM = 90° = \angle MIP$, $BMIP$ is inscribable. Therefore $\angle PBI = \angle PMI = \beta$. Hence BI bisects $\angle ABC$, and, since AI bisects $\angle BAC$, I must be the incenter of ΔABC.

5. Let $\{a_k\}$ $(k = 1,2,3, \ldots, n, \ldots)$ be a sequence of distinct positive integers. Prove that, for all natural numbers n,

$$\sum_{n=1}^{n} \frac{a_k}{k^2} \geq \sum_{k=1}^{n} \frac{1}{n}. \qquad \textit{(France)}$$

Sol. Suppose that for some r and s, $r < s$ and $a_r > a_s$. Then

$$\frac{a_r}{r^2} + \frac{a_s}{s^2} > \frac{a_s}{r^2} + \frac{a_r}{s^2}.$$

It follows that $\sum_{1}^{n} a_k/k^2$ will be least when $a_1 < a_2 < \ldots < a_k$. In this case, $a_k \geq k$ and the conclusion follows easily.

6. An international society has its members from six different countries. The list of members contains 1978 names, numbered 1, 2, ... , 1978. Prove that there is at least one member whose number is the sum of the numbers of two members from his own country, or twice as large as the number of one member from his own country. *(Netherlands)*

Sol. Assume the statement is false. One country, say A, will then have at least 1978/6, or 330, members. Write these as $a_1 < a_2 < \ldots < a_{330}$. The 329 differences $a_{330} - a_1$, $a_{330} - a_2$, ... , $a_{330} - a_{329}$ are not in A, for if $a_{330} - a_i = a_j$ then $a_{330} = a_i + a_j$ which is contrary to our supposition.

Among these 329 differences, there will be at least 66 $(329/5 = 65\frac{4}{5})$ which belong, say, to B. Again, the differences $a_{i_{66}} - a_{i_j}$ are in neither A nor B. There are 65 of these.

In the same manner, we find 17 differences $(65/4 = 16\frac{1}{4})$ that will all be from C, and therefore 16 differences in neither A, B, nor C. Next six in D, $(16/3 = 5\frac{1}{3})$, leaving five differences not in A, B, C or D; then three (5/2) from E, leaving two from F. Since the difference of these two elements cannot belong to A, B, C, D, E or F, we have a contradiction of the supposition that the assertion is false.

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.

Albig, David, *Radford College.*
Anderson, Sabra, *University of Minnesota, Duluth.*
Andrews, George E., *Pennsylvania State University.*
Anton, Howard, *Drexel University.*
Appel, Kenneth I., *University of Illinois, Urbana.*
Barbeau, Edward J., *University of*

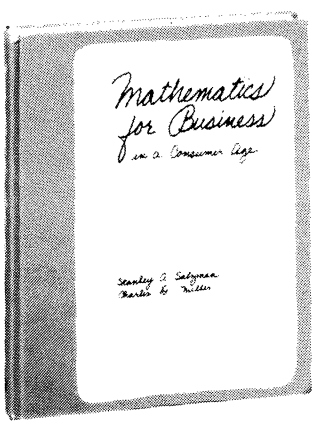
Toronto.
Berndt, Bruce C., *University of Illinois at Urbana-Champaign.*
Bernhart, Frank, *Bloomsburg, Pennsylvania.*
Boas, Mary L., *DePaul University.*
Boas, Ralph P., *Northwestern University.*
Borden, Robert S., *Knox College.*
Brams, Steven J., *New York University.*
Buck, R. Creighton, *University of Wisconsin, Madison.*
Burton, David M., *University of New Hampshire.*
Busenberg, Stavros, *Harvey Mudd College.*
Cable, Charles A., *Allegheny College.*
Callahan, James, *Smith College.*
Cantwell, John, *St. Louis University.*
Chakerian, G.D., *University of California, Davis.*
Chess, Karin, *University of Wisconsin, Eau Claire.*
Coleman, Courtney, *Harvey Mudd College.*
Corzatt, Clifton E., *St. Olaf College.*
deLong, Howard, *Trinity College.*
England, James, *Swarthmore College.*
Feroe, John A., *Vassar College.*
Finkbeiner, Daniel, *Kenyon College.*
Finney, Ross, *UMAP.*
Fisk, Robert, *Pacific Lutheran University.*
Friedberg, Stephen, *Illinois State University.*
Gaffney, Matthew, *University of Massachusetts, Boston.*
Getu, Seymour, *Howard University.*
Goldstein, Larry J., *University of Maryland, College Park.*
Gould, Henry, *West Virginia University.*
Grabiner, Judith V., *California State College.*
Guy, Richard, *University of Calgary.*
Haigh, Tom, *St. John's University.*
Halmos, Paul R., *University of California, Santa Barbara.*
Hanes, Harold, *Earlham College.*
Henriksen, Melvin, *Harvey Mudd College.*
Herzog, John O., *Pacific Lutheran University.*
Heuer, Charles V., *Concordia College.*
Heuer, Gerald, *Concordia College.*
Hilton, Peter, *Battelle Seminar and Study Programs.*
Hoekema, David, *St. Olaf College.*
Insel, A.J., *Illinois State University.*
Isaacs, Rufus, *Baltimore, Maryland.*
Janke, Steven, *Colorado College.*
Kamerud, Dana, *St. Louis University.*
Kirch, Allen, *Macalester College.*
Kleber, Richard S., *St. Olaf College.*
Koch, John, *Wilkes College.*
Krieger, Henry, *Harvey Mudd College.*
Laugwitz, Detlef, *Technische Hochschule, Darmstadt.*
Leach, Ronald, *Howard University.*
Loomer, Norman, *Ripon College.*
Lundgren, Richard, *Allegheny College.*
Luxemburg, W.A.J., *California Institute of Technology.*
Malkevitch, Joseph, *York College (CUNY).*
Meyer, John S., *Cornell College.*
Michaels, Brenda, *Brooklyn, New York.*
Mills, George H., *St. Olaf College.*
Montgomery, Susan, *University of California, Los Angeles.*
Moore, David S., *Purdue University.*
Moore, Thomas, *Marietta College.*
Morris, J. Richard, *Virginia Commonwealth University.*
Nau, Richard W., *Carleton College.*
Ness, Linda, *Institute for Advanced Study.*
Nummela, Eric, *St. Cloud State University.*
Philipp, Walter, *University of Illinois, Urbana.*
Pourciau, Bruce Hunter, *Lawrence University.*
Prichett, Gordon, *Hamilton College.*
Rapoport, Anatol, *University of Toronto.*
Renfrow, J. Thomas, *Beloit College.*
Riordan, John, *Rockefeller University.*
Roberts, Wayne, *Macalester College.*
Roeder, David W., *Colorado College.*
Rosen, David, *Swarthmore College.*
Sabharwal, Chaman, *St. Louis University.*
Savage, Thomas R., *St. Olaf College.*
Schneider, Dennis M., *Knox College.*
Schuster, Seymour, *Carleton College.*
Schwartz, Benjamin L., *Vienna, Virginia.*
Smullyan, Raymond, *Herbert H. Lehman College, (CUNY).*
Snow, Wolfe, *Brooklyn College (CUNY).*
Starr, Norton, *Amherst College.*
Sterling, Daniel J., *Colorado College.*
Stewart, Ian, *University of Warwick.*
Straffin, Philip, Jr., *Beloit College.*
Strang, Gilbert, *Massachusetts Institute of Technology.*
Suppes, Patrick, *Stanford University.*
Tuchinsky, Philip M., *Dearborn Heights, Michigan.*
Ulmer, Milton, *Carleton College.*
Vanden Eynden, Charles, *Illinois State University.*
van Iwaarden, John L. *Hope College.*
Weinstein, Michael L., *Passaic, New Jersey.*
Wolf, Frank, *Carleton College.*
Young, Frank, *Knox College.*

Class Encounters of the Best Kind

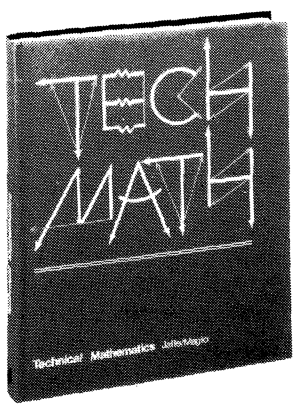
All Available Now

Business Mathematics
Second Edition
Miller / Salzman
384 pp., illus., paperback
Instructor's Guide with Tests / tutorial
Audio Tape cassettes

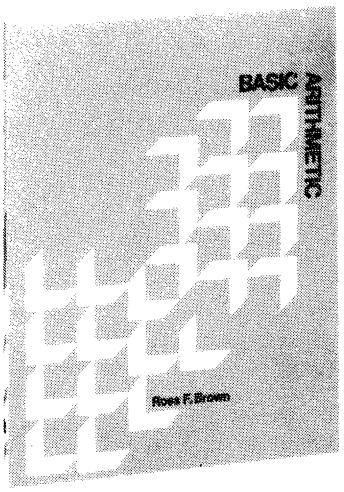
Mathematics with Applications In the Management, Natural, and Social Sciences
Second Edition
Lial / Miller
640 pp., illus., hardbound
Instructor's Guide with Tests / Study Guide with Computer Problems



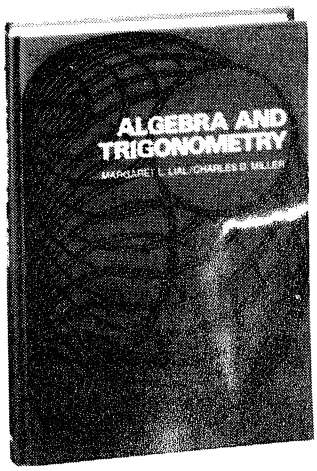
Mathematics for Business
In a Consumer Age
Salzman / Miller
448 pp., illus., hardbound
Instructor's Guide with Tests / tutorial
Audio Tape cassettes



Technical Mathematics
Jaffe / Maglio
576 pp., illus., hardbound
Instructor's Guide with Tests



Basic Arithmetic
Brown
416 pp., illus., paperback
Test Bank with Answer Key



Lial / Mill...

Algebra and Trigonometry
576 pp., hardbound

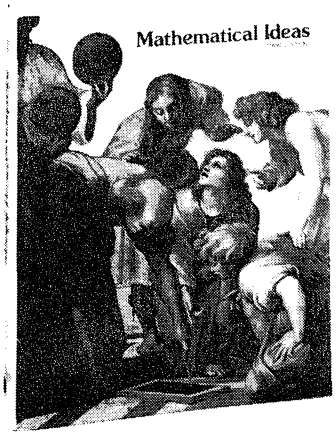
Trigonometry
320 pp., hardbound

College Algebra
Second Edition
384 pp., hardbound

Intermediate Algebra
Second Edition
432 pp., hardbound

Beginning Algebra
Second Edition
334 pp., hardbound

All with Instructor's Guides, Test Banks, Solutions Guides, Study Guides. Tutorial Audio Tape cassettes also available for **Beginning Algebra** and **Intermediate Algebra.**



Mathematical Ideas
Third Edition
Miller / Heeren
512 pp., illus., hardbound
Instructor's Guide with Tests and additional topics

For further information write
Jennifer Toms, Department SA
1900 East Lake Avenue
Glenview, Illinois 60025

Scott, Foresman and Company

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

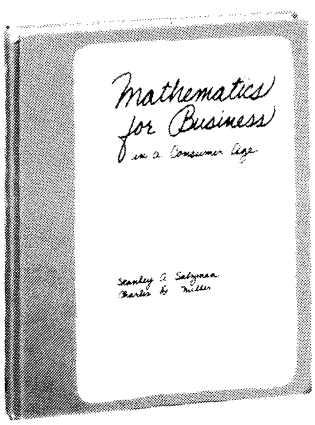
MATHEMATICS MAGAZINE VOL. 51, NO. 5, NOVEMBER, 1978

Class Encounters of the Best Kind

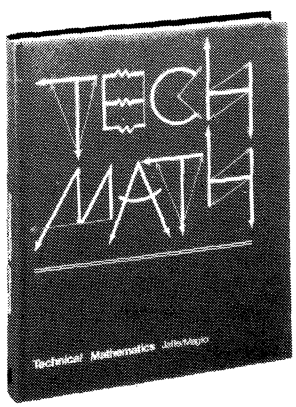
All Available Now

Business Mathematics
Second Edition
Miller / Salzman
384 pp., illus., paperback
Instructor's Guide with Tests / tutorial
Audio Tape cassettes

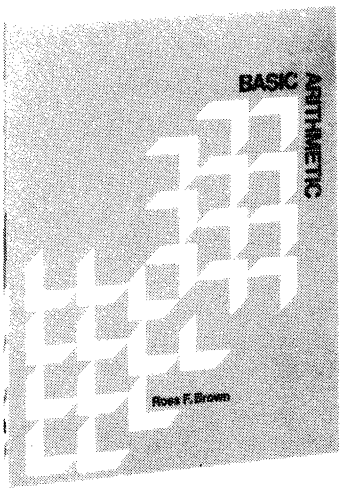
Mathematics with Applications In the Management, Natural, and Social Sciences
Second Edition
Lial / Miller
640 pp., illus., hardbound
Instructor's Guide with Tests / Study Guide with Computer Problems



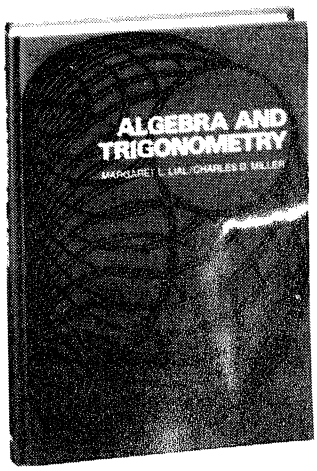
Mathematics for Business
In a Consumer Age
Salzman / Miller
448 pp., illus., hardbound
Instructor's Guide with Tests / tutorial
Audio Tape cassettes



Technical Mathematics
Jaffe / Maglio
576 pp., illus., hardbound
Instructor's Guide with Tests



Basic Arithmetic
Brown
416 pp., illus., paperback
Test Bank with Answer Key



Lial / Mill...

Algebra and Trigonometry
576 pp., hardbound

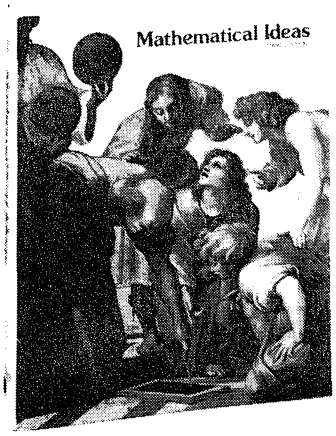
Trigonometry
320 pp., hardbound

College Algebra
Second Edition
384 pp., hardbound

Intermediate Algebra
Second Edition
432 pp., hardbound

Beginning Algebra
Second Edition
334 pp., hardbound

All with Instructor's Guides, Test Banks, Solutions Guides, Study Guides. Tutorial Audio Tape cassettes also available for **Beginning Algebra** and **Intermediate Algebra.**



Mathematical Ideas
Third Edition
Miller / Heeren
512 pp., illus., hardbound
Instructor's Guide with Tests and additional topics

For further information write
Jennifer Toms, Department SA
1900 East Lake Avenue
Glenview, Illinois 60025

Scott, Foresman and Company

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

MATHEMATICS MAGAZINE VOL. 51, NO. 5, NOVEMBER, 1978