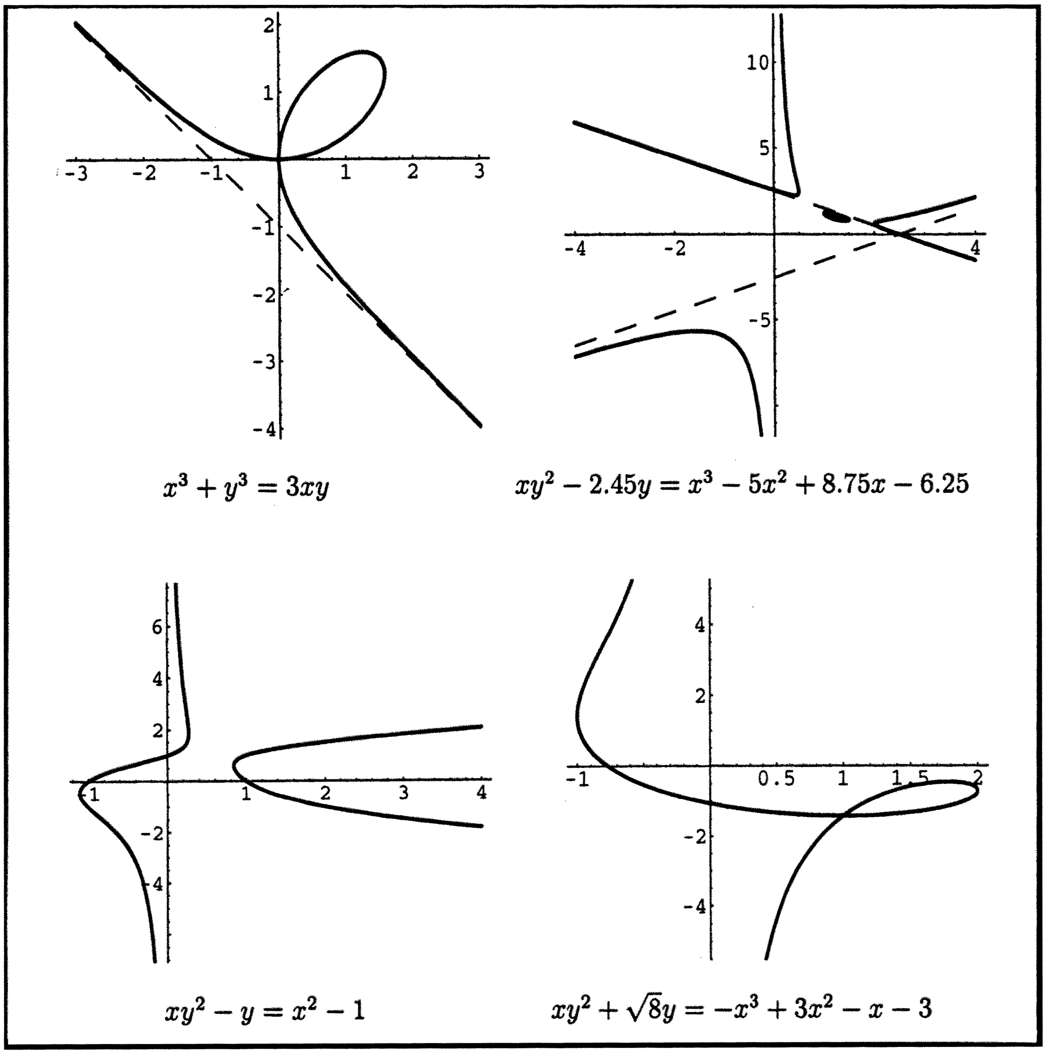
Vol. 72, No. 3, June 1999

MATHEMATICS MAGAZINE



$x^3 + y^3 = 3xy$

$xy^2 - 2.45y = x^3 - 5x^2 + 8.75x - 6.25$

$xy^2 - y = x^2 - 1$

$xy^2 + \sqrt{8}y = -x^3 + 3x^2 - x - 3$

- Asymptotes, Cubic Curves, and the Projective Plane
- Sample Calculus
- Full Rank Matrix Factorization

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

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

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

Send new manuscripts to Paul Zorn, Editor, Department of Mathematics, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098. Manuscripts should be laser-printed, with wide line-spacing, and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit three copies and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Copies of figures should be supplied on separate sheets, both with and without lettering added.

AUTHORS

Richard D. Neidinger is professor of mathematics at Davidson College in North Carolina, his home since 1984. He received his B.A. from Trinity University in San Antonio and his M.A. and Ph.D., specializing in functional analysis, from the University of Texas at Austin. He enjoys using computers to expose interesting topics that are accessible to undergraduates, particularly in numerical methods, fractals, and chaos. His Pólya Award-winning article in the May 1989 *College Mathematics Journal* inspired Walter Spunde's calculus reform project. Their collaboration began at an APL conference in St. Petersburg, Russia.

Jeff Nunemacher entered Oberlin Conservatory as an organ major but, discovering where his true talents and interests lay, he emerged in 1970 with a B.A. in mathematics from Oberlin College. Five years of a broad, pure mathematical training at Yale University produced a Ph.D. with specialization in several complex variables. While there, he lived with a group of mostly mathematician graduate students, who, under the name of N. Bourbaki, Jr., developed a reputation as excellent sabbatical house sitters. He has worked at the University of Texas at Austin, Kenyon College, Oberlin College, and now teaches a mixture of pure and applied mathematics and computer science at Ohio Wesleyan University. This article is part of his continuing endeavor to understand classical algebraic geometry.

Patrick L. Odell graduated with a B.S. in mathematics from the University of Texas in 1952, then did graduate study at Oklahoma State University, ending with a doctorate in 1962. His research has been primarily in multivariate statistical analysis and matrix theory, with applications in space science and engineering. He began his teaching career at the University of Texas at Austin, and then taught at Texas Tech University and the University of Texas at Dallas. He is now professor of mathematics at Baylor University.

Robert Piziak received his B.A. in mathematics from Amherst College in 1964, and his M.A. and Ph.D. degrees from the University of Massachusetts, Amherst, in 1966 and 1969, respectively. He has taught at the University of Florida, Centre College in Kentucky, and, since 1981, at Baylor University. His current research interests are in matrix theory, especially generalized inverses.

Walter Spunde, a graduate of the University of Queensland, spent some years at Queen's (Kingston, Ontario) before joining the University of Southern Queensland, where he has been teaching introductory mathematics for the past 26 years. His early research was on laminar fluid flow, irregular eigenfunction expansions, and spectral operators. The extension of Richard Neidinger's 1989 exposition of automatic differentiation to lists, from which high order derivatives could be graphed without formulae, led to the ideas in *Sample Calculus*. Over the past ten years, Walter Spunde has held several major Commonwealth (of Australia) development grants for improving mathematics instruction, exploiting these ideas and their web-based implementation.

Vol. 72, No. 3, June 1999

MATHEMATICS MAGAZINE

The *MATHEMATICS MAGAZINE* (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.

The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $16 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $68 per year.

Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

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

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

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

Printed in the United States of America

ARTICLES

Sample Calculus

WALTER G. SPUNDE
University of Southern Queensland
Toowoomba, QLD 4350
Australia

RICHARD D. NEIDINGER
Davidson College
Davidson, NC 28036

Introduction

A computational approach to introducing calculus offers new opportunities to focus on the basic operations and their properties. This approach uses lists or vectors of numerical values to describe functions, in contrast with other numerical viewpoints that seek one value. The usual introductions of the derivative at a point and the definite integral can be refocused on the derivative function and an indefinite integral function. The fundamental theorem can be seen as a relationship between functions, without relying on symbolic antiderivatives or abstract theoretical notation. In this computing environment, there can be a clear correspondence between classical Leibniz notation and list manipulation operations. With the numerical list approach, arc length and other line integrals are easy to motivate and approximate. The new perspective opens the door to fresh discussions of fundamental concepts, such as variable, function, continuity, differentiability and the indefinite integral.

The numerical list manipulation approach is built on a partition of a domain interval, say $x = (x_0, x_1, x_2, \dots, x_n)$. Ordinary arithmetic and function evaluation will be applied pointwise, so that $y = f(x) = (f(x_0), f(x_1), f(x_2), \dots, f(x_n))$. Various algorithms will be applied to y to produce lists of values representing (or approximating) derived functions at the same, or related, domain points. To encourage this mode of thinking, we use the word *sample* to refer to any of the numerical lists. Unlike "partition," the word "sample" connotes a representation or, more precisely, one possible set of representative values. Students should think about what is represented and where. Algorithms on samples can then bridge the gap between operations on numbers and the more advanced notions of algebraic operations on variables and functions.

While this numerical list approach to calculus can be implemented in many computing environments (see `http://www.maa.org/pubs/mm_supplements/index.html`), in this paper we use Matlab to illustrate one such environment. No familiarity with Matlab is required and only one quirk of syntax is relevant throughout: a period precedes an operator, as in `.*`, `./`, or `.^`, whenever the operation is to be applied to (respective) components of numerical vector argument(s).

As an introductory example, we numerically differentiate and integrate function f whose values are computed from $f(x) = 1/x$ on $[1, 2]$, using a very rough sample. We begin by partitioning the interval $[1, 2]$ into 5 subintervals of width 0.2 and evaluating the function $1/x$ at these six sample points. The action of most Matlab input lines should be clear from the output.

```
Input        Output
x=(1:0.2:2)  x=1.0000 1.2000 1.4000 1.6000 1.8000 2.0000
y=1./x       y=1.0000 0.8333 0.7143 0.6250 0.5556 0.5000
```

In subsequent sections, finer samples (i.e., ones with more points) will be viewed as graphs, instead of directly viewing the numerical values. Of course, all machine plots are really just views of samples.

To calculate the rates of change, or slopes, use the Matlab operator `diff`, which returns the differences between consecutive points in a sample.

```
dx=diff(x)  dx=0.2000 0.2000 0.2000 0.2000 0.2000
dy=diff(y)  dy=-0.1667  -0.1190  -0.0893  -0.0694  -0.0556
dy./dx      ans=-0.8333 -0.5952 -0.4464 -0.3472 -0.2778
```

These last values are difference quotients of the form $(y_{i+1} - y_i)/(x_{i+1} - x_i)$. They are the average rates of change over each subinterval and, hence, can be associated with the average point (midpoint) of each subinterval. Together, they form a sample (approximate, but good enough for plotting) of the derivative function at the midpoints.

A midpoint sample can also be used to compute the classic midpoint Riemann sum approximation of the definite integral.

```
xm=(1.1:0.2:1.9)       xm=1.1000 1.3000 1.5000 1.7000 1.9000
ym=1./xm               ym=0.9091 0.7692 0.6667 0.5882 0.5263
ym.*dx                 ans=0.1818 0.1538 0.1333 0.1176 0.1053
sum(ym.*dx)            ans=0.6919
Sy=cumsum([0,ym.*dx])  Sy=0 0.1818 0.3357 0.4690 0.5866 0.6919
```

The approximate values, `Sy`, for an antiderivative function, are found by replacing the summation by a cumulative, or partial, sum starting from zero. FIGURE 1 shows a plot of these values against the *original* sample x (simply `plot(x, Sy)` in Matlab) along with an actual graph of the natural logarithm.

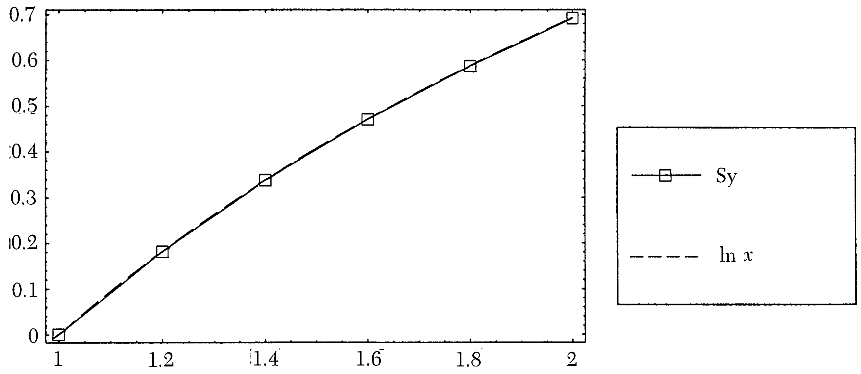


FIGURE 1
Antiderivative of $1/x$ approximated by a cumulative sum.

Leibniz's original writings show his use of the symbols d and $\int$ as operations on lists of numbers in his work on finite series. (Specific references, from **[1]** and **[2]**, are discussed in **[13]**.) Leibniz recognized (what we call) a cumulative sum as the sequence whose differences returned the original sequence. His conception of a curve as a sequence of points infinitely close together led naturally to the fundamental theorem of calculus. Leibniz could have proceeded along the lines of this paper, if computing machinery could have implemented his calculating program. However, work on finite series (operations on lists of numbers) evolved somewhat differently into a wide body of literature known as finite difference calculus (see, e.g., **[7]** or **[5]**). While this sounds equivalent, finite difference calculus concentrates on Δy and Σy but omits the role of dx (or Δx) which is usually normalized to 1. Notation for list operations, such as componentwise division or cumulative sum, was never standardized; hence, we use a modern executable notation (Matlab) and define specific operation names below.

General calculus on samples

Five numerical list operations, along with the ability to plot curves from these lists, are sufficient to allow convenient and general implementation of this method. Table 1 describes these five operations on an interval with endpoints a and b, some number of subintervals n, and a numerical list `x` $= (x_0, x_1, x_2, \dots, x_n)$. The web site `http://www.maa.org/pubs/mm_supplements/index.htm` includes definitions of these operations for J, Matlab, Mathematica, Maple, APL, and on the TI-85 and HP-48G graphics calculators. The remainder of this article assumes that such Matlab definitions (mfiles) make these operations executable, e.g., `S(x)` calls `cumsum([0,x])`.

TABLE 1 List Operations

Operation	Result
`sample(a,b,n)`	$(x_0, x_1, x_2, \dots, x_n)$, where $x_i = a + i(b-a)/n$
`d(x)`	$(x_1 - x_0, x_2 - x_1, \dots, x_n - x_{n-1})$
`mids(x)`	$((x_0 + x_1)/2, (x_1 + x_2)/2, \dots, (x_{n-1} + x_n)/2)$
`sum(x)`	$x_0 + x_1 + x_2 + \cdots + x_n$
`S(x)`	$(0, x_0, x_0 + x_1, \dots, x_0 + x_1 + \cdots + x_n)$

For any domain sample `x` and corresponding sample of function values `y`, a derivative sample at `mids(x)` is given by `d(y)./d(x)` and an antiderivative sample at `x` is `S(mids(y).*d(x))`. The operation names were chosen for pedagogical reasons, emphasizing the correspondence of these algorithmic operations with Leibniz's notation. The operations `d` and `mids` decrease the number of points in a sample from $n+1$ (one for each endpoint) to n (one for each interval). The operations `sum` and `S` usually apply to a list of n values (one area for each interval); `S` returns a list of $n+1$ values (area up to each endpoint).

For a typical example, evaluate $y = 0.25 + (2 - 0.7x + 0.04x^2)\cos^2(2.5 + 1.4x)$ at 101 sample points (100 intervals) over $[0, 10]$. FIGURE 2 displays the Matlab code to create this numerical list and the corresponding derivative and antiderivative samples, along with the graphical output. Using this plot, one can consider common calculus questions such as when the derivative is positive, negative, or zero and why the

Input for Figure 2

```
x=sample (0, 10, 100);
y=.25+(2 - .7*x+.04*x.^2)...
   .*(cos(2.5+1.4*x)).^2;
Dy=d(y)./d(x);
Sy=S(mids(y).*d(x));
plot(x, y, mids(x), Dy, x, Sy)
```

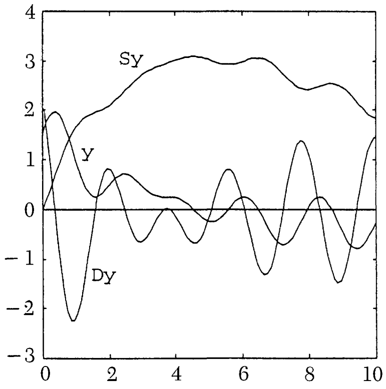


FIGURE 2
A function sample with resulting derivative and antiderivative.

antiderivative is increasing over the first half but not the second. Zooming in on the graph shows why a better estimate of extreme point locations is obtained from the derivative than from the function itself. All of this can be done without ever computing a symbolic representation.

Natural questions arise from this approach. What sample is fine enough? Do finer samples approach ideal functions? Can we find exact formulas for the ideal functions? Students understand that if a sample is too rough, it can distort the graph. Experimentally, they can double the number of sample points and see if the graphs change. They quickly accept that finer samples approach ideal functions; maybe too quickly, since this can motivate deep discussions of limits, continuity, and differentiability. One way to focus on these ideas is to introduce functions with singularities, cusps, or piecewise definitions. For typical functions, students can find it satisfying to find the exact formulas and may be surprised that an elementary antiderivative formula does not always exist. For the particular example above, numerical experimentation and exact symbolic formulae both verify that the plot shown is visually indistinguishable from the true function, derivative, and antiderivative.

The notation S(mids(y).*d(x)) encompasses a couple of algorithmic choices in the numerical integration. First, this is a trapezoidal rule approximation as opposed to the midpoint Riemann sum in the introduction. The difference is that we have used mids(f(x)) as opposed to f(mids(x)) for a representative y-value on each interval (where f is assumed to operate pointwise on the samples). The advantage is certainly not accuracy but rather that an algorithm for values of function f is not required to form mids(y). One may work directly with experimental data or other functions known only from tables. Second, we are approximating one particular antiderivative by the choice of zero as the starting point for the cumulative sum. If x is a partition of interval $[a, b]$, then S creates a sample whose value is zero at the starting endpoint a; i.e., S is approximating $\int_a^x f(t)\,dt$. To form an antiderivative with value C at the starting endpoint, one could begin an accumulation with C (for example, in Matlab use cumsum([C,mids(y).*d(x)])), but we prefer the suggestive notation of simply adding the scalar to each sample value with S(mids(y).*d(x))+C. Such subtleties are not presented to students at the beginning. Indeed, the purpose is to initially avoid complicated symbolic integral or summation notations and to develop intuition from the operations on numerical list examples. This can lead to a motivated discussion of these ideas.

The derivative approximation `d(y)./d(x)` is straightforward; simply a list of slopes over each sample interval. It's important to interpret these as values at the midpoints of each interval. This results in a decent numerical approximation, the central difference formula, with step-size equal to half the interval width.

The fundamental theorem

Continuing with the sample `Sy` from FIGURE 2, differentiate the integral by forming `DSy=d(Sy)./d(x)`. Plotting (not shown here) this sample of slopes `DSy` against `mids(x)` would reveal what looks like the original function values `y`. A zoom would show values at midpoints instead of the endpoints of the original sample. This relationship can be verified by considering how `d` and `S` are related. For any sample `m` $= (m_1, m_2, \dots, m_n)$, we have `d(S(m))=d(`$(0, m_1, m_1 + m_2, \dots, m_1 + m_2 + \cdots + m_n)$`)=m`. Thus `d(Sy)=mids(y).*d(x)`, and `d(Sy)./d(x)` returns `mids(y)`. In general, the values of `DSy`, the derivative of the integral, will be exactly the representative values of y chosen for each interval in the computation of the integral `Sy`. The plot of these representative y values against `mids(x)` approximates the original plot of `y` vs. `x`. While this is certainly not a proof of the fundamental theorem for continuous functions, it can whet the appetite. The difference between consecutive points in `Sy` is an analog of the difference between integral from a to $x + h$ and the integral from a to x.

The integral of the derivative may arouse even more wonder in students, since it will probably not return the original y values. To integrate the sample `Dy` from FIGURE 2, recall that `Dy` gives the average slope over each interval and so may be regarded as representative of the slope values over the intervals. Thus it is unnecessary to compute `mids(Dy)`; simply compute a Riemann sum by `S(Dy.*d(x))`. FIGURE 3 shows the resulting sample `SDy` which appears to be a vertical shift of the original `y`. They are different antiderivatives of the same (derivative) function and so

Input for Figure 3 (cont. from Figure 2)

```
SDy=S(Dy.*d(x));
plot(x, y, x, SDy)
```

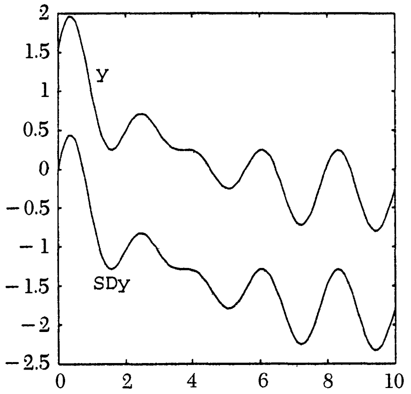


FIGURE 3
The antiderivative of the derivative of sample y.

differ by a constant. This may be verified experimentally by plotting SDy + y_0, or, in Matlab notation, SDy+y(1). Analytically, Dy=d(y)./d(x), so that

$$\texttt{S(Dy.*d(x))=S(d(y))=S((}y_1-y_0, y_2-y_1,\ldots,y_n-y_{n-1}\texttt{))}$$
$$=(0, y_1-y_0, y_2-y_0,\ldots,y_n-y_0)=\texttt{y}-y_0$$

Standard functions

Sample calculus can be used to approximate and study the standard transcendental functions using only arithmetic operations. From samples of related variables, it is a trivial matter to graph different functions by changing the independent and dependent variables. A couple of new sample calculus features also appear.

One way to compute the natural logarithm function using only arithmetic operations is to numerically approximate the formal definition $\ln(x)=\int_1^x 1/t\,dt$. Such an antiderivative function is obtained from a domain sample that must start at 1. The function on [1, 4] is straightforward, beginning with a sample xR of [1, 4]. To also get function values to the left of 1, use a second sample such as xL = (1, 0.95, 0.90, 0.85, . . . , 0.05) where each value in d(xL) is negative. FIGURE 4 displays the integration over both samples. The plot shows the logarithm along with the transpose of the coordinates to obtain the exponential function.

Input for Figure 4

```
xR=sample(1, 4, 20);
yR=S(mids(1./xR).*d(xR));
xL=sample(1,.05, 19);
yL=S(mids(1./xL).*d(xL));
plot(xR, yR, xL,yL, yL,xL, yR,xR)
grid on
```

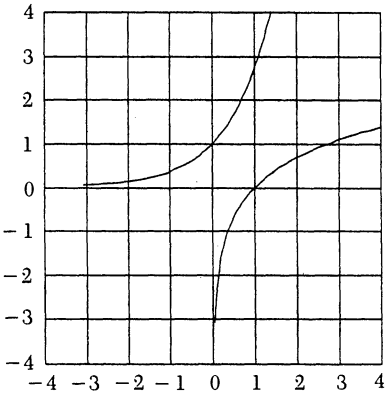


FIGURE 4
Logarithm and exponential functions from samples of $1/x$.

Definitions of trigonometric functions usually begin by defining radian measure as the length of an arc of the unit circle. Unfortunately, a radian measure is only computed by proportionality to the measurement of an angle in degrees and not by the coordinates of points. Computing arc lengths is a challenge that can be conquered directly with sample calculus. If x and y are corresponding samples of coordinates on the unit circle, then the Pythagorean theorem is enough to show that the distance between consecutive points is given by ds=sqrt(d(x).^2+d(y).^2). If the x and y samples progress counterclockwise around the upper semicircle, then sum(ds) is a direct approximation of π. Keeping a cumulative sum of these with t=S(ds) approximates radian measure for every angle in the sample. The trigonometric

Input for Figure 5

```
x=sample(1, -1, 100);
y=sqrt(1-x.^2);
ds=sqrt(d(x).^2+d(y).^2);
t=S(ds);
plot(t,x, t,y)
```

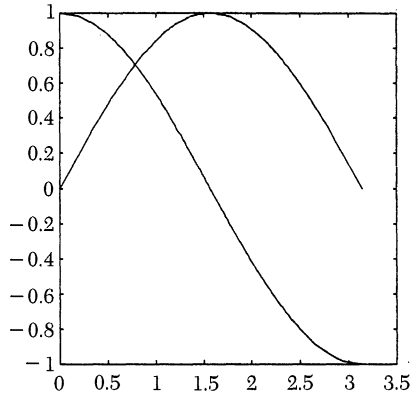


FIGURE 5

Sine and cosine from a sample of $y = \sqrt{1-x^2}$.

functions are then just a matter of which coordinate to plot against which independent. FIGURE 5 shows x against t to form cosine and y against t for sine. Plotting t against x would show the arccos and other algebraic combinations or inverses could also be obtained.

Calculus students often encounter these theoretical definitions but numerically evaluate them at only a few well-chosen points, if any. The direct implementation using samples rewards students with plots of the functions. Of course, the built-in trigonometric and exponential functions are used in all applications.

Parametric curves, integration by parts, and line integrals

Any parametric curve may be manipulated by partitioning the independent variable, say time t, and producing corresponding samples of x and y. For example, samples for the curve

$$x = (3t - t^2)\sin(e^{-t}); \quad y = e^{\cos(3t-t^2)}; \quad 0 \le t \le 3$$

are defined and plotted in FIGURE 6.

Questions about arc length and speed can be directly approximated by using the sample increments dt, dx, dy, and ds. Again, the key approximation is ds= sqrt(dx.^2+dy.^2). The arc length parameter s is simply the integral of ds and the speed is given by $\frac{ds}{dt}$. FIGURE 7 shows the resulting samples s and ds./dt as functions of t, but it would be just as easy to view x or y as functions of s or any other desired combination.

Traditionally, the arc length parameter is seldom used for computation; instead, the differential is replaced by $ds = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt$. While this yields theoretically exact answers, many examples (as above) require a numerical approximation for the arc length integral, so one might as well use the discrete ds approximations. On the other hand, finding the exact answers can sometimes be a challenging exercise that

Input for Figure 6

```
t=sample(0, 3, 100);
x=(3*t-t.^2).*sin(exp(-t));
y=exp(cos(3*t-t.^2));
plot(x, y)
```

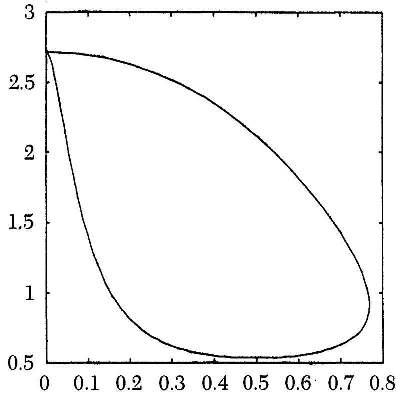


FIGURE 6
A parametric curve traversed clockwise.

Input for Figure 7 (cont. from Figure 6)

```
dt=d(t);
dx=d(x);
dy=d(y);
ds=sqrt(dx.^2+dy.^2);
s=S(ds);
plot(t, s, mids(t), ds./dt)
```

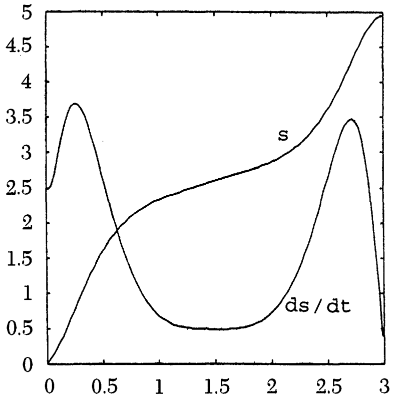


FIGURE 7
Arc length parameter and speed, approximated numerically.

follows the more straightforward numerical approximation. For example, perform the computations of this section for one arch of the cycloid where $x = t - \sin t$ and $y = 1 - \cos t$, and then find exact formulas by symbolic manipulation. Numeric approximations can be used to conjecture exact values.

The area inside the loop of the parametric curve in Figure 6 can be computed directly from the x and y samples. We form the usual Riemann sum from this unusual source of data points where the x increments are not uniform (not even uniform in sign) and the y values are not an explicit function of x.

```
sum(mids(y).*dx)                    ans=1.0548
```

This approximates a line integral ($\int_C y\,dx$) for the area inside the loop. Specifically, as points in the x sample increase from 0 to the maximum, the dx values are positive and the sum computes the area under the top half of the curve. In the rest of the sample, x values decrease so that the dx values are negative and the sum computes the negative of the area under the bottom half of the curve. Similar interpretation shows that sum(mids(x).*dy) approximates the negative of the area inside the loop. While the area approximation can be implemented in a first course, later reflection on the line integrals can introduce the more general Green's theorem.

In general, switching the roles of x and y amounts to integration by parts. For any x and y samples, sum(mids(y).*dx)+sum(mids(x).*dy) simplifies to $x_n y_n - x_0 y_0$, where subscript n indicates the last tabulation point. A symbolic computing environment will do this simplification for symbolic samples. Using plots of numeric samples, observe that the cumulative sum, S(mids(y).*dx) differs from x.*y − S(mids(x).*dy) by a constant, specifically $x_0 y_0$. This is verified numerically for our example by:

```
Sy=S(mids(y).*dx);
intByParts=x.*y-S(mids(x).*dy);
max(abs(Sy-intByParts))      ans=1.3323e−015
```

Samples can also be used to directly approximate line integral formulas without reformulation in terms of the parameter. The classic example is work accomplished along a curve by a varying force field $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$. If samples P and Q are components of the force at every tabulation point, then mids(P).*dx+mids(Q).*dy is a sample representing the increments of work done in taking a particle from each tabulation point to the next. The sum of these terms over all elements of the partition gives the total work done. A cumulative sum returns the work done up to every tabulation point, which may be plotted as a function of t or s or even against x or y. Such values for work are useful in discussing the potential for conservative fields. For example, consider $P = xy^2$ and $Q = yx^2 + y^2$ along the parameterized curve of FIGURE 6. FIGURE 8 plots the resulting work against t and also against y.

Input for Figure 8 (cont. from 6 and 7)

```
P=x.*y.^2;
Q=y.*x.^2+y.^2;
W=S(mids(P).*dx+mids(Q).*dy);
plot(t,W, y, W)
```

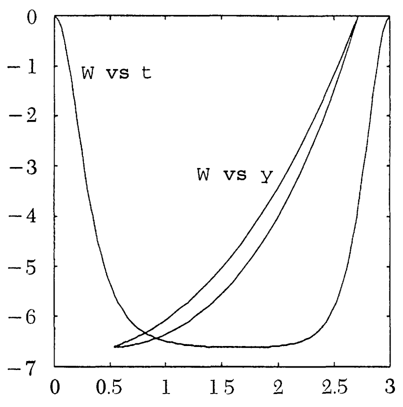


FIGURE 8
Work of force $xy^2\mathbf{i} + (yx^2 + y^2)\mathbf{j}$ along the curve in FIGURE 6.

Derivative rules on samples

Although preceding derivative samples have been approximations, there is an intriguing way to use samples and exact derivative rules to find derivative values. In typical texts, derivative rules are usually established and stated for values of component functions at a point; then the rules are used, almost exclusively, on symbolic expressions. However, the numerical values at a point, or a whole sample of points, may be used directly without forming the symbolic composition. This is the basic idea behind a valuable numerical method known as *automatic (or computational) differentiation* (see **[8]**, **[10]**, and **[3]**). Here, a simple example shows how samples for functions u, v, u' and v' can be combined numerically into a sample of derivative values of $y = u/v$, where $u = \arctan(x)$ and $v = x^4 + 1$. The input in FIGURE 9 is what is called a *code list* in automatic differentiation. Each component function has corresponding derivative values with a suffix p, as in up for u-prime. The output is exact in the sense that the formulas are perfect and values are limited only by machine floating point precision.

Input for Figure 9

```
x=sample(-3, 3, 100);
u=atan(x);
v=x.^4+1;
y=u./v;
up=1./(1+x.^2);
vp=4*x.^3;
yp=(up.*v-u.*vp)./(v.^2);
plot(x,y, x,yp)
```

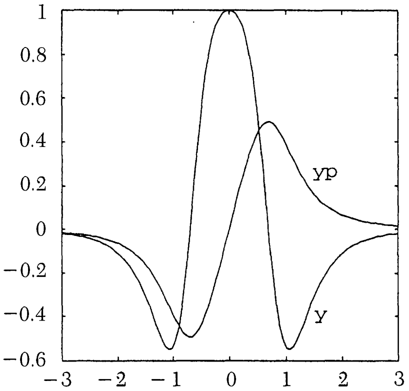


FIGURE 9
Exact derivative values by numeric combination.

Students now have three ways to compute derivative values and they can verify that all yield indistinguishable graphs. First, the above method simply lists component functions and derivative rules that are combined numerically. The second method, instinctive in most calculus courses, is to form the symbolic expression for y' and evaluate this function at each sample point. In this example,

$$y' = \frac{1}{(x^2+1)(x^4+1)} - \frac{4x^3 \arctan x}{(x^4+1)^2}.$$

The values should be identical, although different evaluation order may introduce a tiny machine precision error.

```
    ypcomp=1./((x.^2+1).*(x.^4+1))...
              -((4*x.^3).*atan(x))./(x.^4+1).^2;
    max(abs(yp-ypcomp))                ans= 2.2204e-016
```

Such symbolic manipulation is easy for small expressions but becomes more tedious and error-prone as expressions grow larger; it adds a significant burden of algebraic manipulation on top of the essential skill of identifying component functions and correctly applying the derivative rules. The third approach is to return to first principles and form the slopes `d(y)./d(x)` that can be plotted against `mids(x)`. This introduces an approximation error but, nevertheless, provides a check on the correctness of the work done with the other approaches.

```
Dy=d(y)./d(x);
max(abs(mids(yp) - Dy))          ans=0.0028
```

Conclusion

Sample calculus can enhance a student's perception of derivative and integral and, more fundamentally, function and variable. When x is a sample of numbers from a domain, the abstract symbol becomes a tangible choice from the many possible values of the variable. A functional relationship is understood from a list of range values corresponding to a sample of domain values. The algorithm that produces these values may or may not involve typical elementary function expressions. While experienced concretely, the function *is* this abstract relationship of domain and range values, a notion close to the theoretical definition in upper-level courses. This contrasts with usual approaches where students tend to focus on a function as a formula, such as $f(x)=\tan^2 x$. Whether called a formula, expression, or rule, this mode of thinking tends to identify a function with what, in computer jargon, is called a character string. Then, even on a computer, to differentiate or integrate means to perform a character string manipulation. Numerical methods usually focus on the derivative at a point or the definite integral, either as an introduction of the concepts or as a supplementary technique for accuracy. In the heart of the chapter, discussing derivative or antiderivative functions, there is a gap where numerical intuition is not used. Building small tables by repeated use of numerical methods can help bridge this gap (**[9]**, **[4]**), though this may be tedious and unrealistically small. Sample calculus fills the gap. The cumulative sum is an especially powerful idea that is missing from most calculus courses, though it is theoretically encountered under the guise of a sequence of partial sums. Euler's method for differential equations, introduced early in the calculus sequence by some authors **[11]**, also produces a sample of function values and is a natural part of a sample calculus program.

Sample calculus can naturally motivate otherwise purely theoretical discussions. Students consider domain questions and discover that points of non-differentiability produce artifacts in the approximate derivative just as a plot of function values will at points of discontinuity. In regions of uniform continuity, connecting points in a sufficiently fine sample will accurately capture a faithful representation. Students understand this intuitively and it can be formalized by saying that for any tolerable vertical fluctuation ϵ (e.g. pixel height), there is an interval width δ such that if $|x_1-x_2|<\delta$, then $|f(x_1)-f(x_2)|<\epsilon$. This δ is the theoretical answer to the practical question "what sample is fine enough?" The mean value theorem directly shows that the lists of difference quotients are, in fact, precise samples of the derivative function values at some unknown set of domain sample points (approximated by `mids(x)`). The mean value theorem for integrals says that there exists a sample of function values `yt` (approximated by `mids(y)`) so that `S(yt.*d(x))` would give precise antiderivative values at the original domain sample `x`.

Sample calculus essentially amounts to simple and relevant programming using `sample`, `d`, `mids`, `sum` and `S`. The power of this mini-language is seen in the breadth of applications in the above sections, all the way through line integrals and further. This programming effort reinforces calculus concepts, relates all of the computations, and is simple enough to allow for plotting and analyzing the results. If results are all you want, then commercial programs can provide superior numerics and graphics. As preparation for numerical analysis, perhaps students should program classic numerical formulas with loops. If the focus is on developing calculus concepts, then sample calculus has many advantages over these alternatives.

For many years, the first author has successfully used the approach described here as the basis of a first-year course (**[12]**). The techniques were developed using the numerical-list-based languages APL and J in computer labs. Extension to surface and volume integrals is discussed in **[14]**. The text **[6]** is an independent source that begins to explore the use of lists and first differences (our `d(y)`). Of course, symbolic skills and theoretical limits are still important. The intention of sample calculus is not to illustrate limiting processes but to help students grapple conceptually with the objects of calculus. In particular, classical skills in symbolic manipulation are not required in order to conceive of derivative and antiderivative functions.

REFERENCES

1. F. Cajori, *A History of Mathematical Notations*, Vol. II, Open Court Pub., Chicago, IL, 1929, p. 264 quote from *Leibnizens Math Schriften*, Vol. V (1858), p. 397.
2. J. M. Child, *The Early Mathematical Manuscripts of Leibniz*, Open Court Pub., London, UK, 1920.
3. A. Griewank and G. Corliss (eds.), *Automatic Differentiation of Algorithms: Theory, Implementation and Application*, SIAM, Philadelphia, PA, 1991.
4. D. Hughes-Hallett, et. al., *Calculus*, John Wiley, New York, NY, 1994.
5. C. Jordan, *Calculus of Finite Differences*, 2nd Ed., Chelsea, New York, NY, 1950.
6. D. LaTorre, et al., *Calculus Concepts, An Informal Approach to the Mathematics of Change*, Houghton Mifflin, Boston, MA, 1998.
7. K. S. Miller, *An Introduction to the Calculus of Finite Differences and Difference Equations*, Holt, New York, NY, 1960.
8. R. D. Neidinger, Automatic differentiation and APL, *College Math. J.* 20 (1989), 238–251.
9. A. Ostebee and P. Zorn, *Calculus from Graphical, Numerical and Symbolic Points of View*, Saunders College Pub., Fort Worth, TX, 1997.
10. L. B. Rall, The arithmetic of differentiation, this MAGAZINE 59 (1986), 275–282.
11. D. A. Smith and L. C. Moore, *Calculus: Modeling and Application*, Houghton Mifflin, Boston, MA, 1996.
12. W. G. Spunde, *Mathematics: A Numerical Approach to Algebra and Calculus*, Aline, Toowoomba, QLD, Australia, 1995.
13. W. G. Spunde, What Leibniz might have done... (to introduce calculus with a computer), *Proceedings First Asian Technology Conference in Mathematics*, Assoc. of Math. Educators, Singapore (1995), 511–520.
14. W. G. Spunde, Surface integrals in first year, *Proceedings 2nd Biennial Australian Engineering Mathematics Conference*, Institution of Engineers, Barton, ACT, Australia, 1996, 613–619.

Asymptotes, Cubic Curves, and the Projective Plane

JEFFREY NUNEMACHER
Ohio Wesleyan University
Delaware, OH 43015

1. Introduction

Among the most beautiful and naturally appealing mathematical objects are the various plane curves. It is a pity that our undergraduates encounter so few of them. One extensive class of curves, which played a role in the recent proof of Fermat's Last Theorem, is the class of cubic curves, i.e., curves defined by an equation $P(x, y) = 0$, where P is a polynomial in x and y of total degree three. Famous ancient examples, which can be explored using simple analytic techniques (see, for example, **[8]**), are the folium of Descartes $x^3 + y^3 - 3xy = 0$, the witch of Maria Agnesi $y(1 + x^2) = 1$, the cissoid of Diocles $y^2(2 - x) = x^2$, and the Fermat curve $x^3 + y^3 = 1$. Using a classical formula to express the roots of a cubic equation in terms of its coefficients, it is possible to solve for y in terms of x. The resulting functions are usually not easy to sketch by hand using standard methods of calculus, but software such as *Derive* or *Mathematica* makes it possible to study cubic curves in a computer laboratory. Such a study requires knowledge and care, since the packages often use formulas that select complex branches; hence they can miss certain real branches of the curve.

Newton studied the general cubic equation in two variables and classified irreducible cubic curves into 72 different species. Here *irreducible* means that the polynomial defining the curve does not factor as a product of lower degree polynomials. For example, the curve defined by $x^3 - x^2y - xy + y^2 = 0$ is reducible, since its defining polynomial factors as $(x^2 - y)(x - y)$; this curve is the union of a parabola and a straight line. In fact, Newton missed 6 species—according to his classification scheme (which allows affine coordinate changes), there are a total of 78 species. It makes a good project in a calculus course to explore the diversity of cubic curves and to reconsider Newton's classification. For suggestions on how this might be done making use of both classical algebra and modern technology, see **[6]**.

Newton's classification begins by studying the asymptotic behavior of cubic curves. This approach is very natural, since the behavior "at infinity" is a dominant feature of the shape of any curve. But asymptotes can be far from obvious on a computer-generated graph. The folium of Descartes $x^3 + y^3 - 3xy = 0$ is shown in FIGURE 1 together with its asymptote $x + y + 1 = 0$. If the line were not drawn, would you be confident that the folium has an asymptote, or of the asymptote's exact location? It is an

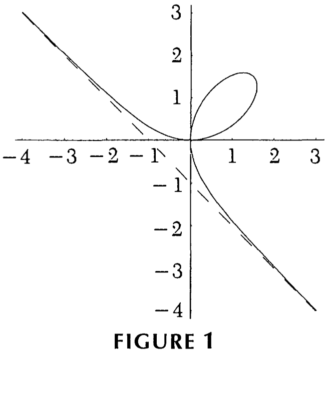


FIGURE 1

interesting and somewhat nonstandard exercise in calculus to verify that this line is asymptotic to the folium (see [8], p. 512). How does one find the asymptotes of a cubic curve or, more generally, of an algebraic curve of degree n? For such curves y is not given explicitly in terms of x so standard limiting techniques do not apply.

There is a theorem involving only polynomial algebra that almost answers this question. It provides a set of at most n lines that are the only possible asymptotes, and these lines will be asymptotic to the curve except in one rare situation. Special cases of this theorem used to be part of the standard repertoire of mathematical techniques that students learned when studying analytic geometry or the theory of equations. Fashions change, however, and this result is now encountered only in algebraic geometry, if at all. One reason for its disappearance is the unfortunate decline in interest and knowledge of geometry in high school and college. Another lies in the fact that the natural domain for thinking about asymptotes is the projective plane, which is often not studied at the undergraduate level. A third is that complications result from looking only at the real portion of an algebraic curve, which is best viewed as an object in complex space. In this article we present this theorem with some background in projective geometry and apply it to study cubic curves.

The theorem specifying the asymptotes of an algebraic curve is not easy to locate in current references, although at one time it must have been well known. A special case occurs explicitly in [7], pp. 8–10, but I have been unable to locate the general case in any reference. The nonsingular case of the theorem is treated nicely in [2]. An approach to finding asymptotes using only calculus can be found in [5], where it is applied to study various examples but not to find a theorem yielding all asymptotes. Classical methods and many examples with beautiful hand-drawn graphs can be found in [3]. Good modern references for many concrete facts in elementary algebraic geometry (but not this one!) are [1] and [4].

2. Statement of the theorem and applications

In this paper *asymptote* always refers to a line that is approached by points (x, y) on a branch of a curve as x or y becomes unbounded. This is the kind of asymptote encountered in a calculus course, but there it is almost always horizontal or vertical. A degenerate case, which we shall exclude from now on, occurs when the curve contains a line as a component, i.e., when the asymptote is actually part of the curve. Thus, in what follows, we assume that the defining polynomial $P(x, y)$ does not vanish identically on any line. The term *asymptotic direction* refers to a vector parallel to such a line so, in particular, the location of the line in the plane is not specified.

The following theorem specifies at most n candidate lines to be real asymptotes to a curve defined by $P(x, y) = 0$, where $P(x, y)$ is a polynomial of total degree n in the variables x and y. Such a curve is called an (affine) algebraic curve of degree n. Let us denote by $P_k(x, y)$ the sum of all terms occurring in $P(x, y)$ of total degree k. Then $P(x, y)$ can be expressed as $\Sigma_{k=0}^{n} P_k(x, y)$. The polynomials $P_k(x, y)$ are homogeneous of degree k; this means that, for any scalar λ, we have $P_k(\lambda x, \lambda y) = \lambda^k P_k(x, y)$. Polynomials such as $P_k(x, y)$, which are homogeneous of some degree k, are sometimes called *forms*.

Each vanishing form $P_k(x,y)$ factors over the complex numbers into a product of k linear factors, which are unique up to scalar multiple. The existence of such a factorization is a direct consequence of the Fundamental Theorem of Algebra. To see this, divide $P_k(x, y)$ by the term x^k, and replace the powers of y/x by powers of a new variable u. The resulting polynomial of one variable $p_k(u)$ has k complex roots and factors completely over the complex numbers. Suppose that $bu + a$ is one of the factors of $p_k(u)$. When u is replaced by y/x, this term gives rise to the factor $ax+by$ of $P_k(x,y)$. If $bu+a$ occurs to exact multiplicity m as a factor of $p_k(u)$, then $P_k(x,y)$ can be expressed as $(ax+by)^m Q(x,y)$, where $Q(x,y)$ is a homogeneous polynomial of degree $k-m$ with $Q(b,-a)\neq 0$.

MAIN THEOREM. *Suppose that $ax+by$ is a factor of the top degree form $P_n(x,y)$ of multiplicity m with a and b real. Let $r\leq m$ denote the largest integer with the property that there exist polynomials $Q_j(x,y)$ for $n-r+1\leq j\leq n$ satisfying the conditions:*

$$P_n(x,y)=(ax+by)^r Q_n(x,y),\ P_{n-1}(x,y)=(ax+by)^{r-1}Q_{n-1}(x,y),\quad \text{(A)}$$
$$\ldots,\ \textit{and finally } P_{n-r+1}(x,y)=(ax+by)Q_{n-r+1}(x,y).$$

Then associated with the factor $ax+by$ is a set of at most r possible asymptotes $ax+by=t_0$, where t_0 is a real root of the equation

$$t^r Q_n(b,-a)+t^{r-1}Q_{n-1}(b,-a)+\cdots+tQ_{n-r+1}(b,-a)+P_{n-r}(b,-a)=0.\quad \text{(B)}$$

All real asymptotes to the curve defined by $P(x,y)=0$ arise in this way as $ax+by$ ranges over the real linear factors of $P(x,y)$. If $r>1$ it may happen that some of these lines are spurious asymptotes.

Equation (B) has at most r roots, which may be complex or have multiplicity greater than one. Since $r\leq m$, the multiplicity of $ax+by$ as a factor of $P_n(x,y)$, the total number of possible asymptotes cannot exceed n. There is an actual asymptote associated with the factor $ax+by$ for each distinct real root except in the case discussed below.

The candidate asymptotes thus satisfy the equation

$$(ax+by)^r Q_n(b,-a)+(ax+by)^{r-1}Q_{n-1}(b,-a)+\cdots$$
$$+(ax+by)Q_{n-r+1}(b,-a)+P_{n-r}(b,-a)=0.\quad \text{(C)}$$

Condition (A) is a divisibility condition requiring that descending powers of $ax+by$ must be factors of the top r forms $P_k(x,y)$. Since r is the largest such integer, $ax+by$ does not further divide $P_{n-k}(x,y)$ or $Q_k(x,y)$ for some k between $n-r+1$ and n, i.e., $P_{n-r}(b,-a)$ or at least one of these $Q_k(b,-a)$'s is nonzero. The most common situation is covered by the following simpler result in which the candidate line is guaranteed to be an asymptote. It is a special case of the Main Theorem.

COROLLARY. *If $ax+by$ is a simple factor of $P_n(x,y)$, i.e., if $P_n(x,y)=(ax+by)Q_n(x,y)$ with $Q_n(b,-a)\neq 0$, then associated with this factor is the single asymptote to $P(x,y)=0$ defined by the equation*

$$(ax+by)Q_n(b,-a)+P_{n-1}(b,-a)=0.\quad \text{(D)}$$

We give some examples of the application of this theorem and its corollary.

Example 1. For the folium of Descartes $x^3+y^3-3xy=0$, which is displayed in FIGURE 1, the sole real linear factor of $P_3(x,y)=x^3+y^3$ is $x+y$ with $Q_3(x,y)=x^2-xy+y^2$. Here we have $r=m=1$ with $a=b=1$ and $P_2(x,y)=-3xy$. Thus the single asymptote is given by (D), namely, $(x+y)Q_3(1,-1)+P_2(1,-1)=0$, i.e., $3x+3y+3=0$.

Example 2. Consider the curve $xy^2-2.45y=x^3-5x^2+8.75x-6.25$; the coefficients have been chosen to present a "typical" nontrivial cubic curve. Here $P_3(x,y)=xy^2-x^3$ which factors as $x(y+x)(y-x)$ and $P_2(x,y)=5\ x^2$. Then the Corollary asserts that there are three asymptotes, namely, $x=0$, $y=x-2.5$, and $y=-x+2.5$. See FIGURE 2 for a graph of this curve and its asymptotes. One

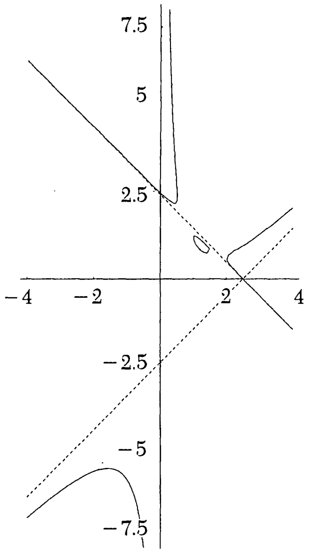


FIGURE 2

limitation of computer graphics is illustrated well by this figure. While the rightmost branch of this curve appears to be asymptotic to the line $x + y = 2.5$ from above, a more careful analysis shows that the curve actually crosses the line at the point $(2.5, 0)$ and is then asymptotic from below. This crossing becomes evident in the graph only with a tenfold increase in scale, so it would not be noticed unless it was deliberately sought.

Example 3. The quartic curve defined by $(x^3 - x)y = 1$ can be analyzed completely using techniques of calculus, since it is easy to solve for y in terms of x. But our theorem applies as well, with the following results. Here $P_4(x, y) = x^3 y$, so the factor x has multiplicity three. Also x^2 divides $P_3(x, y) = 0$ and x divides $P_2(x, y) = -xy$, so $r = 3$ with $a = 1$ and $b = 0$. From (B) we obtain the equation $t^3(-1) + t^2(0) + t(1) + 0 = 0$, i.e., $-t^3 + t = 0$. Solving this equation yields the three parallel candidate asymptotes $x = 0$, $x + 1$, and $x - 1 = 0$, which can easily be verified to be true asymptotes using limits. The other factor of $P_4(x, y) = x^3 y$ is y with multiplicity $m = 1$. Equation (D) now gives $y = 0$ as the only other asymptote. Thus this curve has three asymptotes in the direction $\langle 0, 1\rangle$ and one in the direction $\langle 1, 0\rangle$.

Example 4. Consider the parabola $x^2 - y = 0$, which we know has no asymptotes from basic analytic geometry. The factor x has multiplicity two in $P_2(x, y) = x^2$, but x does not divide $P_1(x, y) = -y$. Thus $r = 1$ with $a = 1$ and $b = 0$. We obtain from (C) the equation $x(0) - 1 = 0$. This linear equation does not describe a line, so we confirm there are no (finite) asymptotes to the parabola.

Example 5. Finally, let us analyze the curve $x^2y^2 - y^2 + 1 = 0$. It is easy to solve this equation for y in terms of x. We find that the curve is the union of the graphs of the two functions $f_{\pm}(x) = \pm 1/\sqrt{1 - x^2}$. Thus, using limits, we see that there are exactly two asymptotes, vertical ones at $x = \pm 1$. (See FIGURE 3 below.) Applying the theorem to this example, we see that the leading term $P_4(x, y) = x^2y^2$ has the two factors x and y, each of which yields possible asymptotes with $r = 2$ (since $P_3 = 0$). The factor x gives rise to the two asymptotes already noticed. The factor y (with $a = 0$ and $b = 1$) produces an equation (B) of the form $t^2\ 1 = 0$. Thus we obtain the candidate line $y = 0$, which is clearly not an asymptote for the curve. This example

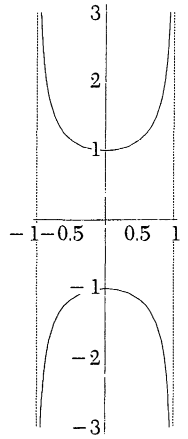


FIGURE 3

shows that the lines provided by the theorem may not be asymptotes if no real branch of the curve approaches them. We shall explore this situation in more detail below.

The conclusions of the theorem become simpler and more pleasing if we move to a space larger than $\mathbb{R}^2$. If we allow a and b to be complex numbers, equations (B) and (D) give a criterion for complex asymptotes, which are two-dimensional real planes in $\mathbb{C}^2$ (see **[3]**, p. 44). For instance, the circle $x^2 + y^2 = 1$ has $x + iy = 0$ and $x - iy = 0$ as complex asymptotes in the four-dimensional space $\mathbb{C}^2$, but these planes intersect the real plane $\mathbb{R}^2$ only at the origin. A different extension is relevant in the situation of Example 4. There the asymptotic equation does not describe a real line. A line is actually present, but it is the line at infinity in the projective plane $\mathbb{RP}^2$ (which contains $\mathbb{R}^2$). We shall return to this situation below when the necessary definitions have been made. A combination of these two extensions is necessary to explain Example 5. In a suitable space there are always exactly n asymptotes, if we interpret "asymptote" correctly and assign multiplicities to asymptotes $ax + by = t_0$ according to the number of times that t_0 occurs as a root when (B) is factored over $\mathbb{C}$. In $\mathbb{R}^2$, however, n provides only an upper bound for the number of finite real asymptotes.

For cubic curves, therefore, there can be no more than three asymptotes. In fact, cubic curves exist with 0, 1, 2, or 3 real asymptotes. The curve $yx(x-1) = 1$ has three asymptotes; $yx^2 = 1$ has two; the folium of Descartes has one, as we saw above; and the polynomial $y = x^3$ has no finite asymptotes.

Notice that for a curve $P(x, y) = 0$ of degree n the possible asymptotic directions $\langle b, -a\rangle$ are defined by the factors $ax + by$ of the top degree term $P_n(x, y)$. It is intuitively reasonable that the dominant term should determine the behavior of the curve at infinity, i.e., the asymptotic directions. However, where the asymptotes are situated in the plane is dependent on some of the lower order terms. In the case of a simple factor $ax + by$ only the next term $P_{n-1}(x, y)$ is relevant. This is the nonsingular case (in a sense to be defined below). For a factor of higher multiplicity $m > 1$, the next r terms in the homogeneous expansion are relevant to the asymptotic behavior, where r is defined by the theorem but is always bounded above by m.

3. Curves and asymptotes in the projective plane

It is natural to regard an asymptote to a curve as a line to which the curve is tangent "at infinity." This idea works very well for curves that are defined by the vanishing of a polynomial, since such curves extend in a simple fashion to a larger space including points at infinity. In addition to simplifying the hunt for asymptotes, the addition of points at infinity to $\mathbb{R}^2$ to form the projective plane $\mathbb{RP}^2$ has various other benefits. Newton's original classification of irreducible cubic curves into 72 species was criticized by later authors as being too complicated to be useful. By regarding the curve as lying in the projective plane and enlarging the group of allowable coordinate changes to include all projective transformations, it can be simplified into a classification containing just 5 different species. The points at infinity, which are sometimes called ideal points, are rather intuitive—we simply introduce one additional point at which all parallel lines in a given direction meet. Thus railroad tracks meet at the horizon, but perhaps less intuitively, they meet at the *same* point in both directions. This point is specified by any nonzero vector $\langle m, n\rangle$ parallel to the line. Notice that for points satisfying $ax+by+c=0$ with x or y large, only the ratio of a to b is important, rather than a, b, or c. This ratio is the information that a direction vector $\langle m, n\rangle$ contains.

There is a clean algebraic way to add points at infinity to $\mathbb{R}^2$. We consider the set of all 3-tuples (X, Y, Z) with real coordinates not all zero, and define an equivalence relation: $(X_1, Y_1, Z_1) \sim (X_2, Y_2, Z_2)$ if each triple is a scalar multiple of the other. These equivalence classes are defined to be the points in the real projective plane $\mathbb{RP}^2$. If the Z-coordinate of the 3-tuple (X, Y, Z) is nonzero, we may divide by it and obtain the equivalent 3-tuple $(X/Z, Y/Z,\ 1)$, which we identify with the Euclidean point (x, y), where $x = X/Z$ and $y = Y/Z$. Only when $Z = 0$ is this not possible, and it is these points which are the points at infinity. Notice that on the line $ax + by + c = 0$ the points escape to infinity as Z approaches 0, so the condition $Z = 0$ for the line at infinity is quite natural. Any nonzero vector $\langle x, y\rangle$ has exactly one point $(x, y, 0)$ associated with it, which we regard as the point at infinity in this direction. An ordinary point (x, y) in $\mathbb{R}^2$ is identified with the point given by the class containing the 3-tuple $(x, y, 1)$ in $\mathbb{RP}^2$.

This algebraic construction of $\mathbb{RP}^2$ has a simple geometric realization. Consider the closed upper unit hemisphere H in $\mathbb{R}^3$ and a plane T tangent to it at the north pole. (See FIGURE 4.) Each equivalence class of 3-tuples (X, Y, Z) in $\mathbb{RP}^2$ when $Z \neq 0$ has a unique representative in H, obtained by dividing all components by $\sqrt{X^2+Y^2+Z^2}$ and multiplying by -1 (if necessary) to make the third component positive. When $Z = 0$ there are two representatives for (X, Y, Z) on the boundary circle B of H, which differ by a factor of -1. If we identify these antipodal points on the boundary unit circle, we have a model for $\mathbb{RP}^2$ that we can visualize easily, although we need to be careful to regard pairs of antipodal points as single entities. The correspondence with real points in $\mathbb{R}^2$ is obtained by regarding T as a copy of $\mathbb{R}^2$ and then projecting from T to H along rays from the origin of $\mathbb{R}^3$. Notice that this projection is

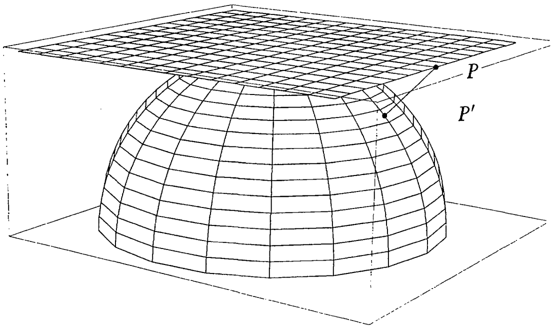


FIGURE 4

compatible with the equivalence relation $\sim$, i.e., each point P of T projects to a unique point P' on the open upper hemisphere contained in H.

Each line L in T projects to a great semi-circle in H which meets B in two antipodal points, which are identified as a single point in $\mathbb{RP}^2$. We regard this single point as the point at infinity on L. Lines parallel to L project to great semi-circles which meet at the same antipodal points on B; thus any two parallel lines meet at a single ideal point in $\mathbb{RP}^2$. Curves in T that are asymptotic to a line L approach the curve at infinity, i.e., they are tangent to the corresponding great semi-circle in H at the point at infinity which lies on it. There is an exceptional case that occurs when the tangent circle is the boundary circle B, which is not the image of any finite line in T. B consists of all points at infinity, so a curve in T whose projection is tangent to B at a point of B has no real asymptote. This happens, for instance, for the parabola $y = x^2$, as we shall see below. Thus we have a nice geometric criterion to detect asymptotes to curves: a curve has a real asymptote if and only if its image in H is tangent at a point of B to a great semicircle different from B.

The extended space $\mathbb{RP}^2$ is in many ways superior to the Euclidean plane $\mathbb{R}^2$. It has a natural topology, which is obtained by forming the quotient of $\mathbb{R}^3$ under the equivalence relation defined above. We obtain this same topology if we form the quotient space of H under the identification of the antipodal points of the boundary circle B. This second approach makes it clear that $\mathbb{RP}^2$ is compact, since it is the continuous image of the compact hemisphere H under the identification map. Each point in $\mathbb{RP}^2$ has a two-dimensional Euclidean neighborhood. This is obvious for all points that are images of the open upper hemisphere in H and is true for the points on B as well, either by noticing that the identification glues together two half discs to create a full Euclidean disc surrounding each such point, or by noticing that in the $\mathbb{R}^3$ construction of $\mathbb{RP}^2$ all points are created equal, so those points that are images of points on B cannot be topologically different from the other points.

The nonsingular linear transformations of $\mathbb{R}^3$ respect the defining equivalence relation, so they define a group of homeomorphisms of $\mathbb{RP}^2$. This group is transitive, since there is such a transformation mapping any nonzero point of $\mathbb{R}^3$ onto any other nonzero point. This transitivity make $\mathbb{RP}^2$ into a homogeneous space. For later work observe that these linear transformations are differentiable with nonvanishing Jacobian, since they are nonsingular. Thus they preserve the tangency of curves (even though they may change the angles at which nontangent curves meet). To establish the theorem specifying asymptotes, we shall use such a linear transformation to map a point at infinity to the origin, where calculations are more familiar. Finally, as noted above, the Euclidean plane $\mathbb{R}^2$ sits naturally in $\mathbb{RP}^2$ as those classes of 3-tuples containing a representative with Z coordinate equal to 1. In summary, $\mathbb{RP}^2$ is a homogeneous compact manifold in which $\mathbb{R}^2$ is naturally embedded.

Any curve in $\mathbb{R}^2$ defined by a polynomial equation $P(x, y) = 0$ of degree n extends naturally to a curve in $\mathbb{RP}^2$ as follows. Replace x by X/Z and y by Y/Z and multiply the entire equation by Z^n to clear the fractions. This procedure produces a homogeneous polynomial in the three variables X, Y, and Z. The resulting equation defines a curve in $\mathbb{RP}^2$, since a homogeneous polynomial has the same zero value at all scalar multiples of any 3-tuple at which it vanishes. For points with $Z \neq 0$ it restricts to the original curve in $\mathbb{R}^2$; thus it defines an extension of the curve $P(x, y) = 0$ into the projective plane. Such curves are called projective algebraic curves. They can be studied using a variety of classical and modern techniques, and form the basic objects of interest in algebraic geometry. The linear equation $ax + by + c = 0$ in $\mathbb{R}^2$, for example, extends to the homogeneous equation $aX + bY + cZ = 0$ in $\mathbb{RP}^2$. As long as not all of a, b, and c are zero, the original line extends to a line in the projective plane by the addition of the single point $(b, -a, 0)$ at infinity. If $a = b = 0$ the line consists entirely of points at infinity and is called the line at infinity.

Our basic problem is to determine all the asymptotes to a general algebraic curve. As noted above, these are the finite lines that are tangent to the curve at a point of infinity in $\mathbb{RP}^2$. For algebraic curves there is an algebraic way to identify such tangent lines. It is based on the idea that a line is tangent to an algebraic curve at a finite point if it has higher "order of contact" at the point than do "generic" lines through the point. When the point is the origin and the curve is defined by a polynomial $P(x, y) = 0$ with $P(x, y) = \Sigma_{k=1}^{n} P_k(x, y)$ as in Section 2 (here we may start the summation at $k = 1$ since $P(0, 0) = 0$), the order of contact with a line is the least k for which $P_k(x, y)$ does not vanish identically on the line. Let l denote this order of contact. Since $P_l(x, y)$ is a form of degree l, at most l distinct lines will have order of contact greater than l, and all others will have order of contact l with the curve. The former set we declare to be the (algebraic) tangent lines to the curve at the origin. This approach is simple algebraically and allows us also to cope with curves that are singular at the origin.

To see the connection with the more familiar approach to tangent lines in calculus, notice that $P_l(x, y)$ is the lth degree Taylor expansion of $P(x, y)$ at $(0, 0)$. When $l = 1$ the curve is said to be nonsingular at $(0, 0)$; otherwise it is singular there. In the nonsingular case let us compute y' at $(0, 0)$ using implicit differentiation. Let $P_1(x, y) = ax + by$. Then we obtain $0 = P_x + P_y y' = a + by' +$ terms that evaluate to 0 at $(0, 0)$ because of the presence of x or y in the term. Thus $y' = -a/b$, so there is a unique tangent line at $(0, 0)$ given by $y = -ax/b$, i.e., by $P_1(x, y) = 0$. This argument justifies this method of finding tangent lines in the nonsingular case. For a discussion of the singular case see **[3]**, pp. 22–24. In this situation the algebraic concept of tangent line does not always agree with our geometric one (because of the artificial restriction that we are looking only at the real portion of our curves). For example, the curve $x^2y^2 = x^4 + y^2$ algebraically has the x-axis ($y^2 = 0$) as a tangent line at the origin, but the curve has the origin as an isolated point on the real graph, since $x^2y^2 \geq y^2$ implies that $x^2 \geq 1$ unless $y = 0$ and the only point on the curve with $y = 0$ is the origin.

As asymptotes for real algebraic curves, we are interested in lines that are real, i.e., we work in $\mathbb{RP}^2$ and not in the corresponding complex projective space $\mathbb{CP}^2$, and in lines that are finite, i.e., not the line at infinity, $Z = 0$. The latter restriction explains the phenomenon that occurred in Example 4 above. It may happen in (B) that all the coefficients $Q_k(b, -a) = 0$ while $P_{n-r}(b, -a) \neq 0$. In this situation there is no finite asymptote in the direction $\langle b, -a \rangle$. If we projectivize the picture, this situation gives rise to the equation $P_{n-r}(b, -a)Z^r = 0$, which does define a line in $\mathbb{RP}^2$, namely, the line at infinity. So there *is* an asymptote in this case, just not a finite asymptote. This is the situation for all curves defined by $y = p(x)$, where $p(x)$ is a polynomial in x of degree greater than one. All such curves have the line at infinity as their only (ideal) asymptote.

We are now in a position to understand what happens in Example 5 above. As noted above, the curve $x^2y^2 = x^4 + y^2$ has the origin as an isolated point, which gives rise to the spurious tangent $y = 0$, obtained by setting its lowest degree form y^2 equal to zero. The curve $x^2y^2 - y^2 + 1 = 0$ of Example 5 was obtained from this curve by the transformation $[X, Y, Z] \rightarrow [Z, Y, X]$, which has the effect of mapping the origin $[0, 0, 1]$ to the point at infinity $[1, 0, 0]$. Thus this curve has a point at infinity as an isolated singular point and it has the x-axis as a spurious asymptote (as we saw in Example 5). Singularities of algebraic curves can be very complicated. But only in this rare situation of an isolated point at infinity in $\mathbb{RP}^2$, which is not isolated if one looks at the entire curve in $\mathbb{CP}^2$, is there no portion of the real curve abutting onto the line that our method has identified. The fact that such candidate lines are not asymptotic will be evident from a machine-drawn graph.

4. Proof of the theorem

Consider now an algebraic curve $P(x, y) = 0$ of degree n, where the polynomial $P(x, y)$ is expressed as a sum of forms $P(x, y) = \Sigma_{k=0}^{n} P_k(x, y)$. Putting this equation into homogeneous coordinates, we obtain the equation

$$F(X,Y,Z) = P_n(X,Y) + ZP_{n-1}(X,Y) + Z^2P_{n-2}(X,Y) + \cdots + Z^nP_0(X,Y) = 0,$$

which extends the affine curve to a projective curve in $\mathbb{RP}^2$. The extended curve, which we shall denote by A, contains points at infinity $(X,Y,0)$ precisely when $F(X,Y,0) = 0$, i.e., when $P_n(X,Y) = 0$. This occurs in those at most n directions $\langle b, -a\rangle$ for which $ax + by$ is a real linear factor of $P_n(x, y)$. We must now determine which of these directions yield lines in $\mathbb{R}^2$ that are possibly tangent to the curve at a point of infinity.

Let us fix a particular factor $ax + by$ of $P_n(x, y)$ to analyze. Without loss of generality we may assume that b is nonzero (since either a or b is nonzero). We shall use a linear transformation to map the point at infinity $[b, -a, 0]$ to $[0, 0, 1]$. This will enable us to do our calculations at the origin. Consider the transformation T: $\mathbb{RP}^2 \to \mathbb{RP}^2$ defined by $T[X,Y,Z] = [bZ, -aZ + Y, X]$. This transformation is invertible, with inverse $T^{-1}[X,Y,Z] = [bZ, bY + aX, X]$, and takes the origin $[0,0,1]$ of the affine plane to the point at infinity $[b, -a, 0]$. This formula for the inverse was obtained by inverting the corresponding matrix and using homogeneity to simplify the expression. It is easy to check that $T^{-1} \circ T[X,Y,Z] = [bX, bY, bZ] \sim [X,Y,Z]$. The curve A defined by $F[X,Y,Z] = 0$ then "pulls back" under T to an associated curve A' defined by $F(T[X,Y,Z]) = F[bZ, -aZ + Y, X] = 0$. In terms of the homogeneous components of $P(X,Y,Z)$, the curve A' is defined by the equation $P_{n-j}(bZ, -aZ + Y)X^j = 0$. With r defined as in the statement of the theorem, we have

$$P_{n-j}(bZ, -aZ + Y) = (bY)^{r-j}Q_{n-j}(bZ, -aZ + Y) \quad \text{for} \quad j = 0, 1, \ldots, r,$$

(where for simplicity we have set Q_{n-r} equal to P_{n-r}). Our equation for A' now takes the form

$$\sum_{j=0}^{r} (bY)^{r-j}Q_{n-j}(bZ, -aZ + Y)X^j + \sum_{j=r+1}^{n} P_{n-j}(bZ, -aZ + Y)X^j = 0.$$

To study the tangents at the origin $[0,0,1]$ of A', we set $Z = 1$, $X = x$, and $Y = y$ to obtain the restriction of the curve to the affine plane $\mathbb{R}^2$. This yields the equation

$$\sum_{j=0}^{r} (by)^{r-j}Q_{n-j}(b, -a + y)x^j + \sum_{j=r+1}^{n} P_{n-j}(b, -a + y)x^j = 0.$$

This equation is a polynomial equation in x and y, and the tangent lines to the curve at the origin are given by those lines that satisfy the equation $L(x, y) = 0$, where $L(x, y)$ is the lowest order nonvanishing form in this polynomial. It is clear from this equation that the degree of $L(x, y)$ is r since at least one of the terms $Q_{n-j}(b, -a + y)$ has nonzero constant term $Q_{n-j}(b, -a)$ (by the definition of r in the statement of the theorem). Thus the lowest order nonvanishing form in the above equation is given by

$$L(x, y) = \sum_{j=0}^{r} (by)^{r-j} x^j Q_{n-j}(b, -a). \tag{E}$$

The tangent lines to A' at the origin are the solutions to the equation $L(x, y) = 0$.

To find the tangent lines to our original curve A at the point at infinity $[b, -a, 0]$, we extend this equation to $\mathbb{RP}^2$ and pull back using T^{-1}. Notice that we are using the fact that T and T^{-1} preserve tangency of curves. Since L is homogeneous, the extension to $\mathbb{RP}^2$ is given by

$$H(X,Y,Z) = \sum_{j=0}^{r} (bY)^{r-j} X^j Q_{n-j}(b, -a) = 0.$$

Under T^{-1} this equation pulls back to the equation

$$\begin{aligned} 0 = H(T^{-1}[X,Y,Z]) &= H[bZ, bY + aX, X] \\ &= \sum_{j=0}^{r} (b(bY + aX))^{r-j} (bZ)^j Q_{n-j}(b, -a) \\ &= b^r \sum_{j=0}^{r} (aX + bY)^{r-j} Z^j Q_{n-j}(b, -a). \end{aligned}$$

Since b is nonzero, this is equivalent to the equation

$$\sum_{j=0}^{r} (aX + bY)^{r-j} Z^j Q_{n-j}(b, -a) = 0.$$

Setting $Z = 1$, $X = x$, and $Y = y$, we obtain $\sum_{j=0}^{r}(ax + by)^{r-j} Q_{n-j}(b, -a) = 0$, which specifies the lines in the direction $\langle b, -a \rangle$ possibly asymptotic to the curve A. This equation is identical to (C) in the statement of the theorem.

Remarks. The ideas in **[5]** can be developed to yield a proof of the Main Theorem which is based entirely on limits and avoids the projective plane. But this route is somewhat circuitous, and does not emphasize the geometric interpretation of an asymptote as a line that is tangent to a curve at infinity. It is also possible to establish the Main Theorem by using an appropriate version of Taylor's theorem at the point at infinity $[b, -a, 0]$. This avoids the transformation T but introduces its own complications. The argument using T is essentially just linear algebra, and was suggested by one of the referees.

Acknowledgment. The author wishes to thank both referees for their careful reading and helpful comments, which resulted in a much improved paper.

REFERENCES

1. Egbert Brieskorn and Horst Knörrer, *Plane Algebraic Curves*, Birkhäuser Verlag, Boston, MA, 1986.
2. Julian Lowell Coolidge, *A Treatise on Algebraic Plane Curves*, Clarendon Press, Oxford, UK, 1931.
3. Percival Frost, *An Elementary Treatise on Curve Tracing*, 2nd ed., Macmillan, London, UK, 1911.
4. Francis Kirwan, *Complex Algebraic Curves*, Cambridge University Press, Cambridge, UK, 1992.
5. Gabriel Klambauer, *Aspects of Calculus*, Springer-Verlag, New York, NY, 1986.
6. Jeffrey Nunemacher, Newton's investigation of cubic curves, in *Problems for Student Investigation*, MAA Notes Number 30, Math. Assoc. of America, Washington, DC, 1993.
7. E. J. F. Primrose, *Plane Algebraic Curves*, Macmillan, London, UK, 1955.
8. George Simmons, *Calculus with Analytic Geometry*, McGraw-Hill, New York, NY, 1985.

Full Rank Factorization of Matrices

R. PIZIAK
P. L. ODELL
Baylor University
Waco, TX 76798

1. Introduction

There are various useful ways to write a matrix as the product of two or three other matrices that have special properties. For example, today's linear algebra texts relate Gaussian elimination to the LU factorization and the Gram–Schmidt process to the QR factorization. In this paper, we consider a factorization based on the rank of a matrix. Our purpose is to provide an integrated theoretical development of and setting for understanding a number of topics in linear algebra, such as the Moore–Penrose generalized inverse and the Singular Value Decomposition. We make no claim to a practical tool for numerical computation—the rank of a very large matrix may be difficult to determine. However, we will describe two applications; one to the explicit computation of orthogonal projections, and the other to finding explicit matrices that diagonalize a given matrix.

2. Rank

Let $\mathbb{C}$ denote the field of complex numbers and $\mathbb{C}^{m\times n}$ the collection of m-by-n matrices with entries from $\mathbb{C}$. If $A \in \mathbb{C}^{m\times n}$, let A^* denote the conjugate transpose (sometimes called the Hermitian adjoint) of A; A^* is formed by taking the complex conjugate of each entry in A and then transposing the resulting matrix.

A very important (but not always easily discoverable) nonnegative integer is associated with each matrix A in $\mathbb{C}^{m\times n}$. The rows of A can be viewed as vectors in $\mathbb{C}^n$, and the columns as vectors in $\mathbb{C}^m$. The rows span a subspace called the *row space* of A; the dimension of the row space is called the *row rank* of A. The *column rank* of A is the dimension of the subspace of $\mathbb{C}^m$ spanned by the columns. Remarkably, the row rank and the column rank are always the same, so we may unambiguously refer to the *rank* of A. Let $r(A)$ denote the rank of A and $\mathbb{C}_r^{m\times n}$ the collection of matrices of rank r in $\mathbb{C}^{m\times n}$. We say that a matrix A in $\mathbb{C}^{m\times n}$ has *full row rank* if $r(A) = m$ and *full column rank* if $r(A) = n$. The following are some basic facts about rank that we will find useful (see **[11]**):

- for $A \in \mathbb{C}^{m\times n}$, $r(A) \leq \min(m, n)$;
- for $A, B \in \mathbb{C}^{m\times n}$, $r(A+B) \leq r(A) + r(B)$;
- for $A \in \mathbb{C}^{m\times n}$, $r(A) = r(A^*) = r(A^*A)$;
- for $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{n\times p}$, $r(AB) \leq \min(r(A), r(B))$;
- if B and C are invertible, then $r(AB) = r(A) = r(CA)$.

3. Full rank factorizations

Let $A \in \mathbb{C}_r^{m\times n}$, with $r > 0$. If we can find F in $\mathbb{C}_r^{m\times r}$ and G in $\mathbb{C}_r^{r\times n}$ such that $A = FG$, then we say that we have a *full rank factorization* of A. It is not difficult to see that every matrix A in $\mathbb{C}_r^{m\times n}$ with $r > 0$ has such a factorization. One approach is to choose for F any matrix whose columns form a basis for the column space of A. Then, since each column of A is uniquely expressible as a linear combination of the columns of F, the coefficients in the linear combinations determine a unique matrix G in $\mathbb{C}^{r\times n}$ with $A=FG$. Moreover, $r=r(A)=r(FG)\leq r(G)\leq r$ so that G is in $\mathbb{C}_r^{r\times n}$.

Another approach, which will lead us to an algorithm, is to apply elementary matrices on the left of A (that is, elementary row operations) to produce the unique row reduced echelon form of A, RREF(A). In other words, we compute an invertible matrix R in $\mathbb{C}^{m\times m}$ with

$$RA=\begin{bmatrix} G_{r\times n} \\ \cdots\cdots\cdots\cdots \\ \mathbb{O}_{(m-r)\times n} \end{bmatrix},$$

where $r=r(A)=r(G)$ and $\mathbb{O}_{(m-r)\times n}$ is the matrix, consisting entirely of zeros, of $m-r$ rows and n columns. Then

$$A=R^{-1}\begin{bmatrix} G \\ \cdots \\ \mathbb{G} \end{bmatrix}.$$

Let R_1 consist of the first r columns of R and R_2 the remaining columns. Then R_1 is m-by-r and R_2 is m-by-$(m-r)$, and

$$A=\begin{bmatrix} R_1 \vdots R_2 \end{bmatrix}\begin{bmatrix} G \\ \cdots \\ \mathbb{O} \end{bmatrix}=R_1G+R_2\mathbb{O}=R_1G.$$

Now take F to be R_1. Since R^{-1} is invertible, its columns are linearly independent so F has r independent columns, and hence has full column rank. This leads us to an algorithm for computing a full rank factorization of a matrix A in $\mathbb{C}_r^{m\times n}$:

Step 1. Use elementary row operations to reduce A to row reduced echelon form RREF(A).

Step 2. Construct a matrix F from the columns of A that correspond to the columns with the leading ones in RREF(A), placing them in F in the same order as they appear in A.

Step 3. Construct a matrix G from the nonzero rows of RREF(A), placing them in G in the same order as they appear in RREF(A). Then $A=FG$ is a full rank factorization of A.

Example. If $A=\begin{bmatrix} 3 & 6 & 13 \\ 2 & 4 & 9 \\ 1 & 2 & 3 \end{bmatrix}$, then RREF$(A)=\begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. So, with $G=\begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $F=\begin{bmatrix} 3 & 13 \\ 2 & 9 \\ 1 & 3 \end{bmatrix}$, we have $A=FG$, a full rank factorization.

Full rank factorizations not only exist, but abound. Indeed, if $A=FG$ is any full rank factorization of A in $\mathbb{C}_r^{m\times n}$ and R is any invertible matrix in $\mathbb{C}^{r\times r}$, then $A=FG=FRR^{-1}G=(FR)(R^{-1}G)$ is another full rank factorization of A.

4. Pseudoinverses from full rank factorizations

The concept of a *pseudoinverse* (now beginning to appear in undergraduate linear algebra texts) has its roots in the central problem of linear algebra: solving systems of linear equations. Let's consider such a system

$$Ax=b,$$

where the coefficient matrix A is m-by-n and has rank r. If $n = m = r$, then $x = A^{-1}b$ is the unique solution. But what if A is not square, or, even if A is square, if A^{-1} does not exist? Note that, for any matrix A, if we can produce another matrix B such that $ABA = A$, then Bb will be one solution to $Ax = b$ if a solution exists. To see this suppose $Ax = b$. Then $BAx = Bb$ so $b = Ax = ABAx = A(Bb)$. Surely, if A is square and invertible then $B = A^{-1}$ has the property $ABA = AA^{-1}A = A$. The matrix B is called a *generalized inverse* of A and is not necessarily unique. The first published work on generalized inverses goes back to Moore **[12]**. However, not until 1955 did the theory blossom, when Penrose **[13]** defined a uniquely determined generalized inverse for any matrix A. Today we use the name *pseudoinverse* or *Moore–Penrose inverse*. Penrose showed that, given any matrix A, there is one and only one matrix B satisfying the following four conditions:

$$ABA = A; \quad BAB = B; \quad (AB)^* = AB; \quad (BA)^* = BA.$$

We can write A^+ for the (unique) solution B to these four equations. Our first goal is to obtain A^+ by beginning with a full rank factorization of A. By the way, it is easy to see from uniqueness that $A^{++} = A$ for any A.

Suppose $A \in \mathbb{C}_r^{m \times n}$, with $r > 0$, and suppose $A = FG$ is a full rank factorization of A. Then $F \in \mathbb{C}_r^{m \times r}$, $G \in \mathbb{C}_r^{r \times n}$, and $r = r(A) = r(F) = r(G)$. Now G has full row rank, so GG^* has full rank in $\mathbb{C}^{r \times r}$, and hence is invertible. Similarly, F has full column rank, so F^*F has full rank in $\mathbb{C}^{r \times r}$ and is therefore invertible. We now have our first main result.

THEOREM 1. *Let $A \in \mathbb{C}_r^{m \times n}$ with $r(A) > 0$, and suppose $A = FG$ is a full rank factorization of A. Then*

(1) $F^+ = (F^*F)^{-1}F^*$
(2) $F^+F = I_r$, *the r-by-r identity matrix*
(3) $G^+ = G^*(GG^*)^{-1}$
(4) $GG^+ = I_r$
(5) $A^+ = G^+F^+$

Proof. Items (2) and (4) are trivial consequences of the definitions. The existence of F^+, G^+, and A^+ follows from the discussion preceding the theorem. Thus it suffices to show F^+, G^+, and A^+ satisfy their respective Moore–Penrose equations. This boils down to "symbol pushing"; we illustrate just a few calculations to suggest the flavor. For example, we need $GG^+G = G$, but

$$GG^+G = G\big(G^*(GG^*)^{-1}\big)G = (GG^*)(GG^*)^{-1}G = IG = G$$

where I is the identity matrix. The next calculation is similar:

$$\begin{aligned}(G^+G)^* &= \big(G^*(GG^*)^{-1}G\big)^* = G^*(GG^*)^{-1*}G^{**} = G^*(GG^*)^{*^{-1}}G \\ &= G^*(GG^*)^{-1}G = G^+G.\end{aligned}$$

The remaining arguments for F^+ and G^+ are similar. For A^+, we compute

$$AA^+A = AG^+F^+A = FGG^+F^+FG = FI_rI_rG = FG = A.$$

The remaining computations are similar; we leave them to the reader.

Equations (2) and (4) above are trivial, but they point out one advantage of computing with full rank factorizations: F^+ is a left inverse of F while G^+ is a right inverse for G.

We can now clean up a loose end. We noted earlier that full rank factorizations are not unique: if $A = FG$ is one full rank factorization and R is invertible of appropriate size, then $A = (FR)(R^{-1}G)$ is another full rank factorization. Does it get any worse than this? The next theorem says no.

THEOREM 2. *Every matrix A in $\mathbb{C}_r^{m\times n}$ with $r(A) > 0$ has infinitely many full rank factorizations. However, if $A = FG = F_1G_1$ are two full rank factorizations of A, then there exists an invertible matrix R in $\mathbb{C}^{r\times r}$ such that $F_1 = FR$ and $G_1 = R^{-1}G$. Moreover, $G_1^+ = (R^{-1}G)^+ = G^+R$ and $F_1^+ = (FR)^+ = R^{-1}F^+$.*

Proof. The first claim is now clear. Suppose $A = FG = F_1G_1$ are two full rank factorizations of A. Then $F_1^+F_1G_1 = F_1^+FG$; since $F_1^*F_1 = I_r$, we have $G_1 = (F_1^+F)G$. Note that F_1^+F is r-by-r and

$$r = r(G_1) = r((F_1^+F)G) \leq r(F_1^+F) \leq r,$$

so F_1^+F has full rank r; therefore F_1^+F is invertible. Similar reasoning shows that GG_1^+ is invertible. Let $S = F_1^+F$ and $GG_1^+ = R$. Then

$$SR = F_1^+FGG_1^+ = F_1^+AG_1^+ = F_1^+F_1G_1G_1^+ = I_r,$$

so $S = R^{-1}$. Therefore, $G_1 = SG = R^{-1}G$ and $F_1 = FGG_1^+ = FR$. To complete the proof, we calculate

$$(FR)^+ = ((FR)^*(FR))^{-1}(FR)^* = (R^*F^*FR)^{-1}R^*F^* = R^{-1}(F^*F)^{-1}R^{*-1}R^*F^*$$
$$= R^{-1}(F^*F)^{-1}F^* = R^{-1}F^+.$$

The computation to show $G_1^+ = G^+R$ is similar.

Example. As in the preceding example, let

$$A = \begin{bmatrix} 3 & 6 & 13 \\ 2 & 4 & 9 \\ 1 & 2 & 3 \end{bmatrix} = FG = \begin{bmatrix} 3 & 13 \\ 2 & 9 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Using Theorem 1, we find

$$G^+ = \begin{bmatrix} 1/5 & 0 \\ 2/5 & 0 \\ 0 & 1 \end{bmatrix}; \quad F^+ = \begin{bmatrix} -3/26 & -11/13 & 79/26 \\ 1/13 & 3/13 & -9/13 \end{bmatrix}.$$

The pseudoinverse of A is the matrix product

$$A^+ = G^+F^+ = \begin{bmatrix} -3/130 & -11/65 & 79/130 \\ -3/65 & -22/65 & 79/65 \\ 1/13 & 3/13 & -9/13 \end{bmatrix}.$$

5. Four fundamental projections

Strang [11] has popularized a structural view of any matrix A in $\mathbb{R}^{m\times n}$; we adapt his approach to $\mathbb{C}^{m\times n}$. He assigns four subspaces to A: the column space of A, denoted $\mathcal{R}(A)$; the null space of A, denoted $\mathcal{N}(A)$; the column space of A^*, $\mathcal{R}(A^*)$; and the null space of A^*, $\mathcal{N}(A^*)$. With the help of a full rank factorization of A and the pseudoinverse, we can easily compute the orthogonal projections onto these subspaces.

A *projection* is a matrix P with $P^2 = P = P^*$; each subspace of $\mathbb{C}^n$ uniquely determines such a projection. The set of fixed vectors ($Px = x$) for P then coincides with the subspace. A standard method of computing projections begins by finding an orthonormal basis for the subspace. This is unnecessary, given a full rank factorization, since the Moore–Penrose equations imply

(i) $AA^+ = FF^+$ is the projection onto $\mathscr{R}(A)$;
(ii) $A^+A = G^+G$ is the projection onto $\mathscr{R}(A^*)$;
(iii) $I_m - AA^+ = I_m - FF^+$ is the projection onto $\mathscr{N}(A^*)$;
(iv) $I_n - A^+A = I_n - G^+G$ is the projection onto $\mathscr{N}(A)$.

Continuing our example from above,

$$G^+G = \begin{bmatrix} 1/5 & 2/5 & 0 \\ 2/5 & 4/5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad I - G^+G = \begin{bmatrix} 4/5 & -2/5 & 0 \\ -2/5 & 1/5 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

are, respectively, the projections onto $\mathscr{R}(A^*)$ and $\mathscr{N}(A)$, while

$$FF^+ = \begin{bmatrix} 17/26 & 6/13 & 3/20 \\ 6/13 & 5/13 & -2/13 \\ 3/26 & -2/13 & 25/26 \end{bmatrix} \quad \text{and} \quad I - FF^+ = \begin{bmatrix} 9/26 & -6/13 & -3/26 \\ -6/13 & 8/13 & 2/13 \\ -3/26 & 2/13 & 1/26 \end{bmatrix}$$

are the projections onto $\mathscr{R}(A^*)$ and $\mathscr{N}(A^*)$, respectively.

6. Special full rank factorizations

Suppose $A \in \mathbb{C}_r^{m \times n}$, with $r > 0$. We have seen that one way to produce a full rank factorization of A is to use as the columns of F a basis for the column space of A. Suppose we pick an orthonormal basis (if a basis isn't orthonormal, the Gram–Schmidt process can be applied). Then we have $A = FG$, where $F^*F = I_r$ because the columns of F are orthonormal. It is easy to check that F^* satisfies the four Moore–Penrose equations, so $F^* = F^+$. Therefore, among the many full rank factorizations $A = FG$, we can always select one for which $F^+ = F^*$. We'll call such a full rank factorization *orthogonal*. A *unitary* matrix U is one for which $U^* = U^{-1}$. Since unitary matrices preserve length and orthogonality, we see that $A = (FU)(U^*G)$ is again orthogonal if $A = FG$ is. We summarize with a theorem.

THEOREM 3. *Every matrix* $A \in \mathbb{C}_r^{m \times n}$ *with* $r(A) > 0$ *has infinitely many orthogonal full rank factorizations.*

Next, we consider the special case of a projection matrix P in $\mathbb{C}_r^{n \times n}$. First we write P in an orthogonal full rank factorization $P = FG$, so $F^+ = F^*$; then

$$P = P^* = (FG)^* = G^*F^* = G^*F^+.$$

For a projection P, the Moore–Penrose equations imply that $P^+ = P^*$, so $P = P^+ = G^+F^+ = G^+F^*$, and $GP = GG^+F^* = F^*$. But then $P = PP = FGP = FF^*$. Actually, this must be the case since $FF^* = FF^+$ is the projection on the range of P, which is P. We have shown:

THEOREM 4. *Every projection* P *in* $\mathbb{C}_r^{n \times n}$ *has a full rank factorization* $P = FG$ *for which* $G = F^* = F^+$.

Returning to our ongoing example, the projection onto $\mathcal{N}(A^*)$ has the full rank factorization

$$I-FF^+=\begin{bmatrix} 9/26 & -6/13 & -3/26 \\ -6/13 & 8/13 & 2/13 \\ -3/26 & 2/13 & 1/26 \end{bmatrix}=\begin{bmatrix} 9/26 \\ -6/13 \\ -3/26 \end{bmatrix}\begin{bmatrix} 1 & -4/3 & -1/3 \end{bmatrix}.$$

This factorization is not orthogonal, but Gram–Schmidt is easy to apply. Normalizing the column vector gives $\begin{bmatrix} \frac{3}{\sqrt{26}} \\ -\frac{4}{\sqrt{26}} \\ -\frac{1}{\sqrt{26}} \end{bmatrix}$. Direct computation shows

$$I-FF^+=\begin{bmatrix} \frac{3}{\sqrt{26}} \\ -\frac{4}{\sqrt{26}} \\ -\frac{1}{\sqrt{26}} \end{bmatrix}\begin{bmatrix} \frac{3}{\sqrt{26}} & -\frac{4}{\sqrt{26}} & -\frac{1}{\sqrt{26}} \end{bmatrix}.$$

7. Matrix equivalence by full rank factorization

Matrices A and B are *equivalent*, and we write $A \sim B$, if B can be obtained from A by applying both elementary row and elementary column operations to A. Thus $A \sim B$ if there exist nonsingular matrices S and T with $SAT=B$.

Let $A=FG$ be a full rank factorization of A. We will show how F and G can be used to construct matrices S and T that will bring A into canonical form for the equivalence relation $\sim$. Consider

$$B=\begin{bmatrix} F^+ \\ \cdots\cdots\cdots\cdots \\ W_1(I-FF^+) \end{bmatrix} A\begin{bmatrix} G^+ \,\vdots\, (I-G^+G)W_2 \end{bmatrix}$$

where W_1 and W_2 are arbitrary matrices of appropriate dimension ($W_1 \in \mathbb{C}^{(m-r)\times m}$ and $W_2 \in \mathbb{C}^{n\times(n-r)}$). Computation gives

$$B=\begin{bmatrix} F^+AG^+ & F^+A(I-G^+G)W_2 \\ W_1(I-FF^+)AG^+ & A(I-GG^+)W_2 \end{bmatrix}=\begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}.$$

In fact, we could take *any* r-by-r matrix M and compute

$$\begin{bmatrix} F^+ \\ \cdots\cdots\cdots\cdots \\ W_1(I-FF^+) \end{bmatrix} A\begin{bmatrix} G^+M \,\vdots\, (I-G^+G)W_2 \end{bmatrix}=\begin{bmatrix} M & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix},$$

again with W_1 and W_2 arbitrary matrices of appropriate size.

We have uncovered the interesting fact that every full rank factorization of A leads to a *diagonal reduction* of A. Of course, the matrices S and T flanking A need not be invertible, but this can be arranged, again using full rank factorizations. Starting with any full rank factorization $A=FG$, we construct the projections

$$I-FF^+=F_1F_1^*=F_1F_1^+ \quad\text{and}\quad I-G^+G=F_2F_2^*=F_2F_2^+,$$

each in an orthogonal full rank factorization.

To make S invertible, we choose W_1 judiciously. A computation gives

$$\begin{bmatrix} F^+ \\ \cdots\cdots\cdots\cdots \\ W_1(I-FF^+) \end{bmatrix}\begin{bmatrix} F \vdots (I-FF^+)W_1^* \end{bmatrix} = \begin{bmatrix} F^+F & F^+(I-FF^+)W_1^* \\ W_1(I-FF^+)F & W_1(I-FF^+)W_1^* \end{bmatrix}$$

$$= \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & W_1(I-FF^+)W_1^* \end{bmatrix}.$$

We need the identity matrix to appear in the lower right-hand corner of the preceding matrix. If we choose $W_1 = F_1^+ = F_1^*$, then

$$W_1(I-FF^+)W_1^* = W_1F_1F_1^*W_1^* = F_1^+F_1F_1^*F_1^{**} = IF_1^+F_1 = I.$$

Thus $S^{-1} = [F \vdots (I-FF^+)F_1] = [F \vdots F_1]$. Similarly, for $T = [G^+ \vdots (I-G^+G)W_2]$ we choose $W_2 = F_2$ and find that

$$T^{-1} = \begin{bmatrix} G \\ F_2^*(I-G^+G) \end{bmatrix} = \begin{bmatrix} G \\ F_2^+ \end{bmatrix}.$$

So we have derived a familiar theorem but with a new twist.

THEOREM 5. *Every matrix A in $\mathbb{C}_r^{m\times n}$ is equivalent to* $\begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$. *If $A = FG$ is a full rank factorization of A and $I - FF^+ = F_1F_1^+$ and $I - G^+G = F_2F_2^+$ are orthogonal full rank factorizations, then*

$$S = \begin{bmatrix} F^+ \\ \cdots \\ F_1^+ \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} G^+ \vdots F_2 \end{bmatrix}$$

are invertible, and

$$SAT = \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}.$$

A little more is true: the matrix S can be chosen to be *unitary*, not just invertible. To see this, we begin with an orthogonal full rank factorization $A = FG$, where $F^* = F^+$. Then

$$SS^* = \begin{bmatrix} F^+ \\ \cdots\cdots\cdots\cdots \\ W_1(I-FF^+) \end{bmatrix}\begin{bmatrix} (F^+)^* \vdots (I-FF^+)W_1^* \end{bmatrix}$$

$$= \begin{bmatrix} F^+(F^+)^* & \mathbb{O} \\ \mathbb{O} & W_1(I-FF^+)W_1^* \end{bmatrix} = \begin{bmatrix} F^+F & \mathbb{O} \\ \mathbb{O} & W_1(I-FF^+)W_1^* \end{bmatrix}$$

$$= \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & W_1(I-FF^+)W_1^* \end{bmatrix}.$$

Now we choose $W_1 = F_1^+$ as before to get $SS^* = I$.

8. The singular value decomposition

We will now see how to derive the singular value decomposition of a matrix beginning with a full rank factorization. We have seen that, for given $A \in \mathbb{C}_r^{m\times n}$, full rank factorizations lead to explicit S and T, with S unitary and T invertible, such that

$SAT = \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$. Can we get T unitary as well? The answer is yes if we replace the matrix I_r with a slightly more general diagonal matrix D.

As before, we begin with an orthogonal full rank factorization $A = FG$ with $F^+ = F^*$. We also use the factorizations $I - FF^+ = F_1 F_1^+ = F_1 F_1^*$ and $I - G^+G = F_2 F_2^* = F_2 F_2^+$. Then

$$\begin{bmatrix} F^+ \\ F_1^+(I - FF^+) \end{bmatrix} A \left[G^+D \,\vdots\, (I - GG^+)W_2 \right] = \begin{bmatrix} D & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}.$$

The matrix $\begin{bmatrix} F^+ \\ F_1^+ \end{bmatrix}$ is unitary; let's call it U^*; the matrix W_2 is arbitrary. Next we consider $V = \left[G^+D \,\vdots\, (I - G^+G)W_2 \right]$, which we would like to make unitary by choice of D and W_2. But

$$\begin{aligned} V^*V &= \begin{bmatrix} (G^+D)^* \\ [(I - G^+G)W_2]^* \end{bmatrix} \left[G^+D \,\vdots\, (I - G^+G)W_2 \right] \\ &= \begin{bmatrix} D^*G^{+*}G^+D & D^*G^{+*}(I - G^+G)W_2 \\ W_2^*(I - G^+G)G^+D & W_2^*(I - G^+G)W_2 \end{bmatrix} \\ &= \begin{bmatrix} D^*G^{+*}G^+D & \mathbb{O} \\ \mathbb{O} & W_2^*(I - G^+G)W_2 \end{bmatrix} \end{aligned}$$

since $G^{+*} = (G^+GG^+)^* = G^{+*}(G^+G)^* = G^{+*}G^+G$ implies that $D^*G^{+*}(I - G^+G)W_2 = \mathbb{O}$. We have already seen that with $W_2 = F_2^+$ we get I_{m-r} in the lower right position. So the problem reduces to solving $D^*G^{+*}G^+D = I_r$ for a suitable D. But $G = F^+A = F^*A$, so

$$D^*G^{+*}G^+D = D^*(GG^*)^+D = D^*(G^{+*})^{-1} = D^*(F^*AA^*F)^{-1}D.$$

To achieve the identity matrix I_r we need $F^*AA^*F = DD^*$ or, equivalently, $AA^*F = FDD^*$. We summarize our results in a theorem.

THEOREM 6. *Let $A = FG$ be an orthogonal full rank factorization. If there exists $D \in \mathbb{C}^{r \times r}$ with $GG^* = DD^*$, then there exist unitary matrices S* and *T* with $\mathrm{SAT} = \begin{bmatrix} D & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$.

One way to exhibit such a matrix D for a given A is to choose for the columns of F an orthonormal basis consisting of the eigenvectors of AA^* corresponding to nonzero eigenvalues. Since AA^* is positive semidefinite we know its eigenvalues are nonnegative. Then $AA^*F = FE$ where E is the r-by-r diagonal matrix of real positive eigenvalues of AA^*. Let D be the diagonal matrix of the positive square roots of these eigenvalues. Then

$$D^*[F^*AA^*F]^{-1}D = DE^{-1}D = I_r.$$

We have captured the classical singular value decomposition in the context of full rank factorization.

THEOREM 7. *Let $A = FG$ be a full rank factorization of A where the columns of F are an orthonormal basis consisting of eigenvectors of AA^* corresponding to nonzero eigenvalues. Suppose*

$$I - FF^+ = F_1F_1^* = F_1F_1^+ \quad \text{and} \quad I - G^*G = F_2F_2^* = F_2F_2^+.$$

Then there exist unitary matrices U and V with $U^*AV = \begin{bmatrix} D_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$, where D_r is an r-by-r diagonal matrix whose diagonal entries are the square roots of the nonzero eigenvalues of AA^*. The matrices U and V can be constructed explicitly from $U^* = \begin{bmatrix} F^* \\ F_1^+ \end{bmatrix}$ and $V = \begin{bmatrix} G^+D \vdots F_2 \end{bmatrix}$.

Conclusion

We have only scratched the surface of what can be obtained from full rank factorizations. More information can be found in the references. Of course, ours is not the only point of view. One could begin with the classical singular value decomposition and derive a full rank factorization from it. Briefly, it goes like this: We write

$$A = U\begin{bmatrix} E_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} V^* = \begin{bmatrix} U_r \vdots U_{m-r} \end{bmatrix} \begin{bmatrix} E_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \begin{bmatrix} V_r \\ \cdots\cdots \\ V_{n-r} \end{bmatrix}^* = U_r E_r V_r^* = U_r(E\, V_r^*).$$

Taking $F = U_r$ and $G = EV_r^*$ we get a full rank (in fact, orthogonal full rank) factorization of A in $\mathbb{C}_r^{m\times n}$. It can be shown that A^+ has full rank factorization $A^+ = V_r(E^{-1}U_r^*)$, $V_rV_r^*$ is the projection on $\mathscr{R}(A^*)$, $V_{n-r}V_{n-r}^*$ is the projection on the $\mathscr{N}(A)$, $U_rU_r^*$ is the projection on $\mathscr{R}(A)$, and $U_{m-r}U_{m-r}^*$ is the projection on $\mathscr{N}(A^*)$. From a numerical linear algebra point of view, starting with the singular value decomposition probably makes more sense since there are effective and stable algorithms available for its direct computation. Also, there are full rank QR and full rank LU factorizations and ways to produce bases for the fundamental subspaces of a matrix. But that is another story.

REFERENCES

1. Anderson, Jr., W. N. and Duffin, R. J., Series and parallel addition of matrices, *J. Math. Anal. Appl.* 26 (1969), 576–595.
2. Ben-Israel, A. and Greville, T. N. F., *Generalized Inverses: Theory and Applications*, John Wiley & Sons, New York, NY, 1974.
3. Boullion, Thomas L., and Odell, Patrick L., *Generalized Inverse Matrices*, Wiley-Interscience, New York, NY, 1971.
4. Campbell, S. L. and Meyer, Jr., C. D., *Generalized Inverses of Linear Transformations*, Dover Publications, Inc., New York, NY, 1979.
5. Halmos, P. R., *A Hilbert Space Problem Book*, Van-Nostrand Co., Princeton, NJ, 1967.
6. Mitra, S. K., Simultaneous diagonalization of rectangular matrices, *Linear Algebra and its Applications* 47 (1982), 139–150.
7. Moore, E. H., On the reciprocal of the general algebraic matrix, *Bull. Amer. Math. Soc.* 26 (1920), 394–395.
8. Odell, P. L. and Boullion, T. L., On simultaneous diagonalization of rectangular matrices, Vol. 33, No. 9, pp. 93–96, (1997), *Computers and Mathematics with Application*, Pergamon Press, Elmsford, NY.
9. Penrose, R., A generalized inverse for matrices, *Proc. Camb. Phil. Soc.* 51 (1955), 406–413.
10. Rao, C. R. and Mitra, S. K., *Generalized Inverse of Matrices and Applications*, John Wiley and Sons, Inc., New York, NY, 1971.
11. Strang, G., *Linear Algebra and its Applications*, 3rd Ed., Harcourt Brace Jovanovich, San Diego, CA, 1988.
12. Sykes, Jeffrey D., Applications of Full Rank Factorization To Solving Matrix Equations, Masters Thesis, Baylor University, December, 1992.
13. Wardlaw, W. P., Minimum and characteristic polynomials of low-rank matrices, this MAGAZINE 68 (1995), 122–127.

NOTES

Those Ubiquitous Archimedean Circles[1]

CLAYTON W. DODGE
University of Maine
Orono, ME 04469-5752

THOMAS SCHOCH
Krefelder Strasse 21
45145 Essen
Germany

PETER Y. WOO
Biola University
La Mirada, CA 90639

PAUL YIU
Florida Atlantic University
Boca Raton, FL 33431

The Bankoff triplet circle The *arbelos*, the figure formed by three mutually tangent semicircles with collinear centers and shown in FIGURE 1, has fascinated geometers since the time of the early Greeks. Also called the shoemaker's knife because it is shaped like that tool, it has been the subject of much study over the centuries. Many amazing and counterintuitive properties have been discovered in this figure, a few of which are described by Bankoff ([1] and [2]). That such a simple figure should be so rich is perhaps not so surprising since the arbelos is, after all, a triangle whose sides are semicircles.

Label the common diameter ACB and let the three semicircles be (O), (O_1), and (O_2) as shown in FIGURE 2. If one erects the common internal tangent line CD to the two interior circles, then the circles (W_1) and (W_2) inscribed in the resulting two regions ACD and BCD are called the *twin circles of Archimedes* and have the same radius. In 1974 Bankoff [1] pointed out that the twin circles of Archimedes are not

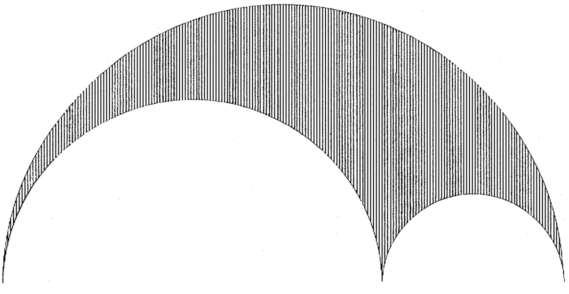


FIGURE 1
The arbelos.

[1] This paper is dedicated to the memory of Leon Bankoff.

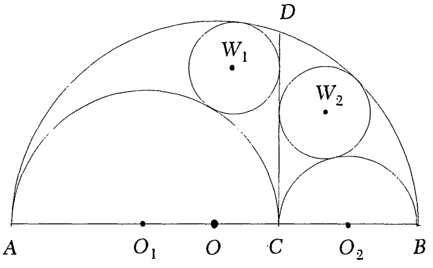


FIGURE 2
The twin circles.

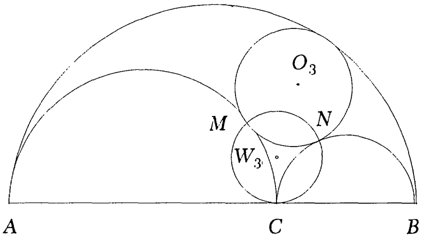


FIGURE 3
The Bankoff triplet circle.

twins, but two of triplets. That is, there is a third circle in the arbelos with the same radius. Inscribe circle (O_3) in the arbelos as shown in FIGURE 3. Then the circle (W_3) that passes through point C and the points of tangency M and N of circle (O_3) with circles (O_1) and (O_2) has the same radius as the twin circles. It is the *Bankoff triplet circle*. We denote certain circles congruent to the twin circles by (W_n) for positive integral n. Proofs of any of these assertions are postponed until after we note some other family members.

One might think that triplets are enough for any one household, but Bankoff discovered yet another member of that famous family. Let EF be the common external tangent to circles (O_1) and (O_2). *The Bankoff quadruplet circle* (W_4) is the circle inscribed in the circular segment of semicircle (O) and the chord EF (extended). See FIGURE 4. Furthermore, it is tangent to circle (O) at point D and is the smallest circle through point D and tangent to line EF. Now draw radius OD to cut EF at G. Then circle (W_4) has diameter GD.

The Dodge circles Bankoff and I (Clayton Dodge) discussed his discoveries, which led me to observe that if we drop a perpendicular CD' from point C to line OD, then $D'D$ is twice the diameter of an Archimedean circle. Furthermore, the two circles shown in FIGURE 5 include (W_4). We label the new circle, whose diameter is $D'G$, (W_5).

FIGURE 6 shows the translations of circles (W_1) and (W_2) that drop their centers onto the common diameter AB as circles (W_6) and (W_7), and the circle (W_8) on their centers as diameter. If these circles were merely translations, their interest would be quite low, but I found other reasons for their consideration. The common external

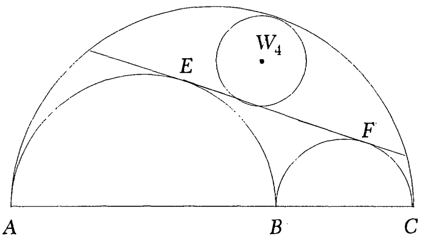


FIGURE 4
The Bankoff circle 4.

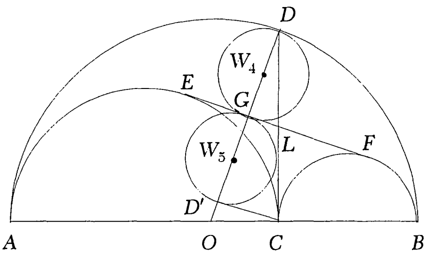


FIGURE 5
Circles 4 and 5.

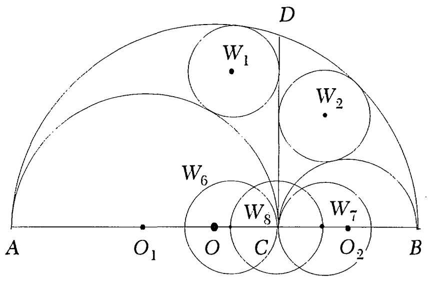


FIGURE 6
Circles 6, 7, and 8.

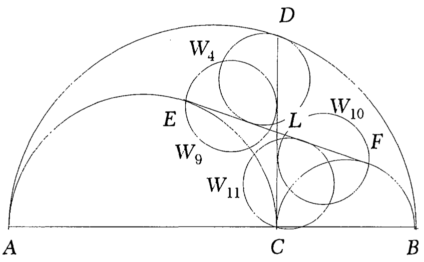


FIGURE 7
Circles 9, 10, and 11.

tangent to circles (O_1) and (W_7), for example, passes through point B and that for circles (O_1) and (W_8) passes through O_2. These properties will be examined more closely in Figure 25, near the end of this article.

Since segments CD and EF are equal and bisect each other at point L, there are three circles (W_9), (W_{10}), and (W_{11}) symmetric to (W_4), the smallest circle through D and tangent to EF. We take (W_9) to be the smallest circle through E and tangent to CD, (W_{10}) through F and tangent to CD, and (W_{11}) through C and tangent to EF. See Figure 7. Figures 5 and 7 show that circle (W_{11}) is also circle (W_5) translated through vector $\mathbf{D'C}$. Of course, circles (W_9) and (W_{10}) also translate down onto (W_6) and (W_7).

When I gave a lecture on these first eleven circles to a student group a few years ago, one of the students, Jonathan Dearing, pointed out circle (W_{12}), the circle whose diameter is the line of centers of circles (W_4) and (W_5), shown in Figure 8. Little did Archimedes realize the size of the family he uncovered! But we are not yet finished.

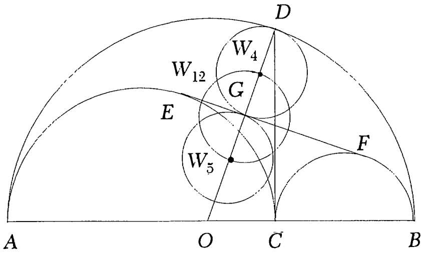


FIGURE 8
The Dearing circle 12.

Schoch's circles The development takes another turn at this point. In 1979 Martin Gardner wrote about Bankoff's triplet circle, inspiring the then student Thomas Schoch of Essen, Germany, to discover several more circles [4]. He sent his work, in German, to Gardner, who forwarded it to Bankoff, who was not familiar with German. Bankoff gave me a copy of it in 1996, when we were discussing the possibility of writing this article. Historically, then, Schoch's work precedes mine, but I shall continue the circle numbering as started above. I recognized the high quality of Schoch's paper and set out to locate him. He, still living in Essen, had not pursued his work on the circles until he found the arbelos website of Peter Woo [5] early in 1998. He then contacted Woo and told him of his findings. Paul Yiu led me to Woo, who had just completed a paper on his infinite family of Archimedean circles [6], and we all decided to combine our separate efforts into this paper.

FIGURE 9 shows Schoch's first two circles (W_{13}) and (W_{14}). They are found by drawing the circles $A(C)$, the circle with center A passing through point C, and $B(C)$ to cut circle (O) at points M_1 and M_2 respectively. Then (W_{13}) and (W_{14}) are the smallest circles through M_1 and M_2 and tangent to line CD. These circles, too, translate down to (W_6) and (W_7).

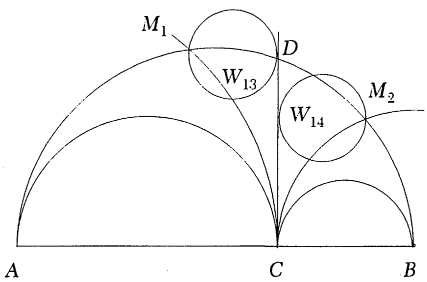


FIGURE 9
Schoch's circles 13 and 14.

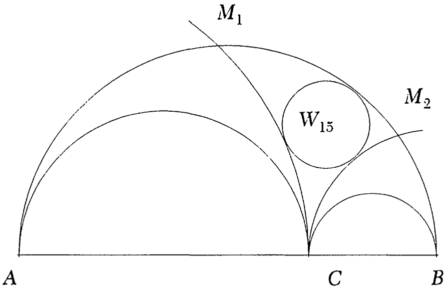


FIGURE 10
The Schoch circle 15.

Circle (W_{15}) is the incircle of curvilinear triangle CM_1M_2. See FIGURE 10. Next, drop a perpendicular $W_{15}K$ from point W_{15}, the center of circle (W_{15}), to line AB. Circle (W_{16}) is the circle centered on that perpendicular and tangent externally to circles (O_1) and (O_2), shown in FIGURE 11.

Let the line KW_{15} cut (O) at V. The smallest circle through V and tangent to the circle on O_1O_2 as diameter, which we denote by (O_1O_2), is circle (W_{17}). Let VK cut the circle (O_1O_2) at U. Then the circle through U, C, and K, denoted (UCK), that is, the circle (UC), is circle (W_{18}). See FIGURE 12.

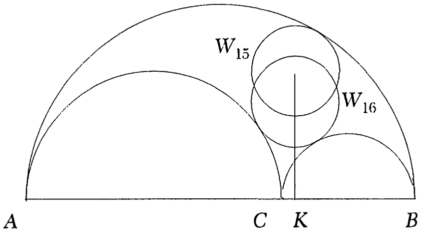


FIGURE 11
The Schoch circle 16.

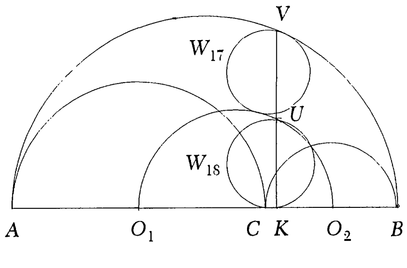


FIGURE 12
Schoch's circles 17 and 18.

If we construct on semicircle (O_1) an arbelos AC_1C similar to the given arbelos ACB, then the semicircle on C_1C as diameter is circle (W_6). Likewise we obtain circle (W_7) by constructing another similar arbelos CC_2B on CB as diameter. We let R_1, R_3, and R_4 be the highest points on circle (O_1), and on the two circles (AC_1) and (C_1C). Then $R_1R_3C_1R_4$ is a rectangle whose sides are in the ratio r_1/r_2. Similarly, $R_2R_5C_2R_6$ and RR_1CR_2 also are such rectangles, where R_2, R_5, R_6, and R are the highest points on circles (O_2), (CC_2), (C_2B), and (O). Furthermore, the lines C_1R_2, C_2R_1, and R_4R_5 all concur at a point Z on line VK. See FIGURE 13. In addition, since

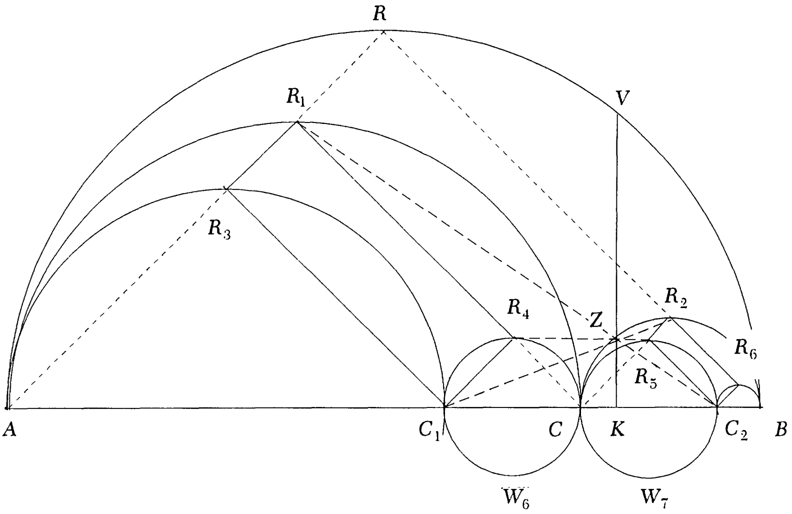


FIGURE 13
Circles 6 and 7 again.

the sides of these rectangles all make 45° angles with line AB, then points A, R_1, R_3, and R are collinear, as also are R_1, R_4, and C, and so forth. Paul Yiu [8] noted that the center W_3 of the Bankoff triplet circle lies at the intersection of the lines R_1O_2 and R_2O_1, thus providing an easy method for constructing that circle.

Locate point P on the circle on O_1O_2 as diameter so that a circle centered at P is externally tangent to both circles (O_1) and (O_2). Circle (W_{19}) is the smallest circle through point P and tangent to line AB. See FIGURE 14.

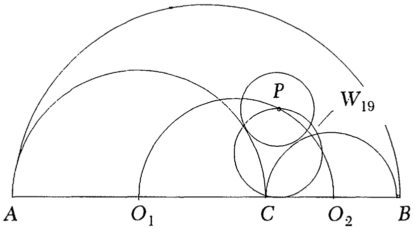


FIGURE 14
The Schoch circle 19.

Refer to FIGURE 15, where R_1, R_2, and R are the highest points on circles (O_1), (O_2), and (O) respectively. Then the circle (R') on R_1R_2 as diameter passes through O, C, and R. Let R_1R_2 cut CD at Y and OR at Q. Then the circles (RQ) and (YC) are circles (W_{20}) and (W_3). Circle (W_{21}) is the circle symmetric to (W_3) in line R_1R_2 and is tangent to circle (O) at point I, which is also the intersection of the circle (R_1R_2) and circle (O).

We note that points Z and K determine two more Archimedean circles, which we shall call (W_{22}) and (W_{23}), the circles $Z(K)$ and $K(Z)$, each centered on one of those points and passing through the other. Although Schoch did not mention these circles, he deserves the credit for them. FIGURE 16 shows these latest circles. Schoch also found the circles (W_4), (W_9), (W_{10}), and (W_{11}).

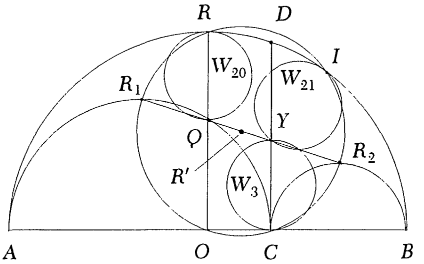


FIGURE 15
Schoch's circles 20 and 21.

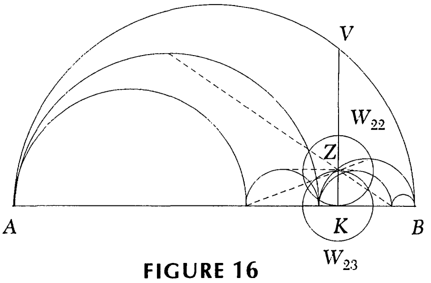


FIGURE 16
Circles 22 and 23.

Some loose ends In assembling proofs for all these circles, I observed three additional Archimedean circles, (W_{24}), (W_{25}), and (W_{26}). Circles (W_{24}) and (W_{25}) are centered at R_4 and R_5 respectively and pass through points W_6 and W_7 respectively. They are shown in FIGURE 17. FIGURE 18 shows circle (W_{26}), the circle symmetric to circle (W_{20}) in line R_1R_2 and also symmetric to circle (W_{21}) in the center of circle $(OCIR)$. If CD cuts circle (R') again at C', then C', Q, and I' are collinear, as are also O, Y, and I. Finally, Schoch found one other circle (W_{27}), the smallest circle through point C and tangent to his circle (W_{15}), shown in FIGURE 19.

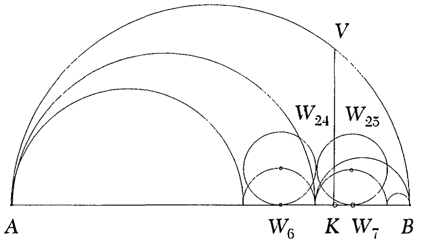


FIGURE 17
Circles 24 and 25.

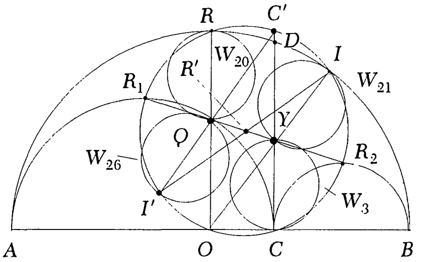


FIGURE 18
Circle 26.

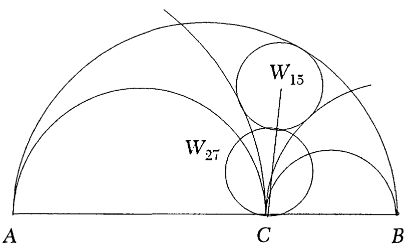


FIGURE 19
The Schoch circle 27.

In November of 1996 Paul Yiu [7] wrote a letter to Bankoff stating that he had just that morning discovered the circle I have called (W_{11}), adding "Maybe you have already known this. But isn't it wonderful?" He noted that its center is at the intersection of O_1F and O_2E [8].

Woo's circles Peter Woo discovered an infinite family of Archimedean circles centered on line KV shown in FIGURES 11 and 12, which he called the *Schoch line*. In FIGURE 10 the Schoch circle (W_{15}) is the incircle of the curvilinear triangle CM_1M_2. The two circles $A(C)$ and $B(C)$, which pass through point C, whose centers lie on AB, and whose radii are twice the radii of circles (O_1) and (O_2) respectively, determine the arcs CM_1 and CM_2. Woo generalized this idea by using any positive multiple n instead of 2, leaving the circles to still pass through C, but moving their centers along line AB. Thus, as shown in FIGURE 20, draw two semicircles (O_n') and (O_n''), each tangent to line CD at point C, with centers O_n' on ray CA and O_n'' on ray CB, and with radii n times the radii of circles (O_1) and (O_2) respectively. Thus we call n the *radius multiplier*. Then the circle (U_n) with radius equal to that of the twin circles and tangent to (O_n') and (O_n'') will surprisingly have its center on the Schoch line. Conversely, any circle (U_n) with twin circle radius and centered sufficiently high up on the Schoch line will be tangent to two such circles (O_n') and (O_n'') for some positive real number n. The Woo circles are a generalization of Schoch's circles (W_{15}) and (W_{16}), also shown in FIGURES 10 and 11. FIGURE 24 shows selected Woo circles: (U_1) tangent to (O_1) and (O_2), $(U_2) = (W_{15})$ tangent to $A(C)$ and $B(C)$, (U_4), (U_7), and the limiting case $(U_0) = (W_{11})$.

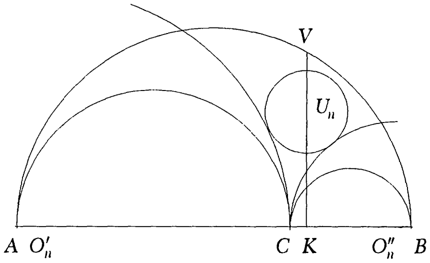


FIGURE 20
A typical Woo circle.

Yiu's second circle When Paul Yiu read Woo's paper, he noted that circle $(W_{15}) = (U_2)$ was tangent internally to circle (O) and observed that there has to be a Woo circle (U_n) that is tangent to (O) externally. He proved that this circle, which we designate as (W_{28}), touches (O) at point D [8]. See FIGURE 27. He commented to me that "Archimedean circles start escaping the shoemaker's knife."

The proofs We now present proofs of some of our assertions. Let the radii and diameters of the circles (O), (O_1), and (O_2) be r and d, r_1 and d_1, and r_2 and d_2 respectively. Then, of course, $r = r_1 + r_2$ and $d = d_1 + d_2$. Let us denote the radius of each circle (W_i) by p_i. Although we shall not prove it, it is helpful in working with circles (W_{15}) and (W_{16}) and any of Woo's circles to know that

$$CK = \frac{r_1 r_2 (r_1 - r_2)}{(r_1 + r_2)^2}.$$

We shall need the fact that DD' of FIGURE 5 is equal to $2d_1d_2/(d_1 + d_2)$, the harmonic mean of the diameters of circles (O_1) and (O_2), so let us first display a delightful figure that shows this fact, along with some other means and their well-known relationship to one another. In the arbelos shown in FIGURE 21, OR is that radius of circle (O) that is perpendicular to the common diameter ACB. We use the notation above and that given in [3] for the means.

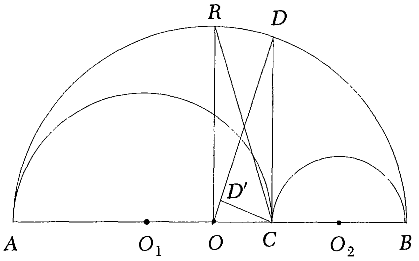


FIGURE 21
A mean figure.

THEOREM 1. *In* FIGURE 21 *we have that*

(i) *CR is the root-mean-square* $M_2(d_1, d_2)$ *of AC and CB.*
(ii) $OR = OD$ *is their arithmetic mean* $M_1(d_1, d_2)$.
(iii) *CD is their geometric mean* $M_0(d_1, d_2)$.
(iv) *DD′, where D′ is the foot of the perpendicular dropped from point C to OD, is their harmonic mean* $M_{-1}(d_1, d_2)$.
(v) *Finally,*

$$d_1 \geq M_2 \geq M_1 \geq M_0 \geq M_{-1} \geq d_2.$$

Furthermore, these inequalities are all strict when $d_1 \neq d_2$.

Proof. Of course, $AC = d_1$, $CB = d_2$, and $AB = d$, so $OD = OR = d/2 = (d_1 + d_2)/2 = M_1$. Now $OC = OB - CB = d/2 - d_2 = (d_1 - d_2)/2$, so by the Pythagorean theorem we have

$$CD = \sqrt{\left(\frac{d_1}{2} + \frac{d_2}{2}\right)^2 - \left(\frac{d_1}{2} - \frac{d_2}{2}\right)^2} = \sqrt{d_1 d_2} = M_0.$$

By similar right triangles, $DD'/CD = CD/OD$ and hence

$$DD' = \frac{CD^2}{OD} = \frac{d_1 d_2}{\frac{1}{2}(d_1 + d_2)} = \frac{2 d_1 d_2}{d_1 + d_2} = M_{-1}.$$

In the desired inequality we see that each internal member is the hypotenuse of a right triangle in which the next member is a leg. Thus triangle COR shows that $M_2 \geq M_1$, triangle OCD shows that $M_1 \geq M_0$, and triangle $CD'D$ produces $M_0 \geq M_{-1}$. Now

$$d_1 = AC = AO + OC = RO + OC \geq RC = M_2$$

and

$$M_{-1} = DD' = OD - OD' = OB - OD' \geq OB - OC = CB = d_2.$$

Equality occurs when and only when all three right triangles reduce to the same straight line segment, that is, when point C coincides with point O, when $d_1 = d_2$. □

Although Theorems 2 and 3 are readily and cleverly proved by inversion, as Bankoff showed in [1], our proof will be by high school geometry.

THEOREM 2. *The radii p_1 and p_2 of circles (W_1) and (W_2) are equal to half the harmonic mean of r_1 and r_2. We denote this common value by p. That is,*

$$p_1 = p_2 = p = \frac{r_1 r_2}{r_1 + r_2} = \frac{r_1 r_2}{r}.$$

Proof. Draw O_1W_1 and OW_1 and drop perpendiculars from W_1 to line CD and to point Q on AB, as shown in FIGURE 22. Then

$$O_1W_1 = r_1 + p_1, \quad OW_1 = r - p_1 = r_1 + r_2 - p_1,$$
$$O_1Q = r_1 - p, \quad \text{and} \quad OQ = r_1 - r_2 - p_1.$$

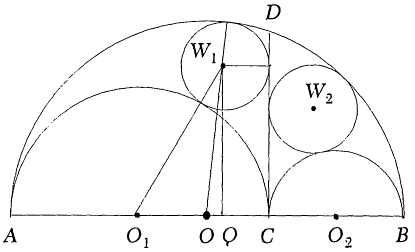


FIGURE 22
The twin circles.

From right triangles O_1W_1Q and OW_1Q we get that

$$QW_1^2 = (r_1 + p_1)^2 - (r_1 - p_1)^2 = (r_1 + r_2 - p_1)^2 - (r_1 - r_2 - p_1)^2,$$

which reduces to

$$4r_1 p_1 = 4(r_1 - p_1)r_2 \text{ and hence } p_1 = \frac{r_1 r_2}{r_1 + r_2} = p.$$

A similar argument shows that $p_2 = p$. □

THEOREM 3. *The radius p_3 of circle (W_3) is equal to p.*

Proof. Let r_3 denote the radius of circle (O_3), h the length of the perpendicular O_3H from O_3 to diameter ACB, and let $x = OH$. See FIGURE 23. For convenience we let $r = r_1 + r_2 = 1$, so that $p = r_1 r_2$. Then, from the three right triangles OO_3H, O_1O_3H, and O_2O_3H we obtain the three equations

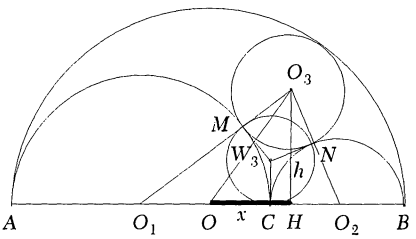


FIGURE 23
Triplet circle proof.

$$x^2 + h^2 = (r_1 + r_2 - r_3)^2, \ (r_2 + x)^2 + h^2 = (r_1 + r_3)^2, \text{ and } (r_1 - x)^2 + h^2 = (r_2 + r_3)^2.$$

Subtract the first equation from each of the other two, obtaining

$$2r_2^2 + 2r_2 x = -2r_1 r_2 + 4r_1 r_3 + 2r_2 r_3$$

and

$$2r_1^2 - 2r_1 x = -2r_1 r_2 + 2r_1 r_3 + 4r_2 r_3.$$

Now multiply the first of these two equations by r_1 and the second by r_2, and then add the resulting equations to get

$$4r_1r_2^{\,2} + 4r_1^{\,2}r_2 = 4r_1^{\,2}r_3 + 4r_2^{\,2}r_3 + 4r_1r_2r_3,$$

which we solve for r_3, finding that

$$r_3 = \frac{(r_1 + r_2)r_1r_2}{r_1^{\,2} + r_2^{\,2} + r_1r_2} = \frac{r_1r_2}{1 - r_1r_2}.$$

The sides of triangle $O_1O_2O_3$ have lengths $O_1O_2 = r_1 + r_2 = 1$, $O_3O_1 = r_3 + r_1$, and $O_2O_3 = r_2 + r_3$, so its semiperimeter is $1 + r_3$. By Heron's formula, its area K is given by

$$K^2 = (1 + r_3)r_1r_2r_3.$$

Since (W_3) is the incircle for that triangle, we also have

$$K = (\tfrac{1}{2})(r_1 + r_2)p_3 + (\tfrac{1}{2})(r_3 + r_1)p_3 + (\tfrac{1}{2})(r_2 + r_3)p_3 = (1 + r_3)p_3.$$

Equating the two expressions for K^2, we get that

$$(1 + r_3)r_1r_2r_3 = (1 + r_3)^2p_3^{\,2},$$

so that

$$p_3^{\,2} = \frac{r_1r_2r_3}{1 + r_3} = \frac{\dfrac{r_1^2r_2^2}{1 - r_1r_2}}{1 + \dfrac{r_1r_2}{1 - r_1r_2}} = r_1^2r_2^2.$$

Hence $p_3 = r_1r_2 = p$. It can be shown that $h = 2r_3$, an example of one of the delightful theorems presented in [**2**]. □

Woo's Theorem. *For any positive number n, draw two semicircles (O_n') and (O_n''), each tangent to line CD at point C, with centers O_n' on ray CA and O_n'' on ray CB, and with radii r_1n and r_2n respectively. Then the circle (U_n) with radius equal to that of the twin circles and externally tangent to (O_n'') and (O_n'') will have its center on the Schoch line. Conversely, any circle U_n with twin circle radius and centered on the Schoch line above height $2r_1r_2\sqrt{r_1r_2}\,/(r_1 + r_2)^2$ will be tangent to two such circles (O_n') and (O_n'') for some nonnegative real number radius multiplier n. See* Figure 24.

Proof. Let C be the origin, ray CB the x-axis, and ray CD the y-axis. Choose the unit of length so that $r_1 + r_2 = 1$ and let the center of (U_n) have coordinates (x, y).

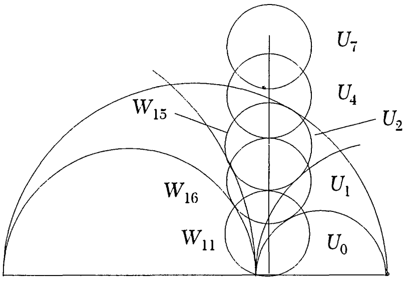


FIGURE 24
Selected Woo circles.

The radius of (U_n) is half the harmonic mean of r_1 and r_2, which is r_1r_2 because $r_1 + r_2 = 1$. Then we have

$$O_n'U_n^2 - O_n''U_n^2 = (nr_1 + r_1r_2)^2 - (nr_2 + r_1r_2)^2 = (nr_1 + x)^2 - (nr_2 - x)^2,$$
$$2nr_1r_2(r_1 - r_2) = 2n(r_1 + r_2)x,$$

and finally,

$$x = r_1r_2(r_1 - r_2),$$

which proves that U_n lies on the Schoch line. One can apply the Pythagorean theorem to triangle CKW_{11} to establish the minimum height requirement, and the converse is established. □

Some additional properties Now we cut short our proofs, having illustrated the techniques by which all circles can be shown to have the same radius. We conclude by stating a few more of the properties that these circles possess and locating one more circle.

Let T_i be the point of contact for circles (O_i) and (W_i) for $i = 1, 2$. Then $BD = BT_1$ and BT_1 is tangent to circles (O_1) and (W_1). Similarly, $AD = AT_2$ and AT_2 is tangent to circles (O_2) and (W_2). See FIGURE 25.

Earlier we stated that (W_6) and (W_7) were more than just translations of the twin circles onto the common diameter AB. We have seen that they are such translations also for (W_9) and (W_{10}), and for (W_{13}) and (W_{14}). Furthermore, as noted by Schoch and seen in FIGURE 13, they are semicircles of the inscribed similar arbelos. Also, FIGURE 26 shows that circle (W_6) is tangent to line AT_2, and circle (W_8) is tangent to the line O_2S_1 drawn from O_2 tangent to circle (O_1). Similarly circle (W_7) is tangent to line BT_1, and circle (W_8) is tangent also to the line O_1S_2 drawn from O_1 tangent to circle (O_2). That is (W_6) is the circle through point C with center lying on segment AC and tangent to the line AT_2 and (W_7) is the circle through point C with center lying on segment BC and tangent to the line BT_1. Finally, (W_8) is the circle centered at point C and tangent to the two lines O_1S_2 and O_2S_1.

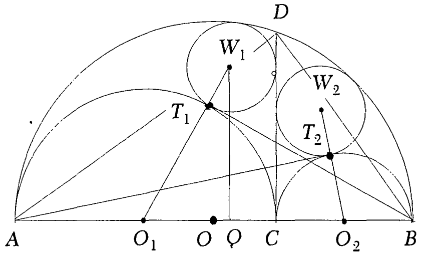


FIGURE 25
Twin circle tangents.

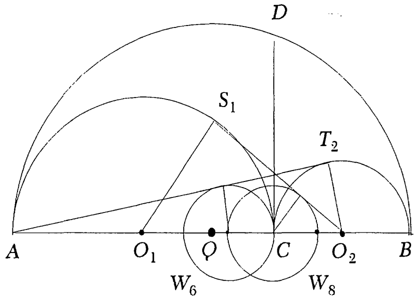


FIGURE 26
Tangents.

Our last circle is another Schoch circle. As shown in FIGURE 27, circles (W_5), (W_{12}), (W_4), and the second Yiu circle (W_{28}), that is, the Woo circle that is tangent externally to circle (O) at point D, all have centers that lie on line OD. Schoch discovered circle

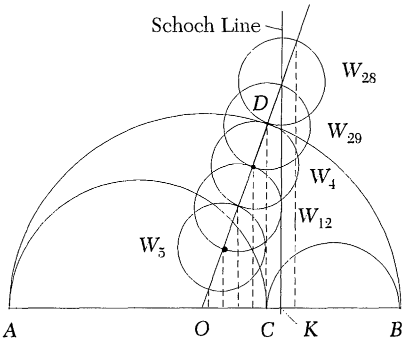


FIGURE 27
Yiu's circle 28 and Schoch's circle 29.

(W_{29}), the Archimedean circle centered at D, which has W_4W_{28} as diameter. Furthermore, the centers of all five of these circles project onto points on the base diameter AB that are spaced distance CK apart from one another, as shown by the dotted lines in the figure.

As a Woo circle (U_n), the second Yiu circle (W_{28}) has the value for its radius multiplier n given by

$$n = 2 + \frac{(r_1 + r_2)^2}{r_1 r_2}.$$

Conclusion Archimedes is credited with finding two delightful congruent circles in the arbelos, and Leon Bankoff opened a door by finding his triplet circle and later the quadruplet circle. Inspired by these masters, we have dramatically extended that family of circles. Although we have not stated as theorems and proved every property we indicated in our opening paragraphs, we have illustrated how to show that the circles (W_1) through (W_{29}) and the infinite family of Woo circles are Archimedean circles and do possess the claimed characteristics. We have achieved our goal of demonstrating that the twin circles of Archimedes are only two members of a huge family, in fact an infinite family, of congruent circles, all neatly hidden in that simple arbelos. When next you have new heels put on your shoes, you might describe some of these curious circles to your local cobbler.

REFERENCES

1. Leon Bankoff, Are the twin circles of Archimedes really twins? this MAGAZINE 47 (1974), 214–218.
2. Leon Bankoff, The marvelous arbelos, *The Lighter Side of Mathematics*, Proceedings of the Eugène Strens Memorial Conference on Recreational Mathematics & its History, edited by Richard K. Guy and Robert E. Woodrow, The Mathematical Association of America, Washington, DC, 1994, 247–253.
3. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge University Press, Cambridge, UK, 1934.
4. Thomas Schoch, Constructions and proofs, unpublished manuscript, 1979.
5. P. Y. Woo, The Arbelos, An Example of What we Teach, Web page with interactive color applet at `http://www.biola.edu/academics/undergrad/math/woopy/arbelos.htm`, 1997.
6. P. Y. Woo, Do arbelos twin circles grow like flowers?, unpublished manuscript, 1998.
7. Paul Yiu, private correspondence to Leon Bankoff, November 19, 1996.
8. Paul Yiu, The Archimedean circles in the shoemaker's knife, lecture at the 31st annual meeting of the Florida Section of the Mathematical Association of America, Boca Raton, FL, March 6–7, 1998.

Should She Switch? A Game-Theoretic Analysis of the Monty Hall Problem

LUIS FERNANDEZ
ROBERT PIRON
Economics Department
Oberlin College
Oberlin, OH 44074

Introduction The Monty Hall problem remains durable even after years of controversy ([4], [5], [6], [7], [8]). A contestant, say Elaine, stands before three curtains on the stage. Elaine is told that behind one curtain is a brand-new car; behind each of the remaining curtains hides a goat; and the car is equally likely to lie behind any curtain. Elaine then chooses one curtain, but before opening that curtain the master of ceremonies, Monty Hall, opens another curtain, revealing a goat. He then offers Elaine the chance to switch her choice to the remaining unopened curtain.

Should Elaine switch?

Our purpose is not to rehash this well-worn problem, but to show how some recent theoretical advances in game theory allows us to model this situation more realistically than has been done heretofore.

The Monty Hall problem has been modeled decision-theoretically by assuming that Monty Hall follows fixed rules, like a blackjack dealer, and that the only decision-maker is Elaine. With this assumption, the question posed above has a clear answer: Elaine should always switch curtains. In an interview ([5]), however, the real Monty Hall argued that the answer is not so simple. Had it been so, the show would have been too predictable and would soon have been pulled off the air. In reality, Monty Hall was an active, unpredictable player, with his own objectives. Sometimes he would not allow a contestant to change curtains; sometimes he offered players cash on the spot if they would ([5]). Modern game theory can be used to model the strategic situation that faced real contestants on the real game show.

The game The Monty Hall game is diagrammed in FIGURE 1. Technically, FIGURE 1 shows what happens if Elaine initially chooses curtain 2. (That is why the arrows that correspond to other choices at nodes B, C, and D have white heads.) The omitted parts of the game are essentially the same as what is shown. Each rectangular box in the diagram represents a decision node, i.e., a moment in the game at which a player must act. The decision nodes are labeled A–H. Each arrow represents a possible move by a player. Decision nodes enclosed within a dotted oval form an information set for a player. Elaine has seven information sets, three of which appear in FIGURE 1, labeled I, II, and III. When Elaine reaches one of her information sets, she does not know which of the nodes in the set has been reached. Monty knows where the car is. This fact is captured in FIGURE 1 by the lack of dotted ovals around any of his decision nodes. Monty has ten possible decision nodes, only four of which are displayed in FIGURE 1.

Monty Hall first hides the car behind one of the curtains (node A). Elaine next chooses a curtain (nodes B, C, or D). For simplicity, FIGURE 1 assumes that Elaine chooses curtain 2. Monty then opens one of the three curtains (nodes E, F, or G). If he opens either Elaine's curtain or the curtain hiding the car, then the game ends. But if he opens a curtain hiding a goat, the car's location remains unknown. At this point

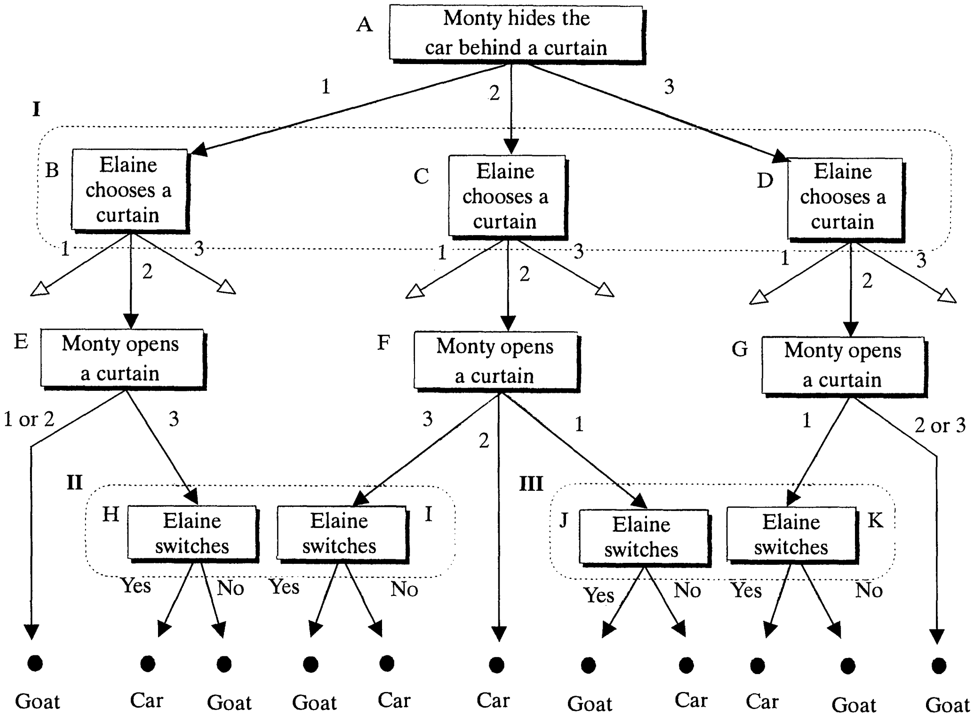


FIGURE 1
The Monty Hall game

Monty can allow Elaine to make her second and final move: she can stay with her initial choice or she can switch to the remaining unopened curtain (nodes *H*, *I*, *J*, or *K*). The game then ends with two possible outcomes: Elaine gets either a new car or a goat. Though this game seems simple, the game tree is rather complicated. Depending on the players' moves, the game can evolve along 39 paths (of which 13 appear in FIGURE 1).

The solution To solve the Monty Hall game we will find its *Nash equilibrium*. A Nash equilibrium is a pair of strategies, one for each player, such that each player's strategy is optimal given the strategy of the other player. (For a discussion of Nash equilibrium, see [1].) To find the Nash equilibrium, we first describe each player's objective. Elaine's objective is obvious: she wants to win the car. Monty's objective is less transparent, but we conjecture that he wants to maximize the number of viewers. Other shows—not Elaine—were Monty's competitors. The show's rules and Monty's behavior are designed to make the show entertaining. We conjecture that the show is more entertaining the harder it is for the audience to predict the outcome.

One way to model this idea is to add a third player—the television audience. While the audience does not move in the game, it does try to predict Monty and Elaine's moves. Monty's payoff is a function of the accuracy of these predictions. Games of this type are called *psychological games*; they differ from normal games in that the players' payoffs can depend not only on what the players do, but also on what they believe others will do. A *psychological Nash equilibrium* is a collection of pairs of strategies and beliefs, one pair for each player, such that the collection of strategies form a Nash equilibrium and the beliefs are consistent with the players' strategies and the laws of probability. A more precise definition of this equilibrium concept can be found [3]. A psychological Nash equilibrium exists under the same conditions as does an ordinary Nash equilibrium in conventional non-cooperative game theory.

The game is easiest to predict if it always results in the same outcome—and hence is uninteresting—and it is hardest to predict if each of the 39 possible decision sequences is equally likely. To this end, clearly, Monty should hide the car behind each curtain with equal probability. Suppose Elaine chooses curtain 2. Now Monty has to decide whether to open a curtain and allow Elaine to switch. He wants to do this in such a way that Elaine will be indifferent between switching and not switching, for then she will be unpredictable. She is indifferent between switching and not switching if and only if the conditional probability that the car is behind curtain 2 given that Monty allows her to switch equals the conditional probability that the car is behind the unopened curtain given that Monty allows her to switch.

Let C_2 denote the event "Monty hides the car behind curtain 2," let C_{13} denote the event "Monty hides the car behind curtains 1 or 3," let M denote the event "Monty allows Elaine to switch curtains," and let $P\{X|Y\}$ denote the conditional probability that the event X occurs given that Y occurs. Also let $P_2 = P\{M|C_2\}$ and let $P_{13} = P\{M|C_{13}\}$. Then Bayes' theorem implies

$$\frac{P\{C_2|M\}}{P\{C_{13}|M\}} = \frac{P\{M|C_2\}P\{C_2\}}{P\{M|C_{13}\}P\{C_{13}\}} = \frac{P_2\cdot\frac{1}{3}}{P_{13}\cdot\frac{2}{3}} = \frac{P_2}{2P_{13}}.$$

Thus Elaine is indifferent between switching and staying with her initial choice if and only if $P_2 = 2P_{13}$. That is, Monty should be twice as likely to let a contestant switch curtains when she initially chooses the curtain hiding the car as when she chooses a curtain hiding a goat.

The game has six possible outcomes, which are displayed in Table 1. Table 1 also gives the probabilities of these outcomes as functions of three parameters P_E, P_{13}, and P_2, where P_E is the probability that Elaine switches curtains if given the chance.

TABLE 1. Possible outcomes and probabilities as functions of P_E, P_{13}, and P_2.

Outcome	Car behind curtain	Monty allows her to switch?	Elaine	Elaine wins	Probability
1	1 or 3	No	—	Goat	$p_1 = \frac{2}{3}(1 - P_{13})$
2	1 or 3	Yes	Switches	Car	$p_2 = \frac{2}{3}P_{13}P_E$
3	1 or 3	Yes	Stands pat	Goat	$p_3 = \frac{2}{3}P_{13}(1 - P_E)$
4	2	No	—	Car	$p_4 = \frac{1}{3}(1 - P_2)$
5	2	Yes	Switches	Goat	$p_5 = \frac{1}{3}P_2P_E$
6	2	Yes	Stands pat	Car	$p_6 = \frac{1}{3}P_2(1 - P_E)$

Elaine's behavior, as encapsulated by the parameter P_E, would seem to be an exogenous behavioral parameter that Monty has to take as given. In fact, Monty has years of experience with contestants, and, therefore, has developed tricks with which to manipulate the contestants' behavior. For example, if Monty senses that Elaine's P_E parameter value is below the optimal level (whose value we have not yet determined), then he can drive up P_E by offering Elaine cash to switch. Similarly, Monty can lower P_E by offering cash for not switching. We will therefore treat all three parameters P_E, P_{13}, and P_2 as decision variables under Monty's control. It remains only to determine their optimal values.

The game is least predictable when each of the six possible outcomes is equiprobable. Such perfect equiprobability cannot be attained. For example, outcomes 5 and 6 are equally likely if and only if $P_E = \frac{1}{2}$. But then $p_6 = \frac{1}{3}P_2(1 - P_E) = \frac{1}{6}$ if and only if $P_2 = 1$, which implies that $p_4 = \frac{1}{3}(1 - P_2) = 0$. Instead, Monty has to settle for getting "close" in terms of some metric. We will use Euclidean distance. We suppose, then, that Monty selects P_E, P_{13}, and P_2 to minimize the total "distance" between the six outcomes probabilities and the (infeasible) *optimum optimorum* of perfect equiprobability. Monty is minimizing

$$\sum_{i=1}^{6}\left(p_i - \frac{1}{6}\right)^2$$

where p_i is the probability that outcome i occurs, as shown in Table 1, and P_2 and P_{13} satisfy the constraint $P_2 = 2P_{13}$. The optimal values can be found using calculus and are $P_E = \frac{1}{2}$, $P_2 = 1$, and $P_{13} = \frac{1}{2}$. So, Monty always allows Elaine to switch curtains when her initial curtain hides the car, but only allows her to switch half the time when her initial curtain hides a goat. When he does allow Elaine to switch, he manipulates her so that she switches half the time. With the three parameters at their optimal values the probabilities of the six outcomes are: $p_1 = \frac{1}{3}$, $p_2 = p_3 = \frac{1}{6}$, $p_4 = 0$, $p_5 = p_6 = \frac{1}{6}$.

Conclusion Why have so many people, including mathematicians, come to such different conclusions about the Monty Hall problem? One possibility is that the concept of conditional probability is poorly understood. Another explanation is that the precise game being played has not been articulated clearly. Careful game-theoretic modeling may help resolve the seemingly endless debate.

Acknowledgment Our thanks to our colleagues Jeff Witmer, Sam Goldberg, and Richard Salter for their helpful comments and suggestions.

REFERENCES

1. H. Scott Bierman and Luis Fernandez, *Game Theory with Economic Applications*, 2nd Edition, Addison-Wesley-Longman, Reading, MA, 1999.
2. Avinnash Dixit and Barry Nalebuff, *Thinking Strategically: The Competitive Edge in Business, Politics, and Everyday Life*, W. W. Norton and Company, New York, NY, 1991.
3. John Geanakoplos, David Pearce, and Ennio Stacchetti, Psychological games and sequential rationality, *Games and Economic Behavior* 1 (1991), 60–79.
4. Leonard Gillman, The 'car' and the 'goats,' *Amer. Math. Monthly* 99 (1992), 3–7.
5. John Tierney, Behind Monty Hall's doors: puzzle, debate and answer, *The New York Times* (July 21, 1991), 1.
6. Marilyn vos Savant, Ask Marilyn, *Parade*, September 9, 1990.
7. Marilyn vos Savant, Ask Marilyn, *Parade*, December 2, 1990.
8. Marilyn vos Savant, Ask Marilyn, *Parade*, February 17, 1991.

Nice Polynomials for Introductory Galois Theory

BARBARA L. OSOFSKY
Rutgers University
Piscataway, NJ 08854-8019

Dick and Jane are students in an advanced undergraduate modern algebra course. They are studying Galois theory. To help them we write down a family of polynomials $\{f_n(x): n > 1\}$ where $f_n(x)$ has degree n. We then show them how to check, using standard methods that they have used before, that the polynomials $f_p(x)$ for primes p have Galois group the symmetric group on p letters. Let's see how.

Dick and Jane are fortunate to have had a good high school background in mathematics. They have learned about prime factorization of integers since grade school. They know that a polynomial $f(x) = \sum_{i=0}^{n} c_i x^i$ of degree n with rational coefficients has a rational root α if and only if the polynomial $(x - \alpha)$ is a factor of $p(x)$; that is, $f(\alpha) = 0 \Leftrightarrow f(x) = (x - \alpha)g(x)$ for some polynomial $g(x)$ with rational coefficients. They also know that if $f(x)$ is a nonconstant polynomial with integer coefficients, then any rational root $\frac{p}{q}$ of $f(x)$, where p and q are relatively prime integers, has the property that p is a factor of the constant term c_0 and q is a factor of the leading coefficient. And they have used the intermediate value theorem, which states that a continuous function, such as a polynomial, must have a zero between any two points where the value of the function takes on different signs. In calculus they learned that a polynomial of degree n with real coefficients can have at most $n - 1$ points where it turns from increasing to decreasing or vice versa, and that at least one such point must lie between any pair of zeros. In their algebra course, if not earlier, they have been told that there are formulas for the roots of any polynomial of degree up to 4, and that there are no such formulas for polynomials of degree 5 or larger over the rationals. They may even have seen an example of a polynomial of degree 5 with integer coefficients that cannot be solved by any formula involving only arithmetic operations and the taking of roots. The polynomials which we define here, $f_p(x)$ of prime degree $p > 3$, show Dick and Jane how to get such polynomials with arbitrarily large degree.

One thing Dick and Jane see in their algebra course is how to get irreducible polynomials over the rational numbers. To show a polynomial of degree 5 is irreducible, one need only check that it has no linear factors and no quadratic factors with rational coefficients. But that will not do for larger degree. So they need some more algebraic results, here listed with hints so that they can do the proofs as exercises.

THEOREM 1. (Eisenstein's Criterion) *Let $f(x) = \sum_{i=0}^{n} c_i x^i$ be a polynomial with integer coefficients. Let p be a prime integer with the property that $p | c_i$ for $0 \leq i \leq n - 1$, but $p^2 \nmid c_0$ and $p \nmid c_n$. Then $f(x)$ is irreducible over the integers.*

Hint: Assume $f(x)$ is a product of two polynomials with integer coefficients. The constant term of precisely one of those polynomials is divisible by p. Show by induction that all of the coefficients of that factor are divisible by p, thereby deriving a contradiction.

THEOREM 2. (Gauss' Lemma) *Let $f(x) = \sum_{i=0}^{n} c_i x^i$ be a polynomial with integer coefficients. Assume $f(x) = g(x)h(x)$, where $g(x) = \sum_{j=0}^{m} a_j x^j$ and $h(x) = \sum_{l=0}^{n-m} b_l x^l$ are nonconstant polynomials with rational coefficients. Then $f(x)$ is a product of two nonconstant polynomials with integer coefficients.*

Hint: Take the factorization over the rationals and express each factor as a polynomial with integer coefficients divided by some integer. In each of the polynomials with integer coefficients, factor out the largest integer that divides all of the coefficients. Show that the product of the resulting polynomials has no prime dividing all of the coefficients, and use this to deduce the theorem.

We can now show Dick and Jane an interesting family of polynomials that are irreducible over the rationals.

Let n be an integer greater than 1. Let $g_n(x)$ be the polynomial

$$g_n(x) = x^2 \prod_{i=1}^{n-2} (x - 2i) x^2(x-2)(x-4) \cdots (x - 2(n-2)).$$

Then $g_n(x)$ has integer coefficients, leading coefficient 1, and degree n. We also know the roots of $g_n(x)$. It has a double root at the origin and single roots at each of the first $n-2$ positive even integers. Clearly, this is not an irreducible polynomial.

If x is close to but not equal to 0, $g_n(x)$ is a product of the positive number x^2 and the $n-2$ negative numbers $x - 2i$ for $1 \leq i \leq n-2$. This product will be negative if n is odd, and positive if n is even. We will modify $g_n(x)$ by moving the local extremum at $(0,0)$ below the x-axis if n is odd, and above the x-axis if n is even. Every integer $n \geq 2$ is of the form $2k$ or $2k+1$ for some $k > 1$, so we can get $f_n(x)$ by

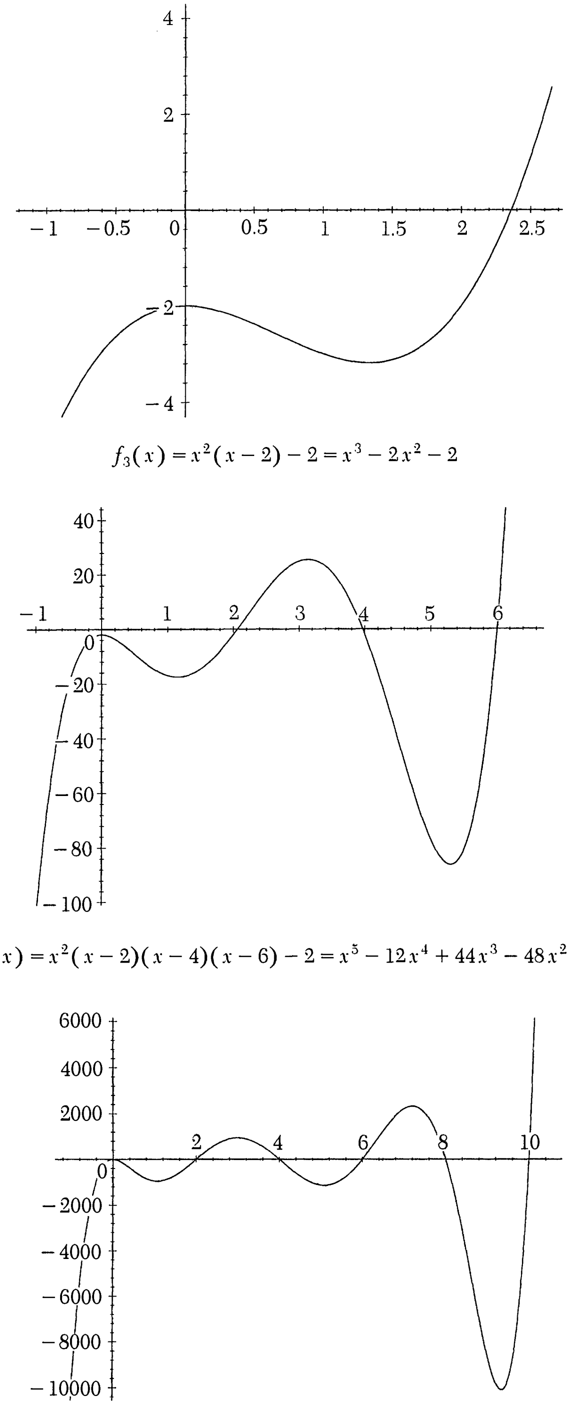
$$f_{2k+1}(x) = g_{2k+1}(x) - 2 = x^2(x-2)(x-4) \cdots (x - 2(2k-1)) - 2$$
$$f_{2k}(x) = g_{2k}(x) + 2 = x^2(x-2)(x-4) \cdots (x - 2(2k-2)) + 2.$$

Now note that every coefficient of $f_n(x)$ except the first is divisible by 2 and the constant term is not divisible by 4. By Eisenstein's criterion, $f_n(x)$ does not factor over the integers, and by Gauss' lemma, $f_n(x)$ does not factor over the rationals. We plot some of these polynomials on intervals containing their roots. Because we are ultimately interested in large prime values of n, we will look first at odd primes (see FIGURE 1). We select the scaling on the y-axis to show the entire graph over our domain intervals.

The graphs for even n would look similar away from the origin, but be positive near the origin. For example, FIGURE 2 shows the case $n = 6$.

These graphs seem to show precisely $n-2$ distinct real roots, since we know that near zero the graph actually lies above or below the x-axis even if the scaling obscures that for $n = 6$ and $n = 7$. We now show Dick and Jane that what the graphs seem to show is indeed correct.

If n is odd, then $g_n(1) < 0$ since it is a product of a square times $n-2$ negative integers, and $f_n(1) = g_n(1) - 2$ is also negative. If n is even, $g_n(1)$ is a product of a square times a product of an even number of negative integers. Thus both $g_n(1)$ and $f_n(1) = g_n(1) + 2$ are positive. For any $n > 2$, what happens when we evaluate $g_n(x)$ at an odd integer $x > 1$? We get x^2 times a nonzero integer. Hence, at each point in $\{x = 2i + 1 : 1 \leq i \leq n-2\}$, $|g_p(x)| \geq (2i+1)^2 > 2$, so adding or subtracting 2 to get $f_n(x)$ cannot cause a sign change. Thus $f_n(2i+1)$ and $g_n(2i+1)$ have the same sign for $0 \leq i \leq n-2$. Since $g_n(x)$ changes sign between $x = 2i - 1$ and $x = 2i + 1$ for

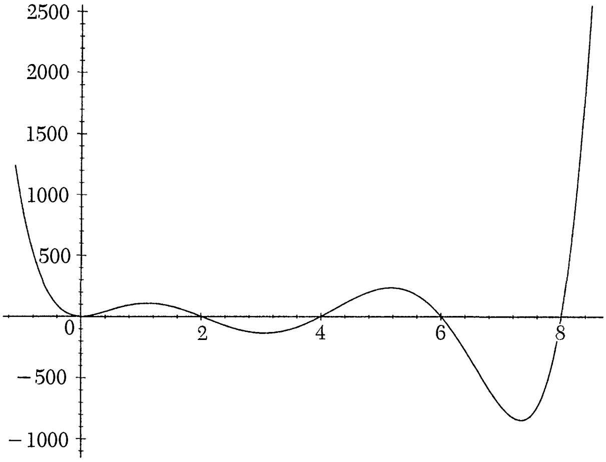


$f_3(x) = x^2(x-2) - 2 = x^3 - 2x^2 - 2$

$f_5(x) = x^2(x-2)(x-4)(x-6) - 2 = x^5 - 12x^4 + 44x^3 - 48x^2 - 2$

$$f_7(x) = x^2(x-2)\cdots(x-10) - 2$$
$$= x^7 - 30x^6 + 340x^5 - 1800x^4 + 4384x - 3840x^2 - 2$$

FIGURE 1
Graphs of $f_n(x)$ for small odd primes.



$$f_6(x) = x^2(x-2)(x-4)(x-6)(x-8) + 2$$
$$= x^6 - 20x^5 + 140x^4 - 400x^3 + 384x^2 + 2$$

FIGURE 2
The graph of f_n for an even value of n.

$1 \leq i \leq n-2$, so must $f_n(x)$. By the intermediate value theorem, $f_n(x)$ has a real zero between $x = 2i - 1$ and $x = 2i + 1$ for $1 \leq i \leq n - 2$. This accounts for $n - 2$ real roots. We note also that $f_n(x)$ does not change sign to the left of $x = 1$, and is positive to the right of $2n - 3$. Since $f_n(x)$ turns from increasing to decreasing at the same x-values where $g_n(x)$ does, $f_n(x)$ can have no real roots other than the $n - 2$ already found.

Why is that significant? Well, Dick and Jane have most likely seen a little about symmetric groups in their algebra course. They have learned elementary group theory from a standard text such as [**1**] or [**2**]. They know cycle notation for the symmetric group, and that the alternating group A_n is nonabelian and simple whenever $n \geq 5$. And they have seen the basic theorems of Galois theory, at least over $\mathbb{Q}$. They know that every polynomial $f(x)$ with coefficients in the field of rational numbers $\mathbb{Q}$ is associated with a group G of permutations of the roots of $f(x)$ in the field of complex numbers $\mathbb{C}$. This group G, called the Galois group of f, consists of those permutations which arise from field automorphisms of $\mathbb{C}$. The polynomial $f(x)$ is solvable by radicals (that is, there is a formula for finding the roots of $f(x)$ in terms of arithmetic operations and taking k^{th} roots) if and only if no quotient of G has a nonabelian simple subgroup. They also know that if $f(x)$ is irreducible, the degree of $f(x)$ divides the order of G. They can be given the following theorem as a computational exercise.

THEOREM 3 . *For p a prime, the symmetric group on p letters is generated by any p-cycle and a transposition (i.e., a 2-cycle).*

Hint: First show that, for any cycle $(i_1, i_2, \ldots, i_k)$ and any permutation σ,

$$\sigma(i_1, i_2, \ldots, i_k)\sigma^{-1} = \left(\sigma(i_1), \sigma(i_2), \ldots, \sigma(i_k)\right).$$

Then show that you can get any transposition if you start with a transposition and an n-cycle. You must use the fact that p is a prime here. If n is not prime, for any n-cycle σ there is a transposition τ such that σ and τ do not generate the symmetric group.

Putting this all together, Dick and Jane can then get the following theorems:

THEOREM 4 . *Let p be a prime. Then the Galois group of*

$$f_p(x) = x^2(x-2)(x-4)\cdots(x-2(p-2))-2$$

(or x^2+2 if $p=2$) is the symmetric group on p letters. If $p \geq 5$, then the roots of the polynomial $f_p(x)$ cannot be expressed in terms of arithmetic operations and taking k^{th} roots.

Proof. If $p=2, f_2(x)$ is irreducible of degree 2 and the Galois group is precisely the group of permutations of the 2 complex roots of this polynomial. If p is an odd prime, then complex conjugation is an automorphism of $\mathbb{C}$ which permutes precisely two roots of $f_p(x)$. Since the degree of $f_p(x)$ divides the order of the Galois group, G must contain an element of order p, which is a p-cycle. Here is the one place where we use the fact that p is prime. Hence, by the previous exercise, G is the entire symmetric group. If $p \geq 5$, then G contains the nonabelian simple group A_n, so $f_n(x)$ is not solvable by radicals, i.e., there is no formula for its roots. ■

We conclude with a corollary of this result using standard Galois theory concepts and results.

THEOREM 5. *Let G be any finite group. Then there exists a finite-dimensional extension field K of $\mathbb{Q}$ and a finite-dimensional normal extension E of K such that the Galois group of E over K is G.*

Proof. Let p be a prime bigger than the order of G. Then the polynomial $f_p(x)$ has Galois group the symmetric group on p letters, and G embeds in this Galois group. Let E be the splitting field of $f_p(x)$ over $\mathbb{Q}$, that is, the smallest subfield of $\mathbb{C}$ containing all of the roots of $f_p(x)$. Let K be the fixed field of the image of G in the Galois group of $f_p(x)$. Then the standard Galois correspondence shows that K and E have the desired properties. ■

Dick and Jane might be interested to know that there are intriguing related questions that are still unanswered. For instance, it is an open question whether every finite group G is the Galois group of some polynomial over $\mathbb{Q}$ itself (as opposed to some finite extension).

REFERENCES

1. J. A. Gallian, *Contemporary Abstract Algebra*, D. C. Heath and Co., Lexington, MA, 1994.
2. I. N. Herstein, *Topics in Algebra*, second edition, Xerox College Publishing, Lexington, MA.-Toronto, Ont., 1975.

An Optimization Problem Involving Binomial Coefficients

ERNEST G. MANES
University of Massachusetts
Amherst, MA 01003

Introduction My colleague Stan Kulikowski was interested in designing "scanners" for severely handicapped persons who could grip a lever and move it either up or down from its neutral position (see **[1]**). Suppose it is desired to access $N = 150$ items with this system. The system starts with the first item on a screen. To access the ith item, press the lever down $i - 1$ times to put the ith item on the screen and then press the lever up to select it. We say that the access time is i because a total of i lever actions was necessary for the access. With this one-dimensional architecture, the maximum access time is N and the average access time is $(1 + 2 + \cdots + N)/N = (N + 1)/2$.

Smaller access times can be achieved with higher-dimensional collection hierarchies. Kulikowski had observed that scanners could be made more efficient by increasing the dimension. I became involved when he approached me for a theoretical foundation. Only basic undergraduate mathematical ideas are needed.

Example. Consider dimension 2 with $N = 12$ and collections $c_1, \dots, c_5$. The following notation indicates how the 12 items might be stored:

$$c_1 : i_1 i_2 i_3 i_4$$
$$c_2 : i_5 i_6 i_7$$
$$c_3 : i_8 i_9 i_{10}$$
$$c_4 : i_{11}$$
$$c_5 : i_{12}$$

The access sequence for i_9 is *Down Down Up Down Up*, for an access time of 5. Here, the first Up selects collection c_3 and exposes the items $i_8 i_9 i_{10}$. There are 3 lever actions in the collection dimension and 2 actions in the item dimension, so the coordinates of the access are $(3, 2)$; the access time is the sum of the coordinates. In dimension 2 the maximum access time is 6 (for i_{10} with coordinates $(3, 3)$) and the average access time is $13/3$—both of which improve the 1-dimensional case.

Example. Following is a 3-dimensional system for 12 items, with 2 levels of collections:

$$\begin{aligned} c_1 &: e_1 : i_1 i_2 i_3 \\ &: e_2 : i_4 i_5 \\ &: e_3 : i_6 i_7 \\ c_2 &: e_4 : i_8 i_9 \\ &: e_5 : i_{10} \\ c_3 &: e_6 : i_{11} i_{12} \end{aligned}$$

Initially, collection c_1 is selected. Levering *Up* exposes the list e_1, e_2, e_3, with e_1 selected. Levering *Up* again exposes the list $i_1 i_2 i_3$, with i_1 selected. Levering *Up* a third time accesses the item i_1. Thus, to get to the closest item i_1 requires *Up Up Up* with coordinates $(1,1,1)$ and so with access time 3, the sum of the coordinates. (By the same reasoning, a d-dimensional system has all access times of least d.) By a similar process, i_9 is accessed with *Down Up Up Down Up* with coordinates $(2,1,2)$ and access time 5. The maximum access time is again 6—for the coordinates $(1,3,2)$ which accesses item i_7 and again for the coordinates $(3,1,2)$ which accesses item i_{12}. Now we have a larger average access time—$19/4$—than in the 2-dimensional case.

Navigational addressing We have just considered how to assign sequences of downs and ups as "addresses" to items in a database. By varying the dimension we found that we could get the maximum access time as low as 6, with average access time $13/3$. By thinking of *Down*, *Up* as binary digits (say 0 and 1) and counting to 12 in binary—$0001, 0010, \ldots, 1100$—all access times are 4. Thus both maximum and average access times are 4, a considerable improvement. But we reject this base-2 approach because it requires the user to know the address of an item before it can be accessed, a big problem if the database is large.

We focus instead on *navigational* systems in which the user knows an algorithm to find an item based on a meaningful standard description of it. A public library provides a familiar example. We may locate virtually every word in the library by imposing a three-dimensional cartesian system, with unit of length the millimeter. This is not a useful system since a patron searching for an article on raccoons will not know the coordinates. The following navigational algorithm is more practical:

(i) Scan each shelf until the desired encyclopedia is found.
(ii) Scan each volume until "R" is found.
(iii) Scan each page head until "rac" is found.
(iv) Scan down the columns to find the desired article.

This example is a 4-dimensional hierarchy in the sense of the introduction. Such navigational systems might be called *scan-and-select* systems because at any stage the user either scans further (lever down) or selects the current collection (lever up). Computer users will recognize navigating to an item through a sequence of menus as an example of scan-and-select. We shall restrict our attention to scan-and-select systems in this note. Our main objective is to choose a dimension for a given database so as to optimize the maximum access time; we shall have something to say about the average access time, too.

A mathematical framework To formalize scan-and-select systems, let's fix some notation and terminology. A *d–item* is a d–tuple $(n_1, \ldots, n_d)$ of positive integers. The *access time* of a d–item is the sum $n_1 + \cdots + n_d$. A *d–system* is a non-empty set of d–items satisfying the constraint that if T is the maximum access time of the system, then all d–items with access time less than T are already in the system.

To justify the constraint, consider a d–system with maximum access time T. If some slot with smaller access time exists, it could have been filled with an item of greater access time, thereby possibly lowering the maximum access time and definitely lowering the average access time. The constraint is satisfied by the preceding examples.

We will use the following notations for a d–system:

d: the dimension;
N: the number of d–items;
$T_{N,d}$: the maximum access time of a d–item;
$a_{N,d}$: the average access time of a d–item.

We seek practical algorithms to compute $T_{N,d}$ and $a_{N,d}$. We shall determine, given N, how to choose at least one d to minimize $T_{N,d}$. We shall also show that for all N there exists d such that $T_{N,d}$ is the smallest maximum access time and $T_{N,d} - a_{N,d} < 2$.

Our first proposition is well known (see, e.g., [**2**, p. 216]).

PROPOSITION 1. *The number of d–items with access time t is $\binom{t-1}{d-1}$. The number of d–items with access time $\leq t$ is $\binom{t}{d}$.*

COROLLARY (THE SANDWICH INEQUALITY). *$T_{N,d}$ is the unique positive integer t such that $\binom{t-1}{d} < N \leq \binom{t}{d}$.*

The sandwich inequality implies that scan-and-select is strictly less efficient than base-2 addressing if $N = 2^n$. Base-2 addressing then requires n binary digits, for a maximum access time of n. Clearly, no address system can access 2^n items with fewer than n bits, so a scan-and-select system with dimension d must have maximum access time $T \geq n$. If $T = n$, then, by the sandwich inequality, $2^n \leq \binom{n}{d}$, which is impossible. We leave it as an exercise to show that base-2 is, in fact, more efficient for any N.

Using elementary calculus we may obtain a quick estimate of $T_{N,d}$ for any N and d. Let $\lceil x \rceil$ denote the smallest integer $\geq x$. Then $T_{N,d} = \lceil r \rceil$, where r is the unique root on $[d, \infty)$ of the polynomial

$$p_{N,d}(x) = x(x-1)\cdots(x-d+1) - Nd!.$$

To see why, note that $p_{N,d}$ has positive derivative on $[d, \infty)$; multiplying the sandwich inequality by $d!$ and subtracting $Nd!$ gives

$$p(T_{N,d} - 1) < 0 \leq p_{N,d}(T_{N,d})$$

so that $T_{N,d} - 1 < r \leq T_{N,d}$. Root-finding algorithms can be used to estimate r, and hence also $T_{N,d}$.

Clearly, $T_{N,1} = N$; by the quadratic formula and the preceding observation,

$$T_{N,2} = \left\lceil \frac{1 + \sqrt{1 + 8N}}{2} \right\rceil.$$

For higher d, a convenient lower-bound for $T_{N,d}$ can be found as follows. By the geometric–arithmetic mean inequality,

$$\sqrt[d]{x(x-1)\cdots(x-d)} \leq x - \frac{d-1}{2}.$$

Thus

$$p_{N,d}(x) \leq q(x) = \left(x - \frac{d-1}{2}\right)^d - Nd!$$

so the unique root of q occurs before that of p. This gives the approximation

$$T_{N,d} \geq \left\lceil \frac{d-1}{2} + \sqrt[d]{Nd!} \right\rceil. \tag{1}$$

To obtain a simple upper bound for $T_{N,d}$, note that the sandwich inequality gives

$$Nd! \leq t(t-1)\cdots(t-d+1) \leq t^d,$$

so $t = T_{N,d} \leq \lfloor \sqrt[d]{Nd!} \rfloor$. Starting with these bounds, the bisection method, using the sandwich inequality as a test, finds the true value of $T_{N,d}$ in at most $\lceil \log_2 \frac{d-1}{2} \rceil$ steps.

The next proposition links the average and maximum access times.

PROPOSITION 2. $a_{N,d} = T_{N,d} - \frac{1}{N}\binom{T_{N,d}}{d+1}$.

Proof. Write T for $T_{N,d}$. The well-known identity (see, e.g., [**2**, p. 217]) gives

$$\binom{t+1}{d+1} = \binom{d}{d} + \binom{d+1}{d} + \cdots + \binom{t}{d}.$$

This and the fact that $u\binom{u-1}{d-1} = d\binom{u}{d}$ imply that the sum of the access times for all items with access time less than T is

$$d\left(\binom{d}{d} + \binom{d+1}{d} + \cdots + \binom{T-1}{d}\right) = d\binom{T}{d+1}.$$

As all $\binom{T-1}{d}$ items with access time less than T are in the d–system, $N - \binom{T-1}{d}$ items have access time T. Hence the average access time of all items in the system is $\frac{d}{N}\binom{T}{d+1} + T - \frac{T}{N}\binom{T-1}{d}$, which simplifies as desired. ■

Choosing a good dimension How does one choose a scan-and-select dimension for a given data base so as to minimize the maximum access time?

THEOREM. *Let* $\alpha_d = \binom{2d-1}{d}$, $\beta_d = \binom{2d}{d-1}$, *and* $\gamma_d = \binom{2d}{d}$. *Then*

$$\alpha_d < \beta_d < \gamma_d < \alpha_{d+1}$$

for all $d \geq 2$. *For* $N \leq 3$, $d = 1$ *gives the smallest possible maximum access time,* $T = N$. *Otherwise* $N > \alpha_2 = 3$, *and exactly one of the following cases obtains*:

Case A: *There exists unique d with* $\alpha_d < N \leq \beta_d$. *The smallest maximum access time for N is* $T = 2d$, *and is attained in dimension* $d - 1$.

Case B: *There exists unique d with* $\beta_d < N \leq \gamma_d$. *The smallest maximum access time for N is* $T = 2d$, *and is attained in dimension d.*

Case C: *There exists unique d with* $\gamma_d < N \leq \alpha_{d+1}$. *The smallest maximum access time for N is* $T = 2d + 1$, *and is attained in dimension d.*

Proof. That $\alpha_d < \beta_d < \gamma_d < \alpha_{d+1}$ is easily checked by direct calculation. For any d–system with $d > 1$ the first three entries have respective access times $d, d+1$, and $d+1$ so it is clear that $d = 1$ minimizes maximum access time for 1, 2, or 3 entries.

The proofs of cases A, B, and C are similar; we do only case B. Let $\beta_d < N \leq \gamma_d$ with $N > 3$. Then

$$\binom{2d-1}{d} = \alpha_d < \beta_d < N \leq \gamma_d = \binom{2d}{d}.$$

By the sandwich inequality, $T_{N,d} = 2d$. Now suppose that $\binom{t-1}{e} < N \leq \binom{t}{e}$, so that N entries in dimension e have maximum access time t. We must show that $t \geq 2d$. If not, then $t < 2d$. If $f = e_0$ is chosen to maximize $\binom{t}{f}$, then $e_0 \leq d$ because the largest entries in each row of Pascal's triangle occur in the middle. Similarly, $\binom{2d-1}{d}$ is the largest number of the form $\binom{2d-1}{f}$. Thus, we have

$$\binom{t}{e} \leq \binom{t}{e_0} \leq \binom{2d-1}{e_0} \leq \binom{2d-1}{d} = \alpha_d < N,$$

which contradicts $N \leq \binom{t}{e}$. ■

The theorem has practical value. With a table of the first hundred values of α_d, a user could quickly locate the d with $\alpha_d < N \leq \alpha_{d+1}$. This treats N up to $\alpha_{100} \approx 4.52 \times 10^{58}$.

It should be stressed that the dimension optimizing the maximum access time is not unique. If, say, $N = 84$, then $\gamma_4 < N \leq \alpha_5 = 126$. By case C, the maximum access time of 9 is realized in dimension 4. But the sandwich inequality shows that dimensions 3, 5, and 6 also give a maximum access time of 9.

Our final proposition compares the average- to the maximum access time.

PROPOSITION 3. *For any N, there exists d such that $T_{N,d}$ is the smallest maximum access time, and such that*

$$\frac{d}{d+1} \leq T_{N,d} - a_{N,d} < 2.$$

Proof. For arbitrary N and d, set $T = T_{N,d}$ and $a = a_{N,d}$. By Proposition 2,

$$T - a = \frac{1}{N}\binom{T}{d+1}.$$

Calculation gives

$$\frac{\binom{T}{d+1}}{\binom{T-1}{d}} = \frac{T}{d+1} \quad \text{and} \quad \frac{\binom{T}{d+1}}{\binom{T}{d}} = \frac{T-d}{d+1};$$

combining these with the sandwich inequality gives

$$\frac{T-d}{d+1} \leq T - a < \frac{T}{d+1},$$

which is an interesting relation in its own right. Now choose d by the previous theorem so that T is the smallest possible maximum access time according to one of cases A, B, and C. Since T is one of $2d + 2$, $2d$, and $2d + 1$, we're done. ■

REFERENCES

1. S. Kulikowski, II, Efficiency limits on scanning structures: expert systems in nonvocal communication prosthesis, *Computer Technology for the Handicapped: APPLICATIONS '85, Selected Proceedings of Closing the Gap's* 1985 *National Conference*, Minneapolis, Oct. 30–Nov. 2 (1985), 89–97.
2. A. Tucker, *Applied Combinatorics*, Third Ed., John Wiley, New York, NY, 1995.

Extended Euclid's Algorithm via Backward Recurrence Relations

S. P. GLASBY
The University of the South Pacific
Suva
Fiji Islands

Introduction Given elements a and b of a Euclidean ring R, Euclid's algorithm is most useful for computing the greatest common divisor $\gcd(a, b)$, or when $\gcd(a, b)$ is invertible, computing the inverse of $a + bR$ in the quotient ring R/bR. The second problem is usually solved by computing $x, y \in R$ satisfying $ax + by = \gcd(a, b)$ via the familiar backward substitution method. It seems less well known that solving $ax + by = \gcd(a, b)$ can be performed more efficiently using a "backward" recurrence relation. Many books, such as [**1**, **2**, **3**], use a "forward" recurrence relation method. I shall argue that the backward recurrence relation method is both pedagogically more natural for students, and more efficient for hand computations.

The ring $F[X]$ of polynomials over a field F and the ring $\mathbb{Z}$ of integers are examples of Euclidean rings. There is a "degree" function $\delta : R \to \mathbb{N} \cup \{-\infty\}$. Given $a, b \in R$ where $b \neq 0$ there exist a "quotient" q, and a "remainder" r in R such that $a = qb + r$ where $\delta(r) < \delta(b)$. The reader unfamiliar with Euclidean rings in general need not worry: in our examples, $R = \mathbb{Z}$ and q and r are the usual quotients and remainders, i.e., $q = \lfloor a/b \rfloor$ (where $\lfloor\ \rfloor$ denotes the greatest integer function), $r = a - qb$, and $\delta(r) = |r|$.

Given $a_0, a_1 \in R$ where $a_1 \neq 0$, there exist quotients $q_1, \dots, q_r$ and remainders $a_2, \dots, a_{r+1}$ in R such that $a_2, \dots, a_r$ are nonzero and

$$a_0 = q_1 a_1 + a_2, \quad a_1 = q_2 a_2 + a_3, \quad \dots \quad a_{r-1} = q_r a_r + a_{r+1}, \tag{1}$$

where $\delta(a_{r+1}) < \cdots < \delta(a_2) < \delta(a_1)$. Since $\mathbb{N} \cup \{-\infty\}$ has no infinite descending sequences, this process cannot continue indefinitely, so we shall assume that $a_r \neq a_{r+1} = 0$. It follows from $\gcd(a_{i-1}, a_i) = \gcd(a_i, a_{i+1})$ that $\gcd(a_0, a_1) = a_r$. Furthermore, by writing equations (1) in the form $a_{i+1} = a_{i-1} - q_i a_i$ and using backward substitution, one can show that there exist $x, y \in R$ such that $a_0 x + a_1 y = a_r$. That is, starting with $a_r = a_{r-2} - q_{r-1} a_{r-1}$, one substitutes $a_{r-1} = a_{r-3} - q_{r-2} a_{r-2}$ and then, after collecting terms, $a_{r-2} = a_{r-4} - q_{r-3} a_{r-3}$ is substituted, etc.

When writing the equations (1) one sees the same terms written repeatedly: $a_2, \dots, a_{r-1}$ each appear three times, and a_1, a_r appear twice. This suggests abbreviating these equations by

$$\begin{array}{cccccc} & q_1 & q_2 & \cdots & q_r & \\ \hline a_0 & a_1 & a_2 & \cdots & a_r & 0 \end{array} \tag{2}$$

A backward recurrence relation One can start with a_0, a_1 and generate higher a_i via the *forward recurrence* $a_{i+1} = -q_i a_i + a_{i-1}$, or less conventionally, start with $a_{r+1} = 0$ and a_r and generate lower a_i via a *backward recurrence*: $a_{i-1} = q_i a_i + a_{i+1}$. The backward substitution method corresponds to solving the equations $a_{i-1} x_i + a_i y_i = a_r$ for x_i and y_i until one finds x_1 and y_1. It turns out to be more convenient to consider the following related equations: $a_{i-1} x_{i-1} + a_i(-1)^{r+i} y_{i-1} = a_r$. Our initial conditions are then $x_r = 1$, $y_r = 0$ and $x_{r-1} = 0$, $y_{r-1} = 1$. A recurrence relation can be determined as follows:

$$\begin{aligned} a_r &= a_i x_i + a_{i+1}(-1)^{r+i+1} y_i \\ &= a_i x_i + (a_{i-1} - q_i a_i)(-1)^{r+i+1} y_i \\ &= a_{i-1}(-1)^{r+i+1} y_i + a_i(-1)^{r+i}\left(q_i y_i + (-1)^{r+i} x_i\right). \end{aligned}$$

Take $x_{i-1} = (-1)^{r+i+1} y_i$ and $y_{i-1} = q_i y_i + (-1)^{r+i} x_i$. Eliminating x_i gives the following recurrence relation:

$$y_r = 0, \quad y_{r-1} = 1 \quad \text{and} \quad y_{i-1} = q_i y_i + y_{i+1}.$$

Hence we may add an extra row to table (2) to compute the y's:

	q_1	$\cdots$		q_i		$\cdots$	q_r	
a_0	a_1	$\cdots$	$q_i a_i + a_{i+1}$	a_i	a_{i+1}	$\cdots$	a_r	0 .
y_0	y_1	$\cdots$	$q_i y_i + y_{i+1}$	y_i	y_{i+1}	$\cdots$	y_r	

Recall that

$$\begin{aligned} a_r &= a_{i-1} x_{i-1} + a_i(-1)^{r+i} y_{i-1} \\ &= a_{i-1}(-1)^{r+i+1} y_i + a_i(-1)^{r+i} y_{i-1}. \end{aligned}$$

Putting $i = 1$ gives $a_0 y_1 - a_1 y_0 = (-1)^r a_r$. As an illustration, we compute gcd(74, 54) and solve $74x + 54y = \gcd(74, 54)$:

	1	2	1	2	3	
74	54	20	14	6	2	0
11	8	3	2	1	0	

Therefore $74 \times 8 - 54 \times 11 = (-1)^5 2$, so $x = -8$ and $y = 11$ is a solution.

The above calculation has advantages over the familiar method:

$$\begin{aligned} 2 &= 1 \times 14 - 2 \times 6 \\ &= 1 \times 14 - 2 \times (20 - 1 \times 14) = -2 \times 20 + 3 \times 14 \\ &= -2 \times 20 + 3 \times (54 - 2 \times 20) = 3 \times 54 - 8 \times 20 \\ &= 3 \times 54 - 8 \times (74 - 1 \times 54) = -8 \times 74 + 11 \times 54. \end{aligned}$$

Forward recurrence relations An alternative approach is to set $a_0 X_i + a_1 Y_i = a_i$ and seek the values of X_r and Y_r. Our initial conditions are $X_0 = 1, Y_0 = 0$ and $X_1 = 0, Y_1 = 1$. Furthermore,

$$\begin{aligned} a_{i+1} &= a_{i-1} - q_i a_i \\ &= (a_0 X_{i-1} + a_1 Y_{i-1}) - q_i(a_0 X_i + a_1 Y_i) \\ &= a_0(X_{i-1} - q_i X_i) + a_1(Y_{i-1} - q_i Y_i) \end{aligned}$$

so we may take $X_{i+1} = X_{i-1} - q_i X_i$ and $Y_{i+1} = Y_{i-1} - q_i Y_i$. Unlike the backward method, these recurrence relations are not coupled: one may solve for the X's independently of the Y's, and vice versa. Hence we may add two extra rows to table

(2) to compute the X's and the Y's:

	q_1	$\cdots$		q_i		$\cdots$	q_r	
a_0	a_1	$\cdots$	a_{i-1}	a_i	$a_{i-1}-q_i a_i$	$\cdots$	a_r	0
X_0	X_1	$\cdots$	X_{i-1}	X_i	$X_{i-1}-q_i X_i$	$\cdots$	X_r	
Y_0	Y_1	$\cdots$	Y_{i-1}	Y_i	$Y_{i-1}-q_i Y_i$	$\cdots$	Y_r	

For example, if $a_0 = 74$ and $a_1 = 54$, this method gives

	1	2	1	2	3	
74	54	20	14	6	2	0
1	0	1	-2	3	-8	
0	1	-1	3	-4	11	

Therefore $74 \times (-8) + 54 \times 11 = 2$, yielding the same answer as before.

Comparing the methods The backward recurrence method was motivated by backward substitution which is taught to most students, and so is pedagogically preferable to the forward recurrence method. From the point of view of hand calculations, the backward recurrence method requires half the effort, and students are less likely to make a sign error. The X_i, Y_i, and y_i can be expressed in terms of polynomials in $q_1, \ldots, q_{r-1}$, called *continuants* (see **[2]**). Computing the y's instead of the X's and the Y's has half the computational complexity, as the polynomials arising are very similar. It follows from a symmetry property of continuants that $y_0 = (-1)^{r-1} Y_r$ and $y_1 = (-1)^r X_r$.

The forward method is good for computers when long computations are involved as the q's need not be stored—once q_i has been used to compute a_{i+1}, X_{i+1} and Y_{i+1}, it can be discarded. This makes writing a computer program for the forward method slightly easier than the backward method. This is no advantage for hand calculation, as the student will invariably have recorded each q_i on paper.

The point made in the first paragraph needs qualification. If one uses the forward recurrence method to compute $d = \gcd(a, b)$ and X, then Y can be computed, using three divisions, from the equation $Y = (1 - a'X)/b'$ where $a' = a/d, b' = b/d$. When computing the inverse of $a + bR$ in the quotient ring R/bR, the value of Y is irrelevant, and so one need only compute d and X.

When $R = \mathbb{Z}$ and a_0, a_1 are positive, it is annoying that the signs of the X's and the Y's alternate. This can be avoided by redefining X_i and Y_i as follows: $a_0 X_i - a_1 Y_i = (-1)^i a_i$. Then the forward recurrence relations become

$$X_0 = 1, \quad X_1 = 0 \quad \text{and} \quad X_{i+1} = X_{i-1} + q_i X_i,$$
$$Y_0 = 0, \quad Y_1 = 1 \quad \text{and} \quad Y_{i+1} = Y_{i-1} + q_i Y_i.$$

This, however, is undesirable from the view point of teaching as then different rows of our table are computed via different rules (namely $a_{i+1} = a_{i-1} - q_i a_i$, $X_{i+1} = X_{i-1} + q_i X_i$, and $Y_{i+1} = Y_{i-1} + q_i Y_i$).

REFERENCES

1. Henri Cohen, *A Course in Computational Algebraic Number Theory*, Graduate Texts in Mathematics 138, Springer-Verlag, New York, NY, 1995.
2. Donald E. Knuth, *The Art of Computer Programming, Vol. 2: Semi-Numerical Algorithms*, Addison-Wesley, Reading, MA, 1995.
3. M. Pohst and H. Zassenhaus, Algorithmic algebraic number theory, in *Encyclopedia of Mathematics and its Applications*, Addison-Wesley, Reading, MA, 1989.

On the Diagonalization of Quadratic Forms

T. Y. LAM
University of California
Berkeley, CA 94720

Introduction An undergraduate course in linear algebra sometimes includes a treatment of the elementary theory of quadratic forms. This is not surprising since a quadratic form (over a field F of characteristic not 2) is essentially the same as a symmetric bilinear form, which is in turn the same as a symmetric matrix. The main theorem for quadratic forms proved in a typical linear algebra course is that any quadratic form $q(x_1, \ldots, x_n)$ can be "diagonalized," i.e., after a linear change of variables $\{x_1, \ldots, x_n\} \to \{y_1, \ldots, y_n\}$, q can be written as $\Sigma_{i=1}^n a_i y_i^2$. Here, the a_i's are elements of F, some of which may be equal to zero. The *rank* of q is defined to be the number of nonzero a_i's, and (in case $F = \mathbb{R}$) the *signature* of q is defined to be $r - s$, where r and s are respectively the number of positive and negative a_i's. Both the rank and the signature depend only on the isometry class of the quadratic form q (and does not depend on the particular diagonalization taken); see, e.g., [**1**, §5.3], [**3**, Ch. 9].

While most textbooks offer exercises for the diagonalization of quadratic forms in a small number of variables (say $n \leq 5$), there seem to be few good examples for the diagonalization of quadratic forms in n variables. In teaching a course in quadratic form theory, I recently came across the following four explicit n-ary forms:

$$q_1(\mathbf{x}) = \sum_{i,j=1}^{n} \min\{i,j\} x_i x_j, \qquad q_2(\mathbf{x}) = \sum_{i,j=1}^{n} \max\{i,j\} x_i x_j,$$

$$q_3(\mathbf{x}) = \sum_{i,j=1}^{n} (i+j) x_i x_j, \qquad q_4(\mathbf{x}) = \sum_{i,j=1}^{n} |i-j| x_i x_j.$$

For $n = 4$ (for example), these quadratic forms have the following symmetric matrices:

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 2 & 3 & 4 \\ 3 & 3 & 3 & 4 \\ 4 & 4 & 4 & 4 \end{pmatrix}, \quad \begin{pmatrix} 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \\ 5 & 6 & 7 & 8 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{pmatrix}.$$

The diagonalization for the quadratic forms q_i above turns out to be an interesting exercise, the solution of which also leads to some nontrivial conclusions about the forms themselves. I shall offer my solution(s) below as an illustration for the general process of diagonalizing quadratic forms, for those colleagues who may want to present some challenging examples in a linear algebra class.

Throughout this note, we work over the field of rational numbers, although our main computations are valid over any field of characteristic not 2.

The forms q_1, q_2 We start with the form q_1. To diagonalize it, note that

$$\begin{aligned}
q_1(\mathbf{x}) &= x_1^2 + 2x_1(x_2 + \cdots + x_n) + \sum_{i,j=2}^{n} \min\{i,j\} x_i x_j \\
&= (x_1 + x_2 + \cdots + x_n)^2 - (x_2 + \cdots + x_n)^2 + 2x_2^2 + 4x_2(x_3 + \cdots + x_n) \\
&\quad + \sum_{i,j=3}^{n} \min\{i,j\} x_i x_j \\
&= (x_1 + \cdots + x_n)^2 + (x_2 + \cdots + x_n)^2 - 2(x_3 + \cdots + x_n)^2 \\
&\quad + \sum_{i,j=3}^{n} \min\{i,j\} x_i x_j .
\end{aligned}$$

Repeating this process to the end, we get

$$\begin{aligned}
q_1(\mathbf{x}) &= (x_1 + x_2 + \cdots + x_n)^2 + (x_2 + \cdots + x_n)^2 + \cdots + (x_{n-1} + x_n)^2 \\
&\quad - (n-1)x_n^2 + nx_n^2 \\
&= (x_1 + x_2 + \cdots + x_n)^2 + (x_2 + \cdots + x_n)^2 + \cdots + (x_{n-1} + x_n)^2 + x_n^2 .
\end{aligned}$$

Of course, once we derived this equation, it can also be checked directly by comparing the coefficients of $x_i x_j$ on the two sides. It follows that q_1 has rank n and signature n. The above sum-of-squares expression for q_1 has appeared in **[2]**. The fact that q_1 is positive definite was used in **[2]** in an ingenious way to show that a certain representation ring of a cyclic p-group has no nonzero nilpotent elements.

For the form q_2, we start the diagonalization process from the last variable:

$$\begin{aligned}
q_2(\mathbf{x}) &= nx_n^2 + 2nx_n(x_1 + \cdots + x_{n-1}) + \sum_{i,j=1}^{n-1} \max\{i,j\} x_i x_j \\
&= n(x_1 + \cdots + x_n)^2 - n(x_1 + \cdots + x_{n-1})^2 \\
&\quad + (n-1)(x_1 + \cdots + x_{n-1})^2 - (n-1)(x_1 + \cdots + x_{n-2})^2 \\
&\quad + \sum_{i,j=1}^{n-2} \max\{i,j\} x_i x_j \\
&= n(x_1 + \cdots + x_n)^2 - (x_1 + \cdots + x_{n-1})^2 - (n-1)(x_1 + \cdots + x_{n-2})^2 \\
&\quad + \sum_{i,j=1}^{n-2} \max\{i,j\} x_i x_j .
\end{aligned}$$

Repeating this process to the end, we arrive at the diagonalization

$$\begin{aligned}
q_2(\mathbf{x}) &= n(x_1 + \cdots + x_n)^2 - (x_1 + \cdots + x_{n-1})^2 - \cdots - (x_1 + x_2)^2 - 2x_1^2 + x_1^2 \\
&= n(x_1 + \cdots + x_n)^2 - (x_1 + \cdots + x_{n-1})^2 - \cdots - (x_1 + x_2)^2 - x_1^2 ,
\end{aligned}$$

which, again, can be checked directly by comparing coefficients. In particular, q_2 has rank n and signature $1 - (n-1) = 2 - n$. Here, we get as a bonus a criterion for $q_2(x_1, \ldots, x_n)$ to be nonnegative for real numbers $x_1, \ldots, x_n$, namely:

$$n(x_1 + \cdots + x_n)^2 \geq (x_1 + \cdots + x_{n-1})^2 + (x_1 + \cdots + x_{n-2})^2 + \cdots + x_1^2 .$$

An alternative way to obtain a diagonalization for q_2 is to use the following interesting "reciprocal" relation between q_1 and q_2:

$$q_2(x_1,\dots,x_n)+q_1(x_n,\dots,x_1)=(n+1)(x_1+\cdots+x_n)^2. \qquad (*)$$

This can be proved either by using the explicit forms of the symmetric matrices for the two forms on the left-hand side, or by relabelling the variables backwards in $q_2(\mathbf{x})$ and using the max-min relation:

$$\max\{n-i+1,n-j+1\}=n+1-\min\{i,j\}.$$

Given the equation $(*)$, we get

$$\begin{aligned}q_2(x_1,\dots,x_n)&=(n+1)(x_1+\cdots+x_n)^2-q_1(x_n,\dots,x_1)\\&=(n+1)(x_1+\cdots+x_n)^2-\Big((x_n+\cdots+x_1)^2+(x_{n-1}+\cdots+x_1)^2\\&\qquad+\cdots+(x_2+x_1)^2+x_1^2\Big)\\&=n(x_1+\cdots+x_n)^2-(x_1+\cdots+x_{n-1})^2-\cdots-(x_1+x_2)^2-x_1^2,\end{aligned}$$

as before.

The explicit diagonalizations for q_1 and q_2 above also lead to various formulas relating the n-ary forms q_i to their $(n-1)$-ary versions. The exploration of these further relations will be left to the reader.

The forms q_3, q_4 Coming now to the form q_3, we first note, in the spirit of $(*)$, that $q_3(\mathbf{x})=q_1(\mathbf{x})+q_2(\mathbf{x})$. However, the two diagonalizations obtained for q_1 and q_2 are *not* mutually compatible, so just adding them does not lead to a diagonalization of q_3. Thus, we must deal with the diagonalization problem from scratch.

Fortunately, there is a good way out. Upon noting that $\Sigma_{i,j}ix_ix_j=\Sigma_{i,j}jx_ix_j$, we can write $q_3=2f$ for the quadratic form $f(\mathbf{x})=\Sigma_{i,j}ix_ix_j$. Now f factors into $(\Sigma_i ix_i)(\Sigma_j x_j)$. Therefore, using the identity, $uv=[(u+v)^2-(u-v)^2]/4$, we get

$$\begin{aligned}q_3(\mathbf{x})&=\frac{2}{4}\left[\left(\sum_i ix_i+\sum_i x_i\right)^2-\left(\sum_i ix_i-\sum_i x_i\right)^2\right]\\&=\frac{1}{2}\left[(2x_1+3x_2+\cdots+(n+1)x_n)^2-(x_2+2x_3+\cdots+(n-1)x_n)^2\right].\end{aligned}$$

This yields the diagonalization $q_3=\frac{1}{2}(z_1^2-z_2^2)$ with $z_1=2x_1+3x_2+\cdots+(n+1)x_n$ and $z_2=x_2+2x_3+\cdots+(n-1)x_n$. In particular, q_3 has rank 2 and signature 0, and we get the interesting bonus conclusion that, for real variables $x_1,\dots,x_n$, we have $\Sigma_{i,j}(i+j)x_ix_j\geq 0$ if and only if

$$|2x_1+3x_2+\cdots+(n+1)x_n|\geq|x_2+2x_3+\cdots+(n-1)x_n|.$$

We observe in passing that the above method can actually be used to handle the more general form $q:=\Sigma_{i,j}(a_i+b_j)x_ix_j$, for arbitrary elements a_i, b_j in the field. One sets $c_i=(a_i+b_i)/2$ to rewrite q as $2g$ for $g=\Sigma_{i,j}c_ix_ix_j$, after which we can proceed as before.

Finally, we come to the form q_4. The diagonalization of q_4 would have been pretty difficult without the work above. Given what we have already done, however, a solution can be worked out. We start out again by noting that $q_4(\mathbf{x})=q_2(\mathbf{x})-q_1(\mathbf{x})$. This in itself does not yield a diagonalization for q_4, but we can further replace $q_2(\mathbf{x})$

by $q_3(\mathbf{x}) - q_1(\mathbf{x})$ to write $q_4(\mathbf{x})$ as $q_3(\mathbf{x}) - 2q_1(\mathbf{x})$. Now introduce the new variables $y_i = x_i + x_{i+1} + \cdots + x_n$ $(1 \le i \le n)$. Using our earlier results on q_1 and q_3, we can write

$$\begin{aligned} q_4(\mathbf{x}) &= 2(x_1 + \cdots + x_n)(x_1 + 2x_2 + \cdots + nx_n) - 2(y_1^2 + \cdots + y_n^2) \\ &= 2y_1(y_1 + y_2 + \cdots + y_n) - 2(y_1^2 + \cdots + y_n^2) \\ &= 2y_1(y_2 + \cdots + y_n) - 2(y_2^2 + \cdots + y_n^2) \\ &= -2\sum_{i=2}^{n}\left[\left(\frac{y_1}{2}\right)^2 - y_1 y_i + y_i^2\right] + 2(n-1)\left(\frac{y_1}{2}\right)^2 \\ &= \frac{n-1}{2} y_1^2 - \frac{1}{2}\sum_{i=2}^{n}(y_1 - 2y_i)^2 \\ &= (n-1)(x_1 + \cdots + x_n)^2/2 - (x_1 - x_2 - \cdots - x_n)^2/2 \\ &\quad -(x_1 + x_2 - x_3 - \cdots - x_n)^2/2 - \cdots - (x_1 + x_2 + \cdots + x_{n-1} - x_n)^2/2. \end{aligned}$$

This gives the desired diagonalization of $q_4(\mathbf{x})$ (over the rationals). In particular, we conclude that, like q_2, q_4 has rank n and signature $2-n$. Note that the above diagonalization of q_4 also leads to a criterion for $q_4(\mathbf{x})$ to be nonnegative for a given real vector $\mathbf{x}$.

As further exercises, the reader may try to diagonalize the following quadratic forms:

$$x_1x_2 + x_2x_3 + \cdots + x_{n-1}x_n, \qquad x_1x_2 + x_2x_3 + \cdots + x_{n-1}x_n + x_nx_1, \qquad \sum_{i<j} x_ix_j.$$

Epilogue After I mentioned the above diagonalization of the form $q_4(\mathbf{x})$ in a lecture at the University of Illinois, Professor Kenneth Stolarsky pointed out to me a number of references concerning $q_4(\mathbf{x})$. To my great surprise (as well as delight), it turned out that the form q_4 had been studied by my late Berkeley colleague Raphael M. Robinson sixty-five years ago! Robinson **[4]** posed the computation of the determinant

$$D_n := \det(q_4(\mathbf{x})) = \det(|i-j|)_{1 \le i, j \le n}$$

as Problem 3667 in the *American Mathematical Monthly* in 1934, and a little later, posed as Problem 3705 the computation of the signature of $q_4(\mathbf{x})$ over the real numbers. The latter problem was solved by G. Szegö in **[6]**. Szegö obtained the recurrence relation

$$D_n = -4(D_{n-1} + D_{n-2}),$$

whereby he proved by induction on n that $D_n = (-1)^{n-1}2^{n-2}(n-1)$. Using a formula of Frobenius, Szegö then showed that $q_4(\mathbf{x})$ has real signature $2-n$. Szegö also considered the Hermitian form $\sum_{i,j}|i-j|x_i\bar{x}_j$, and related it to the Fourier series expansions of periodic functions. The two solutions Szegö presented to Robinson's problems occupied six pages of the *Monthly*, and read like a paper in itself. This was followed by an even longer "Editorial Note," which occupied another eight pages. The total number of pages, 14, must have been a record for the amount of space ever devoted by the *Monthly* to the solution of a single problem. However, the problem of an explicit diagonalization of $q_4(\mathbf{x})$ *over the rationals* was not considered by Robinson or Szegö.

A few years later, in 1937, I. J. Schoenberg [**5**, §6] considered the quadratic form

$$\sum_{1 \le i,j \le n} |P_i - P_j|^{\alpha} x_i x_j \quad (0 < \alpha < 2),$$

where $P_1, \dots, P_n$ are distinct points in an euclidean space, and $|P_i - P_j|$ denote their euclidean distances. Schoenberg showed that this form is nonsingular with real signature $2 - n$, and that it is negative definite on the hyperplane $\sum_i x_i = 0$. In the case when $|P_i - P_j| = |i - j|$ and $\alpha = 1$, this, of course, also follows from our diagonalization of the form $q_4(\mathbf{x})$. I am very much indebted to Professor K. Stolarsky for pointing out the above pertinent references.

REFERENCES

1. B. Jacob, *Linear Algebra*, W. H. Freeman and Co., New York, NY, 1990.
2. T. A. Hannula, T. G. Rally, and I. Reiner, Modular representation algebras, *Bull. Amer. Math. Soc.* 73 (1967), 100–101.
3. K. Hoffman and R. Kunze, *Linear Algebra*, Prentice Hall, Englewood Cliffs, NJ, 1971.
4. R. M. Robinson, Problems 3667, 3705, *Amer. Math. Monthly* 41 (1934), p. 193 and p. 581.
5. I. J. Schoenberg, On certain metric spaces arising from euclidean spaces by a change of metric and their imbedding in Hilbert space, *Annals of Math.* 38 (1937), 787–793.
6. G. Szegö, Solution to Problem 3705, *Amer. Math. Monthly* 43 (1936), 246–259.

Math Bite: Irrationality of $\sqrt{m}$

A recent *Gazette* article [**3**] reminded me of Theodor Estermann's proof of the irrationality of $\sqrt{2}$. As his proof is not well known, I take the liberty of publicizing it with a slight "generalization." The really interesting thing about this proof is that it doesn't use divisibility, just mathematical induction in its "$\mathbb{Z}$ is well-ordered" form.

THEOREM. *Suppose m is not a perfect square. Then $\sqrt{m}$ is irrational.*

Proof. Let n be the integer with $n < \sqrt{m} < n + 1$. It suffices to prove that $\alpha = \sqrt{m} - n$ is irrational. Suppose not. As $0 < \alpha < 1$, we have $\alpha = p/q$ where $0 < p < q$. We may assume that q is *as small as possible* (Estermann's key idea). Then we have

$$\frac{q}{p} = \frac{1}{\sqrt{m} - n} = \frac{\sqrt{m} + n}{m - n^2} = \frac{\alpha + 2n}{m - n^2}.$$

We solve for α:

$$\alpha = \frac{(m - n^2)q}{p} - 2n = \frac{(m - n^2)q - 2np}{p} = \frac{r}{p}.$$

Thus α is a fraction with even smaller denominator, a contradiction.

REFERENCES

1. T. Estermann, The irrationality of $\sqrt{2}$, *Math. Gazette* 59 (1975), 110.
2. K. Roth and R. C. Vaughn, Obituary of Theodor Estermann, *Bull. London Math. Soc.* 26 (1994), 593–606.
3. M.-K. Siu, Estermann and Pythagoras, *Math. Gazette* 82 (1998), 92–93.

—HARLEY FLANDERS
JACKSONVILLE UNIVERSITY AND THE UNIVERSITY OF NORTH FLORIDA
JACKSONVILLE, FL

A few years later, in 1937, I. J. Schoenberg [**5**, §6] considered the quadratic form

$$\sum_{1 \le i, j \le n} |P_i - P_j|^{\alpha} x_i x_j \quad (0 < \alpha < 2),$$

where $P_1, \dots, P_n$ are distinct points in an euclidean space, and $|P_i - P_j|$ denote their euclidean distances. Schoenberg showed that this form is nonsingular with real signature $2 - n$, and that it is negative definite on the hyperplane $\sum_i x_i = 0$. In the case when $|P_i - P_j| = |i - j|$ and $\alpha = 1$, this, of course, also follows from our diagonalization of the form $q_4(\mathbf{x})$. I am very much indebted to Professor K. Stolarsky for pointing out the above pertinent references.

REFERENCES

1. B. Jacob, *Linear Algebra*, W. H. Freeman and Co., New York, NY, 1990.
2. T. A. Hannula, T. G. Rally, and I. Reiner, Modular representation algebras, *Bull. Amer. Math. Soc.* 73 (1967), 100–101.
3. K. Hoffman and R. Kunze, *Linear Algebra*, Prentice Hall, Englewood Cliffs, NJ, 1971.
4. R. M. Robinson, Problems 3667, 3705, *Amer. Math. Monthly* 41 (1934), p. 193 and p. 581.
5. I. J. Schoenberg, On certain metric spaces arising from euclidean spaces by a change of metric and their imbedding in Hilbert space, *Annals of Math.* 38 (1937), 787–793.
6. G. Szegö, Solution to Problem 3705, *Amer. Math. Monthly* 43 (1936), 246–259.

Math Bite: Irrationality of $\sqrt{m}$

A recent *Gazette* article [**3**] reminded me of Theodor Estermann's proof of the irrationality of $\sqrt{2}$. As his proof is not well known, I take the liberty of publicizing it with a slight "generalization." The really interesting thing about this proof is that it doesn't use divisibility, just mathematical induction in its "$\mathbb{Z}$ is well-ordered" form.

THEOREM. *Suppose m is not a perfect square. Then $\sqrt{m}$ is irrational.*

Proof. Let n be the integer with $n < \sqrt{m} < n + 1$. It suffices to prove that $\alpha = \sqrt{m} - n$ is irrational. Suppose not. As $0 < \alpha < 1$, we have $\alpha = p/q$ where $0 < p < q$. We may assume that q is *as small as possible* (Estermann's key idea). Then we have

$$\frac{q}{p} = \frac{1}{\sqrt{m} - n} = \frac{\sqrt{m} + n}{m - n^2} = \frac{\alpha + 2n}{m - n^2}.$$

We solve for α:

$$\alpha = \frac{(m - n^2)q}{p} - 2n = \frac{(m - n^2)q - 2np}{p} = \frac{r}{p}.$$

Thus α is a fraction with even smaller denominator, a contradiction.

REFERENCES

1. T. Estermann, The irrationality of $\sqrt{2}$, *Math. Gazette* 59 (1975), 110.
2. K. Roth and R. C. Vaughn, Obituary of Theodor Estermann, *Bull. London Math. Soc.* 26 (1994), 593–606.
3. M.-K. Siu, Estermann and Pythagoras, *Math. Gazette* 82 (1998), 92–93.

—HARLEY FLANDERS
JACKSONVILLE UNIVERSITY AND THE UNIVERSITY OF NORTH FLORIDA
JACKSONVILLE, FL

PROBLEMS

GEORGE T. GILBERT, *Editor*
Texas Christian University

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

Proposals

To be considered for publication, solutions should be received by November 1, 1999.

1574. *Proposed by Larry Hoehn, Austin Peay University, Clarksville, Tennessee.*

Let $ABCD$ be a convex quadrilateral. Prove or disprove: There exists a point E in the plane of $ABCD$ such that $\triangle ABE \sim \triangle CDE$.

1575. *Proposed by Peter Y. Woo, Biola University, La Mirada, California.*

Given a convex quadrilateral $ABCD$, find the ratio of the area of $\triangle ABD$ to that of $\triangle BCD$ in terms of $\angle BAC$, $\angle BCA$, $\angle DAC$, *and* $\angle DCA$.

1576. *Proposed by Mircea Radu, Bielefeld University, Bielefeld, Germany.*

For $\mathscr{P}$ a convex polygon, prove that the following are equivalent:

(1) The straight lines that divide $\mathscr{P}$ into two polygons of equal perimeter have a common point.
(2) The straight lines that divide $\mathscr{P}$ into two polygons of equal area have a common point.
(3) $\mathscr{P}$ is centrally symmetric.

1577. *Proposed by Philip Korman, University of Cincinnati, Cincinnati, Ohio.*

Consider the differential equation $x^n(t) + a(t)x^3(t) = 0$ on $0 \le t < \infty$, where $a(t)$ is continuously differentiable and $a(t) \ge \kappa > 0$.

(a) If $a'(t)$ has only finitely many changes of sign, prove that any solution $x(t)$ is bounded.
(b) If one does not assume that $a'(t)$ has only finitely many sign changes, is $x(t)$ necessarily bounded?

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

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

Solutions and new proposals should be mailed to George T. Gilbert, Problems Editor, Department of Mathematics, Box 298900, Texas Christian University, Fort Worth, TX 76129, or mailed electronically (ideally as a LATEX file) to `g.gilbert@tcu.edu`. *Readers who use e-mail should also provide an e-mail address.*

1578. *Proposed by Chu Wenchang, Paris, France, and Alberto Marini, IAMI-CNR-Milano, Italy.*

For nonnegative integers m, n, and p and complex numbers a, b, and z, define

$$S_{m,n,p}(a,b,z) := \sum_{k=0}^{m} (-1)^k \binom{m}{k} \frac{z}{z+k} \cdot \frac{(a+bk)^p}{\binom{z+k+n}{n}}.$$

For $p \leq \max\{m,n\}$, show that $S_{m,n,p}(a,b,z) = S_{n,m,p}(a-bz, -b, z)$.

Quickies

Answers to the Quickies are on page 242.

Q891. *Proposed by Keivan Mallahi, student, Sharif University of Technology, Tehran, Iran.*

Prove that a finite ring R for which $x^2 = x$ for $x \in R$ (a finite Boolean ring) has a multiplicative identity.

Q892. *Proposed by Jon Florin, Chur, Switzerland.*

Let $r \neq 1$ be a given positive real number. Let A and B be randomly chosen points on a rectangular piece of paper (based on the uniform distribution). What is the probability that the point X on the line AB but outside the segment AB which satisfies $XA/XB = r$ also lies on the paper?

Solutions

Pythagorean Triangles with Sides in an Interval June 1998

1549. *Proposed by K. R. S. Sastry, Bangalore, India.*

Given a positive integer k, prove that for all sufficiently large x, there exist at least k primitive Pythagorean triangles whose sides all have lengths in the interval $[x, 2x]$.

I. *Solution by Kathleen E. Lewis, SUNY Oswego, Oswego, New York.*

We will look at Pythagorean triples of the form $(p^2 - q^2,\ 2pq, p^2 + q^2)$ with $p = 4n + 1$ and $q = 2n$. We will refer to n as the *generator* of this triple. All such triples are primitive because $\gcd(p, q) = 1$. We have

$$p^2 - q^2 = 12n^2 + 8n + 1, \qquad 2pq = 16n^2 + 4n, \qquad p^2 + q^2 = 20n^2 + 8n + 1.$$

For all values of n greater than 1, $2pq$ is strictly between $p^2 - q^2$ and $p^2 + q^2$, so we need only ensure that these latter two values lie in the appropriate range.

For sufficiently large x, we will find that the triples generated by $n + 1, \ldots, n + k$ all lie in $[x, 2x]$. Suppose that $x > 12(11k)^2 + 8(11k) + 1$. Let n be the largest integer for which $x > 12n^2 + 8n + 1$. Observe that $n \geq 11k$. Then the values of $p^2 - q^2$ generated by $n+1, \ldots, n+k$ lie in $[x, \infty)$, so that it suffices to prove that the value of $p^2 + q^2$ generated by $n+k$ does not exceed $2x$. This holds because

$$\begin{aligned}
&2x - \left(20(n+k)^2 + 8(n+k) + 1\right)\\
&\quad > 2(12n^2 + 8n + 1) - \left(20(n+k)^2 + 8(n+k) + 1\right)\\
&\quad = 4n(n - 10k) - 20k^2 + 8n - 8k + 1\\
&\quad \geq 4(11k)(11k - 10k) - 20k^2 + 8(11k) - 8k + 1 = 24k^2 + 80k + 1 > 0.
\end{aligned}$$

II. *Solution by Kevin Ford, University of South Carolina, Columbia, South Carolina.*

All primitive Pythagorean triangles have sides of the form $(p^2 - q^2, 2pq, p^2 + q^2)$ where p and q are relatively prime positive integers of opposite parity. There is a one-to-one correspondence between the primitive Pythagorean triangles and such pairs (p, q) with $p > q$. Setting $u = px^{-1/2}$ and $v = qx^{-1/2}$, the conditions of the problem stipulate that each quantity $u^2 - v^2, 2uv, u^2 + v^2$ lies in $[1, 2]$. The inequalities define a region S in the first quadrant of the uv-plane. We will show that the number, $N(x)$, of primitive Pythagorean triangles with each side length in $[x, 2x]$ is given by the asymptotic formula

$$\begin{aligned}
N(x) &= \frac{4}{\pi^2} Ax + O(x^{1/2} \ln x)\\
&= \frac{4}{\pi^2}\left(\frac{\pi}{12} - \frac{1}{2}\ln\left(\frac{2+\sqrt{3}}{1+\sqrt{2}}\right)\right)x + O(x^{1/2} \ln x). \qquad (1)
\end{aligned}$$

Here A is the area of S. The factor $4/\pi^2$ is the "probability" that two random integers are relatively prime and have opposite parity. To prove (1), first let T denote the set of pairs of integers (a, b) with $(ax^{-1/2}, bx^{-1/2}) \in S$, let T_d denote the set of $(a, b) \in T$ with $d | a, d | b$, and let T_d' be the set of $(a, b) \in T_d$ with both a and b odd. Also let $\mu(n)$ denote the Möbius function. Then we have

$$\begin{aligned}
N(x) &= \sum_{\substack{(p,q)\in T\\ (p,q)=1\\ p\not\equiv q \pmod 2}} 1 = \sum_{\substack{(p,q)\in T\\ (p,q)=1}} 1 - \sum_{\substack{(p,q)\in T\\ (p,q)=1\\ p\equiv q\equiv 1 \pmod 2}} 1\\
&= \sum_{(p,q)\in T} \sum_{d|(p,q)} \mu(d) - \sum_{\substack{(p,q)\in T\\ p\equiv q\equiv 1 \pmod 2}} \sum_{d|(p,q)} \mu(d)\\
&= \sum_{d\leq\sqrt{2x}} \mu(d)|T_d| - \sum_{\substack{d\leq\sqrt{2x}\\ d \text{ odd}}} \mu(d)|T_d'|.
\end{aligned}$$

Let P be the perimeter of S. Since S is convex, we obtain

$$|T_d| = \frac{A}{d^2}x + O\left(\frac{P}{d}\sqrt{x}\right) = \frac{Ax}{d^2} + O\left(\frac{\sqrt{x}}{d}\right)$$

and likewise, for d odd,

$$|T_d'| = \frac{A}{4d^2}x + O\left(\frac{P}{d}\sqrt{x}\right) = \frac{Ax}{4d^2} + O\left(\frac{\sqrt{x}}{d}\right).$$

Therefore

$$
\begin{aligned}
N(x) &= \sum_{d \le \sqrt{2x}} \mu(d)\frac{Ax}{d^2} - \frac{1}{4}\sum_{\substack{d \le \sqrt{2x} \\ d \text{ odd}}} \mu(d)\frac{Ax}{d^2} + O\left(\sum_{d \le \sqrt{2x}} \frac{\sqrt{x}}{d}\right) \\
&= Ax\left(\sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} - \frac{1}{4}\sum_{\substack{d=1 \\ d \text{ odd}}}^{\infty} \frac{\mu(d)}{d^2}\right) + O(x^{1/2}\ln x + x^{1/2}) \\
&= Ax\left(\frac{1}{\zeta(2)} - \frac{1}{3\zeta(2)}\right) + O(x^{1/2}\ln x) = Ax\frac{4}{\pi^2} + O(x^{1/2}\ln x),
\end{aligned}
$$

where $\zeta(s)$ denotes the Riemann zeta function. To determine the area of S, we first note that S is the region lying between the curves $uv = \frac{1}{2}$, $u^2 + v^2 = 2$ and $u^2 - v^2 = 1$. Then

$$
\begin{aligned}
A &= \int_{\sqrt{(1+\sqrt{2})/2}}^{\sqrt{3/2}} \left(\sqrt{u^2 - 1} - \frac{1}{2u}\right) du + \int_{\sqrt{3/2}}^{(1+\sqrt{3})/2} \left(\sqrt{2 - u^2} - \frac{1}{2u}\right) du \\
&= \frac{\pi}{12} - \frac{1}{2}\ln\left(\frac{2+\sqrt{3}}{1+\sqrt{2}}\right) \approx 0.04400723.
\end{aligned}
$$

Also solved by the proposer. There were two incorrect solutions.

An Inequality in n Variables — June 1998

1550. *Proposed by Mihály Bencze, Braşov, Romania.*

Let z_i, $1 \le i \le n$, be complex and let $s_i = z_1 + z_2 + \cdots + z_i$, $1 \le i \le n$. Prove that

$$\sum_{1 \le i \le j \le n} |s_j - z_i| \le \sum_{k=1}^{n} \left((n+1-k)|z_k| + (k-2)|s_k|\right).$$

Solution by Michel Bataille, Rouen, France.

Suppose that we have already proved the inequality

$$\sum_{i=1}^{j} |s_j - z_i| \le \left(\sum_{i=1}^{j} |z_i|\right) + (j-2)\,|s_j| \qquad \text{for } j = 1, 2, \ldots, n. \tag{1}$$

Then the result follows from

$$
\begin{aligned}
\sum_{1 \le i \le j \le n} |s_j - z_i| &= \sum_{j=1}^{n}\sum_{i=1}^{j} |s_j - z_i| \\
&\le \sum_{j=1}^{n}\sum_{i=1}^{j} |z_i| + \sum_{j=1}^{n} (j-2)|s_j| \\
&= \sum_{k=1}^{n} (n+1-k)|z_k| + \sum_{k=1}^{n} (k-2)|s_k|.
\end{aligned}
$$

Inequality (1) is immediate for $j = 1, 2$, so we may suppose $j \ge 3$. Let

$$t_j := |z_1| + |z_2| + \cdots + |z_j|.$$

We obtain

$$\sum_{i=1}^{j}\left(t_j-|z_i|-|s_j-z_i|\right)\left(t_j-|z_i|+|s_j-z_i|\right)=\sum_{i=1}^{j}\left(t_j^2-2t_j|z_i|+|z_i|^2-|s_j-z_i|^2\right)$$

$$=(j-2)t_j^2+\sum_{i=1}^{j}\left(|z_i|^2-|s_j-z_i|^2\right)=(j-2)t_j^2-\sum_{i=1}^{j}\left(|s_j|^2-s_j\bar{z}_i-\bar{s}_j z_i\right)$$

$$=(j-2)\left(t_j^2-|s_j|^2\right)=(j-2)\left(t_j-|s_j|\right)\left(t_j+|s_j|\right).$$

On the other hand, the triangle inequality implies $0\le(t_j-|z_i|+|s_j-z_i|)\le t_j+|s_j|$. Because $t_j+|s_j|>0$ except in the trivial case $z_1=z_2=\cdots=z_n=0$, we get

$$(j-2)\left(t_j-|s_j|\right)\le\sum_{i=1}^{j}\left(t_j-|z_i|-|s_j-z_i|\right)=(j-1)t_j-\sum_{i=1}^{j}|s_j-z_i|,$$

from which (1) follows immediately.

Comment. Heinz-Jürgen Seiffert reports that inequality (1) appears in D. S. Mitrinović, J. E. Pečarič, and A. M. Fink, *Classical and New Inequalities in Analysis*, Kluwer (1993), 521–522.

Also solved by Heinz-Jürgen Seiffert and the proposer.

A Stirling-Type Approximation **June 1998**

1551. *Proposed by Howard Morris, Germantown, Tennessee.*

For which values of a is

$$\lim_{n\to\infty} n^2\ln\frac{\sqrt{2\pi}\,(n+a)^{n+1/2}\,e^{-n-a}}{n!}$$

finite?

Solution by Heinz-Jürgen Seiffert, Berlin, Germany.

The values of a are $(3\pm\sqrt{3})/6$.

Stirling's formula is

$$\ln(n!)=\left(n+\frac{1}{2}\right)\ln n-n+\ln\sqrt{2\pi}+\frac{1}{12n}+O(n^{-3}).$$

From

$$\ln\left(1+\frac{a}{n}\right)=\frac{a}{n}-\frac{1}{2}\left(\frac{a}{n}\right)^2+\frac{1}{3}\left(\frac{a}{n}\right)^3+O(n^{-4}),$$

we find

$$\left(n+\frac{1}{2}\right)\ln\left(1+\frac{a}{n}\right)=a+\frac{a-a^2}{2n}+\frac{4a^3-3a^2}{12n^2}+O(n^{-3}).$$

Now, it follows easily that

$$n^2\ln\frac{\sqrt{2\pi}\,(n+a)^{n+1/2}\,e^{-n-a}}{n!}=\frac{6a-6a^2-1}{12}n+\frac{4a^3-3a^2}{12}+O(n^{-1}).$$

Hence, the limit under consideration exists if and only if $6a - 6a^2 - 1 = 0$, which has roots $a = (3 \pm \sqrt{3})/6$, in which case the limit is $\pm \sqrt{3}/216$.

Also solved by Robert A. Agnew, Reza Akjlaghi, Daniele Donini (Italy), Kazuo Goto (Japan), Hans Kappus (Switzerland), Kee-Wai Lau (China), José H. Nieto (Venezuala), Sam Speed, Tiberiu V. Trif (Romania), Western Maryland College Problems Group, and the proposer.

A Functional Equation Satisfied by Complex Conjugation June 1998

1552. *Proposed by Wu Wei Chao, Guang Zhou Normal College, Guang Zhou City, Guang Dong Province, China.*

Find all functions $f\colon \mathbb{R} \to \mathbb{R}$ that satisfy

$$f(x + yf(x)) = f(x) + xf(y)$$

for all x and y.

Solution by Bassem Ghalayini and Ajaj Tarabay, Notre Dame University, Zouk Mikael, Lebanon.

We prove that f must be the zero function or the identity function on $\mathbb{R}$.

Both the zero function and the identity function satisfy the functional equation. Now assume that f satisfies the functional equation and is not identically zero. Putting $y = 0$ implies that $xf(0) = 0$ for all x, hence $f(0) = 0$. Conversely, if $f(x) = 0$, then $0 = xf(y)$ for all y; therefore $x = 0$.

With $x = 1$, we get

$$f(1 + yf(1)) = f(1) + f(y) \quad \forall y \in \mathbb{R}. \tag{1}$$

If $f(1) \neq 1$, then choosing $y = 1/(1 - f(1))$ in (1), we obtain $f(y) = f(1) + f(y)$, hence $f(1) = 0$, a contradiction. Thus $f(1) = 1$ and

$$f(1 + y) = 1 + f(y) \quad \forall y \in \mathbb{R}. \tag{2}$$

In particular, $f(n) = n$ for all $n \in \mathbb{Z}$. Substituting $x = n \in \mathbb{Z}$ and $y = z - 1$ into the functional equation, we get

$$f(nz) = f(n + (z-1)f(n)) = n + nf(z-1) = nf(z) \quad \forall n \in \mathbb{Z},\ z \in \mathbb{R}.$$

It follows immediately that

$$f(rz) = rf(z) \quad \forall r \in \mathbb{Q},\ z \in \mathbb{R}. \tag{3}$$

We next show that f is additive. It is clear from (3) that $f(a) + f(-a) = 0 = f(a - a)$. For $a + b \neq 0$, write

$$f(a) + f(b) = f\left(\frac{a+b}{2} + \frac{\frac{a-b}{2}}{f\left(\frac{a+b}{2}\right)} f\left(\frac{a+b}{2}\right) \right) + f\left(\frac{a+b}{2} + \frac{\frac{b-a}{2}}{f\left(\frac{a+b}{2}\right)} f\left(\frac{a+b}{2}\right) \right).$$

Applying the functional equation and (3), we obtain

$$\begin{aligned} f(a) + f(b) &= f\left(\frac{a+b}{2}\right) + \frac{a+b}{2} f\left(\frac{\frac{a-b}{2}}{f\left(\frac{a+b}{2}\right)} \right) + f\left(\frac{a+b}{2}\right) + \frac{a+b}{2} f\left(\frac{\frac{b-a}{2}}{f\left(\frac{a+b}{2}\right)} \right) \\ &= 2f\left(\frac{a+b}{2}\right) = f(a+b). \end{aligned}$$

Applying additivity to the functional equation, we get $f(yf(x)) = xf(y)$. Setting $y = 1$ yields $f(f(x)) = x$, so that f is bijective. Replacing x with $f(x)$, we get

$$f(xy) = f(x)f(y) \quad \forall x, y \in \mathbb{R}. \tag{4}$$

Therefore, f is a field automorphism. It is well-known that f must be the identity. It follows easily here by noting that (4) with $y = \pm x$ implies that $f(z) > 0$ if and only if $z > 0$. Because

$$f(x - f(x)) = f(x) - x = -(x - f(x)),$$

f must be the identity function.

Also solved by Michael Bataille (France), Mansur Boase (student, England), and the proposer. There were four incorrect solutions and two incomplete solutions.

Roots of Polynomials with Positive Coefficients **June 1998**

1553. *Proposed by Paul Zorn, St. Olaf College, Northfield, Minnesota.*

What complex numbers are the root of some polynomial with positive coefficients?

Solution by Grand Valley State University Problem Group, Allendale, Michigan.

We show that every complex number that does not lie on the nonnegative real line is a root of such a polynomial. (If we allow coefficients of terms of degree less than the degree of the polynomial to be 0, then only positive real numbers are excluded.)

First, observe that if w is a nonnegative real number and $P(z)$ is a polynomial with positive coefficients, then clearly $P(w) > 0$. Thus no $w \geq 0$ is the root of a polynomial with positive coefficients.

Every $w = \alpha + \beta i \in \mathbb{C}$ is a root of the real polynomial

$$q(z) = (z - w)(z - \overline{w}) = z^2 - 2\alpha z + \alpha^2 + \beta^2.$$

If $\alpha < 0$, then $q(z)$ is a polynomial with positive coefficients. This shows that every complex number in the open left half-plane is the root of some such polynomial.

Now assume that $\alpha \geq 0$ and $\beta \neq 0$. Thus w lies in the right half-plane, but not on the real axis, so $0 < |\arg(w)| \leq \pi/2$. If n is the smallest positive integer for which $\pi/2 < |\arg(w^n)|$, then w^n lies in the open left half-plane. By our work above, there exists a quadratic polynomial q with positive coefficients such that $q(w^n) = 0$, say $q(z) = z^2 + c_1 z + c_0$. We now let

$$P(z) = \left(z^{2n} + c_1 z^n + c_0\right)\left(z^{n-1} + z^{n-2} + \cdots + z + 1\right).$$

It follows that $P(z)$ has positive coefficients and $P(w) = 0$.

Also solved by Roy Barbara (Lebanon), Michel Bataille (France), Brian D. Beasley, John Christopher, Charles R. Diminnie, Daniele Domini (Italy), Mordechi Falkowitz (Canada), Micah Fogel, Kevin Ford, Gerald A. Heuer, Dean Larson, Tiberiu V. Trif (Romania), Western Maryland College Problems Group, and the proposer.

Answers

Solutions to the Quickies on page 237.

A891. It is a standard exercise to show that $x^2 = x$ for all $x \in R$ implies that R is commutative. Let $R = \{x_1, x_2, \ldots, x_n\}$, and set

$$e = \sum_{1 \leq i \leq n} x_i - \sum_{1 \leq i < j \leq n} x_i x_j + \cdots + (-1)^n \sum_{1 \leq i \leq n} x_1 \cdots x_{i-1} x_{i+1} \cdots x_n$$
$$+ (-1)^{n+1} x_1 x_2 \cdots x_n.$$

Then

$$\begin{aligned} x_k e = {} & \left(x_k + \sum_{\substack{1\le i\le n\\ i\ne k}} x_k x_i\right) - \left(\sum_{\substack{1\le i\le n\\ i\ne k}} x_k x_i + \sum_{\substack{1\le i<j\le n\\ i\ne k,\, j\ne k}} x_k x_i x_j\right) + \cdots \\ & + (-1)^n \left(\sum_{\substack{1\le i\le n\\ i\ne k}} x_1 \cdots x_{i-1} x_{i+1} \cdots x_n + x_1 x_2 \cdots x_n\right) + (-1)^{n+1} x_1 x_2 \cdots x_n \\ = {} & x_k . \end{aligned}$$

A892. The probability is $(1-r)^2$ for $r<1$ and $(1-1/r)^2$ for $r>1$.

Let R denote the rectangle. First assume $r<1$. Observe that X satisfies $XB/AB = 1/(1-r)$. Fix B and let R_B denote the image of R under the centric extension with center B and dilation factor $1-r$. Thus for X to lie inside R, the point A has to lie inside R_B. Noting that R_B lies entirely inside R, the probability of choosing an appropriate A is the ratio of the areas of R_B and R, which is $(1-r)^2$. The probability for $r>1$ is the same as that for $1/r$ because of symmetry, hence the result.

The same probabilities hold for a piece of paper in any convex shape.

REVIEWS (continued from page 244)

Walker, Thomas J., Free internet access to traditional journals, *American Scientist* 86 (5) (September-October 1998). Text plus additional author's notes at `http://www.amsci.org/amsci/articles/98articles/walker.html` .

What is the future of academic journals? Some of our colleagues urge that mathematicians follow the lead of physicists, who now archive more than 50% of their preprints in electronic form ("e-prints") at `http://xxx.lanl.gov` . Since the beginning of 1998, this archive has also been trying to attract mathematics papers in all fields. The idea is that journals would become "overlay" electronic journals. They would operate by mathematicians requesting that the journal consider their e-prints that are already archived, the editor would ask referees to review the papers (no need to send paper copies by mail), and the journal would then publish electronically a table of contents of accepted papers (with whatever revisions). What's missing here? There's still paper (print your own copy), but no publishers, no page charges (you typeset your own paper), no mailing costs, no cost to libraries, no $100 million in revenues to publishers and costs to libraries, no need to visit a library. Author Walker has contemplated the "toll-gate" approaches to the future for journals (online subscriptions, site licenses, and pay per view) and rejected them. He proposes a different model, used by his Florida Entomological Society. It still publishes a printed journal but devotes a small portion of the page charges to producing electronic copies (in Adobe Acrobat PDF format). In lieu of ordering reprints, authors may order "e-prints": They can pay an additional page fee to make the PDF versions of their articles available immediately to everyone for free (otherwise, the PDF version goes online at the journal's site one year later.) As Walker notes, going solely electronic would vastly reduce costs. He argues "free access to traditional journals is affordable and achievable. It is the right thing to do for those who pay for the research and for those who do it." You can join a discussion of these issue at `http://www.amsci.org/amsci/editors/forum.html` .

REVIEWS

PAUL J. CAMPBELL, *editor*
Beloit College

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

Fink, Thomas M., and Yong Mao, Designing tie knots by random walks *Nature* 398 (4 March 1999) 31–32. Chang, Kenneth, A puzzle fit to be tied: Applying math to knots for the neck, ABC Evening News (4 March 1999), `http://www.abcnews.go.com/sections/science/DailyNews/tieknots990303.html` .

Knot theory finally reaches the man on the street. Two physicists (why wasn't it mathematicians?) at Cambridge University have enumerated all of the knots that can be tied in a necktie in at most nine moves. Of the total of 85, they find 10 to be esthetically pleasing; these include the traditional four-in-hand, Windsor, and half-Windsor, and the newer Shelby, plus six new and as yet unnamed knots. The researchers did the enumeration by considering walks on a triangular lattice (the "active" tie end goes to the left, to the right, or through the center).

Eglash, Ron, *African Fractals: Modern Computing and Indigenous Design*, Rutgers University Press, 1999; xi + 258 pp, \$60, \$25 (P). ISBN 0–8135–2614–0, 0–8135–2613–2.

Author Eglash identifies a spectrum for the presence of mathematics in culture. Unconscious structures (such as beehives) "do not count as mathematical knowledge, even though we can use mathematics to describe them." Intentional structures, such as decorative design, may involve mathematics only implicitly (as in scaling), or else explicitly (as in named patterns and rules), or even as rules for how patterns can be combined (as in applied mathematics of various kinds). Finally, "pure mathematics" comprises abstract theories of why the rules work and how to find new patterns. With these distinctions in mind, Eglash investigates the fractal nature of geometric patterns in African art, architecture, hairstyles, religion, and symbol systems. He also looks into the social implications and political consequences of this fractalism, and makes comparison to fractal structures in European culture.

Kunzig, Robert, The physics of ... traffic, *Discover* 20 (3) (March 1999) 31–32. Why does traffic jam? Segment of "Life's Little Questions ... and Some Very Big Answers," show 904 of the TV series *Scientific American Frontiers*, aired 24 February 1999; transcript at `http://www.pbs.org/saf/8_resources/83_transcript_904b.html#part4` . Hayes, Brian, *E pluribus unum*, *American Scientist* 87 (1) (January/February 1999) `http://www.amsci.org/amsci/issues/Comsci99/compsci1999-01.html` . Cellular Automaton Traffic Simulators, `http://www.theo2.physik.uni-stuttgart.de/helbing/RoadApplet/` .

Traffic jams sometimes seem to appear out of nowhere. New computer simulations of traffic flow by cellular automata offer insights that help to explain this observation and also reveal many distinct phases of traffic flow between congested traffic and a traffic jam. "[M]ost types of congestion are avoidable—they aren't caused by overloading of the highway but by small disturbances that grow and at some point cause the traffic to break down," says Dirk Helbinn (University of Stuttgart), who offers an online and downloadable Java applet for traffic simulation.

Then

$$
\begin{aligned}
x_k e &= \Bigg(x_k + \sum_{\substack{1 \le i \le n \\ i \ne k}} x_k x_i \Bigg) - \Bigg(\sum_{\substack{1 \le i \le n \\ i \ne k}} x_k x_i + \sum_{\substack{1 \le i < j \le n \\ i \ne k, j \ne k}} x_k x_i x_j \Bigg) + \cdots \\
&\quad + (-1)^n \Bigg(\sum_{\substack{1 \le i \le n \\ i \ne k}} x_1 \cdots x_{i-1} x_{i+1} \cdots x_n + x_1 x_2 \cdots x_n \Bigg) + (-1)^{n+1} x_1 x_2 \cdots x_n \\
&= x_k .
\end{aligned}
$$

A892. The probability is $(1-r)^2$ for $r < 1$ and $(1 - 1/r)^2$ for $r > 1$.

Let R denote the rectangle. First assume $r < 1$. Observe that X satisfies $XB/AB = 1/(1-r)$. Fix B and let R_B denote the image of R under the centric extension with center B and dilation factor $1 - r$. Thus for X to lie inside R, the point A has to lie inside R_B. Noting that R_B lies entirely inside R, the probability of choosing an appropriate A is the ratio of the areas of R_B and R, which is $(1-r)^2$. The probability for $r > 1$ is the same as that for $1/r$ because of symmetry, hence the result.

The same probabilities hold for a piece of paper in any convex shape.

REVIEWS (continued from page 244)

Walker, Thomas J., Free internet access to traditional journals, *American Scientist* 86 (5) (September-October 1998). Text plus additional author's notes at `http://www.amsci.org/amsci/articles/98articles/walker.html` .

What is the future of academic journals? Some of our colleagues urge that mathematicians follow the lead of physicists, who now archive more than 50% of their preprints in electronic form ("e-prints") at `http://xxx.lanl.gov` . Since the beginning of 1998, this archive has also been trying to attract mathematics papers in all fields. The idea is that journals would become "overlay" electronic journals. They would operate by mathematicians requesting that the journal consider their e-prints that are already archived, the editor would ask referees to review the papers (no need to send paper copies by mail), and the journal would then publish electronically a table of contents of accepted papers (with whatever revisions). What's missing here? There's still paper (print your own copy), but no publishers, no page charges (you typeset your own paper), no mailing costs, no cost to libraries, no $100 million in revenues to publishers and costs to libraries, no need to visit a library. Author Walker has contemplated the "toll-gate" approaches to the future for journals (online subscriptions, site licenses, and pay per view) and rejected them. He proposes a different model, used by his Florida Entomological Society. It still publishes a printed journal but devotes a small portion of the page charges to producing electronic copies (in Adobe Acrobat PDF format). In lieu of ordering reprints, authors may order "e-prints": They can pay an additional page fee to make the PDF versions of their articles available immediately to everyone for free (otherwise, the PDF version goes online at the journal's site one year later.) As Walker notes, going solely electronic would vastly reduce costs. He argues "free access to traditional journals is affordable and achievable. It is the right thing to do for those who pay for the research and for those who do it." You can join a discussion of these issue at `http://www.amsci.org/amsci/editors/forum.html` .

NEWS AND LETTERS

Twenty-Seventh Annual USA Mathematical Olympiad – Problems and Solutions

1. Suppose that the set $\{1,2,\cdots,1998\}$ has been partitioned into disjoint pairs $\{a_i,b_i\}$ $(1 \le i \le 999)$ so that for all i, $|a_i - b_i|$ equals 1 or 6. Prove that the sum

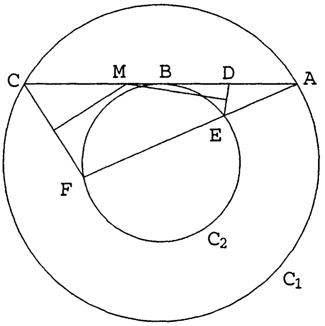
$$|a_1 - b_1| + |a_2 - b_2| + \cdots + |a_{999} - b_{999}|$$

ends in the digit 9.

Solution. Let k denote the number of pairs $\{a_i,b_i\}$ with $|a_i - b_i| = 6$. Then the sum in question is $k \cdot 6 + (999 - k) \cdot 1 = 999 + 5k$, which ends in 9 provided k is even. Hence it suffices to show that k is even.

Write $k = k_{\text{odd}} + k_{\text{even}}$, where k_{odd} (resp. k_{even}) is equal to the number of pairs $\{a_i,b_i\}$ with a_i, b_i both odd (resp. even). Since there are as many even numbers as odd numbers between 1 and 1998, and since each pair $\{a_i,b_i\}$ with $|a_i - b_i| = 1$ contains one number of each type, we must have $k_{\text{odd}} = k_{\text{even}}$. Hence $k = k_{odd} + k_{even}$ is even as claimed.

2. Let C_1 and C_2 be concentric circles, with C_2 in the interior of C_1. From a point A on C_1 one draws the tangent AB to C_2 $(B \in C_2)$. Let C be the second point of intersection of AB and C_1, and let D be the midpoint of AB. A line passing through A intersects C_2 at E and F in such a way that the perpendicular bisectors of DE and CF intersect at a point M on AB. Find, with proof, the ratio AM/MC.



Solution. Writing the power of A with respect to C_2 we get $AE \cdot AF = AB^2$. On the other hand, $AD \cdot AC = (AB/2) \cdot 2AB = AB^2$. Hence $AE \cdot AF = AD \cdot AC$. This shows that triangles ADE and AFC (with the shared angle at A) are similar. Thus, $\angle AED = \angle ACF$, so $DEFC$ is cyclic. Since M is the intersection of the perpendicular bisectors of DE and CF, it must be the circumcenter of $DEFC$. Consequently, M also lies on the perpendicular bisector of CD. Since M is on AC, it must be the midpoint of CD. Hence $AM/MC = 5/3$.

3. Let $a_0, a_1, \cdots, a_n$ be numbers from the interval $(0, \pi/2)$ such that

$$\tan(a_0 - \frac{pi}{4}) + \tan(a_1 - \frac{\pi}{4}) + \cdots + \tan(a_n - \frac{\pi}{4}) \geq n - 1.$$

Prove that

$$\tan a_0 \tan a_1 \cdots \tan a_n \geq n^{n+1}.$$

First Solution. Let $b_k = \tan(a_k - \pi/4)$, $k = 0, 1, \cdots, n$. It follows from the hypothesis that for each k, $-1 < b_k < 1$, and

$$1 + b_k \geq \sum_{0 \leq l \neq k \leq n} (1 - b_l). \tag{1}$$

Applying the Arithmetic-Geometric-Mean Inequality to the positive numbers $1-b_l$, $l = 0, 1, \ldots, k-1, k+1, \ldots, n$, we obtain

$$\sum_{0 \leq l \neq k \leq n} (1 - b_l) \geq n \left(\prod_{0 \leq l \neq k \leq n} (1 - b_l) \right)^{1/n}. \tag{2}$$

From (1) and (2) it follows that

$$\prod_{k=0}^{n} (1 + b_k) \geq n^{n+1} \left(\prod_{l=0}^{n} (1 - b_l)^n \right)^{1/n},$$

and hence that

$$\prod_{k=0}^{n} \frac{1 + b_k}{1 - b_k} \geq n^{n+1}.$$

Because

$$\frac{1 + b_k}{1 - b_k} = \frac{1 + \tan(a_k - \frac{\pi}{4})}{1 - \tan(a_k - \frac{\pi}{4})} = \tan\left(\left(a_k - \frac{\pi}{4}\right) + \frac{\pi}{4}\right) = \tan a_k,$$

the conclusion follows.

Second Solution. We first prove a short lemma:

Let w, x, y, z be real numbers with $x + y = w + z$ and $|x - y| < |w - z|$. Then $wz < xy$.

Proof: Let $x + y = w + z = 2L$. Then there are non-negative numbers r, s with $r < s$ and

$$wz = (L - s)(L + s) < (L - r)(L + r) = xy.$$

We now use this lemma in solving the problem. For $0 \leq k \leq n$, let $b_k = \tan(a_k - \pi/4)$ and let

$$t_k = \tan a_k = \frac{1 + b_k}{1 - b_k}.$$

Then $-1 < b_k < 1$ and

$$t_j t_k = \left(\frac{1 + b_j}{1 - b_j}\right)\left(\frac{1 + b_k}{1 - b_k}\right) = 1 + \frac{2}{\frac{1 + b_j b_k}{b_j + b_k} - 1}. \tag{1}$$

First note that because $-1 < b_k < 1$ and $b_0 + b_1 + \cdots + b_n \geq n - 1$, it follows that $b_j + b_k > 0$ for all $0 \leq j, k \leq n$ with $j \neq k$. Next note that if $b_j + b_k > 0$ and $b_j \neq b_k$, then it follows from the lemma applied to (1) that the value of $t_j t_k$ can be made smaller by replacing b_j and b_k by two numbers closer together and with the same sum. In particular, if $b_j < 0$, then replacing b_j and b_k by their average reduces the problem to the case where $b_i > 0$ for all i.

We may now successively replace the b_i's by their arithmetic mean. As long as the b_i are not all equal, one is greater than the mean and another one is less than the mean. We can replace one of this pair by the arithmetic mean of all of the b_i's, and the other by a positive number chosen so that the sum of the pair does not change. Each such change decreases the product of the t_i's. It follows that for a given sum of the b_i's, the minimum product is attained when all of the b_i's are equal. In this case we have $b_i \geq \frac{n-1}{n+1}$, for each i, so

$$t_0 t_1 \cdots t_n \geq \left(\frac{1 + \frac{n-1}{n+1}}{1 - \frac{n-1}{n+1}}\right)^{n+1} = \left(\frac{2n}{2}\right)^{n+1} = n^{n+1}.$$

This completes the proof.

Third Solution. We present a solution based on calculus. Though all Olympiad problems can be solved without calculus, solutions based on calculus are acceptable and may be instructive. We set

$$a = b_0 + b_1 + \cdots + b_n,$$

where $-1 < b_i < 1$, and assume that $a \geq n - 1$. We then show that the product

$$\prod_{k=0}^{n} \frac{1 + b_k}{1 - b_k}$$

attains its minimum when all of the b_k's are equal, that is, their common value is $a/(n+1)$. The desired inequality will follow immediately.

We proceed by induction. The case $n = 1$ was established in the discussion of (1) in the previous solution. For $n \geq 2$, set

$$\sum_{k=0}^{n-1} b_k = a' = a - b_n > n - 2.$$

The last inequality follows from $a \geq n - 1$ and $b_n < 1$. Set $b = b_n$ and $c = a'/n$, so $b + nc = a$. By the induction hypothesis,

$$\left(\prod_{k=0}^{n-1} \frac{1 + b_k}{1 - b_k}\right) \frac{1 + b_n}{1 - b_n} \geq \left(\frac{1 + c}{1 - c}\right)^n \frac{1 + b}{1 - b}.$$

Thus we need to prove that

$$\left(\frac{1 + c}{1 - c}\right)^n \left(\frac{1 + b}{1 - b}\right) \geq \left(\frac{n + 1 + a}{n + 1 - a}\right)^{n+1}, \tag{1}$$

where the right hand side is obtained by substituting $a/(n+1)$ for each b_k, $k = 0, 1, \ldots, n$, in the product. Next, recall that a is fixed, and that $b + nc = a$. Thus we can eliminate b from (1) to obtain the equivalent inequality

$$\left(\frac{1 + c}{1 - c}\right)^n \left(\frac{1 + a - nc}{1 - a + nc}\right) \geq \left(\frac{n + 1 + a}{n + 1 - a}\right)^{n+1}. \tag{2}$$

Now bring all terms in (2) to the left side of the inequality, clear denominators, and replace c by x. Let the expression on the left define a function f with

$$f(x) = (1+x)^n(1+a-nx)(n+1-a)^{n+1} - (1-x)^n(1-a+nx)(n+1+a)^{n+1}.$$

To establish (2) it is sufficient to show that for $0 \leq x < 1$, $f(x)$ attains its minimum value at $x = a/(n+1)$. Towards this end we differentiate to obtain

$$\begin{aligned} f'(x) &= n(a-(n+1)x)\left((1+x)^{n-1}(n+1-a)^{n+1} - (1-x)^{n-1}(n+1+a)^{n+1}\right) \\ &= n(a-(n+1)x)g(x), \end{aligned}$$

where $g(x) = (1+x)^{n-1}(n+1-a)^{n+1} - (1-x)^{n-1}(n+1+a)^{n+1}$. It is clear that $f'\left(\dfrac{a}{n+1}\right) = 0$, so we check the second derivative. We find

$$f''\left(\frac{a}{n+1}\right) = -n(n+1)g\left(\frac{a}{n+1}\right) > 0,$$

so f has a local minimum at $x = a/(n+1)$. But $f'(x)$ could have another zero, t, obtained by solving the equation $g(x) = 0$. Because

$$g'(x) = (n-1)(1+x)^{n-2}(n+1-a)^{n+1} + (n-1)(1-x)^{n-2}(n+1+a)^{n+1}$$

is obviously positive for all $x \in [0,1)$, there is at most one solution to the equation $g(x) = 0$ in this interval. It is easy to check that $g(a/(n+1)) < 0$ and $g(1) > 0$. Thus there is a real number t, $a/(n+1) < t < 1$, with $g(t) = 0$. For this t we have

$$f''(t) = n(a-(n+1)t)g'(t) < 0.$$

Thus, t is a local maximum for f, and no other extrema exist on the interval $(0,1)$.

The only thing left is to check that $f(1) \geq f(a/(n+1))$. Note that the case $x = 1$ is also an extreme case with $b_0 = b_1 = \cdots = b_{n-1} = 1$. This case does not arise in our problem, but we must check it to be sure that on the interval $0 \leq x < 1$, $f(x)$ has a minimum at $x = a/(n+1)$. We have

$$f(1) = 2^n(1+a-n)(n+1-a)^{n+1} \geq 0,$$

since $n-1 \leq a \leq n+1$, and $f(a/(n+1)) = 0$ (by design). Thus $f(x)$ indeed attains a unique minimum at $x = a/(n+1)$.

4. A computer screen shows a 98 × 98 chessboard, colored in the usual way. One can select with a mouse any rectangle with sides on the lines of the chessboard and click the mouse button: as a result, the colors in the selected rectangle switch (black becomes white, white becomes black). Find, with proof, the minimum number of mouse clicks needed to make the chessboard all one color.

Solution. More generally, we show that the minimum number of selections required for an $n \times n$ chessboard is $n-1$ if n is odd, and n if n is even. Consider the $4(n-1)$ squares along the perimeter of the chessboard, and at each step, let us count the number of pairs of adjacent perimeter squares which differ in color. This total begins at $4(n-1)$, ends up at 0, and can decrease by no more than 4 each turn (If the rectangle touches two adjacent edges of the board, then only two pairs can be affected. Otherwise, the rectangle either touches no edges, one edge, or two opposite edges, in which case 0, 2 or 4 pairs change, respectively). Hence at least $n-1$ selections are always necessary.

If n is odd, then indeed $n-1$ selections suffice, by choosing every second, fourth, sixth, etc. row and column. However, if n is even, then $n-1$ selections cannot suffice: at some point a corner square must be included in a rectangle (since the corners do not all begin having the same color), and such a rectangle can only decrease the above count by 2. Hence n selections are needed, and again by selecting every other row and column, we see that n selections also suffice.

5. Prove that for each $n \geq 2$, there is a set S of n integers such that $(a-b)^2$ divides ab for every distinct $a, b \in S$.

Solution. We will prove by induction on n, that we can find such a set, all of whose elements are *nonnegative*. For $n = 2$, we may take $S = \{0, 1\}$.

Now suppose that for some $n \geq 2$, the desired set S_n of n nonnegative integers exists. Let L be the least common multiple of $(a-b)^2$ and ab, with (a, b) ranging over pairs of distinct elements from S_n. Define $S_{n+1} = \{L + a \ : \ a \in S\} \cup \{0\}$. Then S_{n+1} consists of $n+1$ nonnegative integers, since $L > 0$. If $\alpha, \beta \in S_{n+1}$ and either α of β is zero, then $(\alpha - \beta)^2$ divides $\alpha\beta$. If $L + a, L + b \in S_{n+1}$, with a, b distinct elements of S_n, then

$$(L+a)(L+b) \equiv ab \equiv 0 \ (\mathrm{mod}(a-b)^2),$$

so $[(L+a)-(L+b)]^2$ divides $(L+a)(L+b)$, completing the inductive step.

6. Let $n \geq 5$ be an integer. Find the largest integer k (as a function of n) such that there exists a convex n-gon $A_1A_2 \ldots A_n$ for which exactly k of the quadrilaterals $A_iA_{i+1}A_{i+2}A_{i+3}$ have an inscribed circle. (Here $A_{n+j} = A_j$.)

Solution. The maximum is $\lfloor \frac{n}{2} \rfloor$. We first establish the upper bound by showing that if A, B, C, D, and E are consecutive vertices of the n-gon, then the quadrilaterals $ABCD$ and $BCDE$ cannot both have inscribed circles.

Assume the contrary. By equal tangents,

$$AB + CD = BC + AD$$
$$BC + DE = CD + BE$$

and so $AB + DE = AD + BE$. On the other hand, if O is the intersection of AD and BE, then by the triangle inequality, $AO + OB > AB$ and $OD + OE > DE$, so

$$AD + BE = AO + OB + OD + OE > AB + DE,$$

a contradiction.

Now we give a construction to show that $\lfloor \frac{n}{2} \rfloor$ circumscribing quadrilaterals are possible. First suppose that n is even. Draw an isosceles trapezoid with base angles $\frac{2\pi}{n}$ and admitting an inscribed circle. Let x be the length of the shorter base and y the length of either leg in the trapezoid. Then an equiangular n-gon with side lengths $x, y, x, y, \ldots$ clearly gives $n/2$ circumscribing quadrilaterals.

Now suppose that n is odd. Construct an $(n+1)$-gon $A_1 \dots A_{n+1}$ yielding $(n+1)/2$ circumscribing quadrilaterals as described in the previous paragraph. Now erase A_{n+1} and move A_n to a new position so that the quadrilateral $A_{n-2}A_{n-1}A_nA_1$ has an inscribed circle; this gives $(n-1)/2$ circumscribing quadrilaterals, as desired.

Thirty-Ninth Annual International Mathematical Olympiad – Problems

1. In the convex quadrilateral $ABCD$, the diagonals AC and BD are perpendicular and the opposite sides AB and DC are not parallel. Suppose that the point P, where the perpendicular bisectors of AB and DC meet, is inside $ABCD$. Prove that $ABCD$ is a cyclic quadrilateral if and only if the triangles ABP and CDP have equal areas.

2. In a competition, there are a contestants and b judges, where $b \geq 3$ is an odd integer. Each judge rates each contestant as either "pass" or "fail". Suppose k is a number such that, for any two judges, their ratings coincide for at most k contestants. Prove that

$$\frac{k}{a} \geq \frac{b-1}{2b}.$$

3. For any positive integer n, let $d(n)$ denote the number of positive divisors of n (including 1 and n itself). Determine all positive integers k such that

$$\frac{d(n^2)}{d(n)} = k$$

for some n.

4. Determine all pairs (a, b) of positive integers such that $ab^2 + b + 7$ divides $a^2b + a + b$.

5. Let I be the incenter of triangle ABC. Let the incircle of ABC touch the sides BC, CA, and AB at K, L, and M, respectively. The line through B parallel to MK meets the lines LM and LK at R and S, respectively. Prove that angle RIS is acute.

6. Consider all functions f from the set $\mathbb{N}$ of all positive integers into itself satisfying $f(t^2f(s)) = s(f(t))^2$ for all s and t in $\mathbb{N}$. Determine the least possible value of $f(1998)$.

Notes

The top eight students on the 1998 USAMO were (in alphabetical order):

Reid Barton	Arlington, MA
Gabriel Carroll	Oakland, CA
Kevin Lacker	Cincinnati, OH
Alexander Schwartz	Radnor, PA
David Speyer	Wallingford, CT
Paul Valiant	Milton, MA
David Vickrey	Vermillion, SD
Melanie Wood	Indianapolis, IN

Melanie Wood and Alexander Schwartz tied as winners of the Greitzer-Klamkin award, given to the top scorer on the USAMO. Members of the USA team at the 1998 IMO (Taipei, Taiwan) were Reid Barton, Gabriel Carroll, Kevin Lacker, Alexander Schwartz, Paul Valiant, and Melanie Wood. Barton, Carroll, and Schwartz received gold medals, while Lacker, Valiant, and Wood received silver medals. In terms of total score, the highest ranking of the seventy-six participating teams were as follows:

Iran	211	Russia	175
Bulgaria	195	India	174
United States	186	Ukraine	166
Hungary	186	Vietnam	158
Taiwan	184	Yugoslavia	156

The 1998 USA Mathematical Olympiad was prepared by Titu Andreescu (chair), Elgin Johnston, Jim Propp, Alexander Soifer, Richard Stong, and Paul Zeitz. The training program to prepare the USA team for the IMO (the Mathematical Olympiad Summer Program) was held at the University of Nebraska, Lincoln, NE. Titu Andreescu (Director), Zuming Feng, Razvan Gelca, Elgin Johnston, Kiran Kedlaya, and Zvezdelina Stankova-Frankel served as instructors, assisted by Carl Bosley and Noam Shazeer.

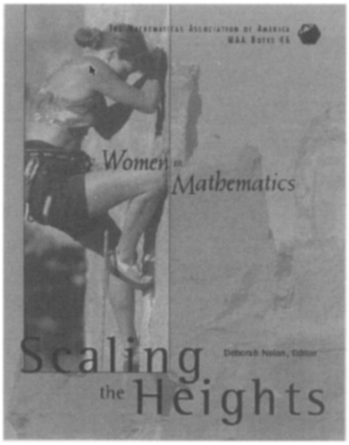
The booklet *Mathematical Olympiads 1998* presents additional solutions to problems on the 27th USAMO and solutions to the 39th IMO. Such a booklet has been published every year since 1976. Copies are $5.00 for each year 1976–1998. They are available from:

> Titu Andreescu, American Mathematics Competitions, 1740 Vine Street,
> Lincoln, NE 68588-0658.

The USA Mathematical Olympiad, participation of the US team in the International Mathematical Olympiad, and the sequence of examinations leading to qualification for these Olympiads are under the administration of the M.A.A. Committee on American Mathematical Competitions, and these activities are sponsored by eight organizations of professional mathematicians. For further information about this sequence of examinations, contact the Executive Director of the Committee, Titu Andreescu, at the above address.

This report was prepared by Titu Andreescu and Elgin Johnston.

THE MATHEMATICAL ASSOCIATION OF AMERICA



Women in Mathematics

Scaling the Heights

Deborah Nolan, Editor

Series: MAA Notes

Motivate your students to study advanced mathematics

Women in Mathematics: Scaling the Heights will provide you with a wealth of ideas and examples you can use to develop programs and courses that will motivate undergraduates who want to study advanced mathematics. The articles in this collection make a unique and valuable contribution to upper-division undergraduate mathematics education. They show us what talented undergraduates can do.

The heart of this book presents the insights of eight individuals who have taught at the Summer Mathematics Institute at Mills College. They share their course materials and give pedagogical tips on how to teach topics in mathematics that are not ordinarily part of the undergraduate curriculum, and in ways not often found in the undergraduate classroom. Although the courses described here were designed to encourage talented undergraduate women to pursue advanced degrees in mathematics, the good ideas found in them are gender free and can be used equally well with male as well as female students.

Exercises, class handouts, lists of research projects, and references are included. Topics covered are algebraic coding theory, hyperplane arrangements, p-adic numbers, quadratic reciprocity, stochastic processes, and linear optimization.

The book rounds out the material presented by the Summer Mathematics Institute instructors, with perspectives from mathematicians who have been active in the promotion of women in the field. Results from a survey of undergraduate mathematics majors in which they tell us what they think about the major and their future in mathematics complements these essays.

Catalog Code: NTE-46/JR
146 pp., Paperbound, 1997
ISBN 0-88385-156-3
List: $29.95 MAA Member: $23.95

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

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

Name ____________________

Address ____________________

City ____________ State _____ Zip ________

Phone ____________________

Qty.	Catalog Code	Price	Amount
______	NTE-46/JR	______	______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Shipping & handling ______

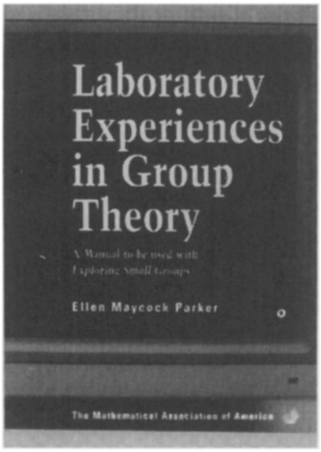
TOTAL ______

Payment □ Check □ VISA □ MasterCard

Credit Card No. ____________________ Expires ___/___

Signature ____________________

THE MATHEMATICAL ASSOCIATION OF AMERICA


Laboratory Experiences in Group Theory

A Manual to be used with Exploring Small Groups

Ellen Maycock Parker

The Mathematical Association of America

Laboratory Experiences in Group Theory

A Manual to be Used with *Exploring Small Groups*

Ellen Maycock Parker

Series: Classroom Resource Materials

A lab manual with software for introductory courses in group theory or abstract algebra

Laboratory Experiences in Group Theory is a workbook of 15 laboratories designed to be used with the software *Exploring Small Groups* as a supplement to the regular textbook in an introductory course in group theory or abstract algebra. Written in a step-by-step manner, the laboratories encourage students to discover the basic concepts of group theory and to make conjectures from examples that are easily generated by the software. The labs can be assigned as homework or can be used in a structured laboratory setting. Since the software is user-friendly and the laboratories are complete, students and faculty should have no difficulty in using the labs without training.

Most students find that the laboratories provide an enjoyable alternative to the "theorem-proof-example" format of a standard abstract algebra course. At the end of the semester, one student wrote in his evaluation of the course:

I am truly grateful for the laboratory component...Work on the computer helped to make the abstract theory more concrete... One of the best things about the labs was that we formed our own conjectures about the patterns we saw...I believe that the progression of (1) lab, (2) conjecture, (3) class discussion, and (4) proof was highly beneficial in gaining understanding of the abstract material of the course.

Table of Contents: 1. Groups and Geometry; 2. Cayley Tables; 3. Cyclic Groups and Cyclic Subgroups; 4. Subgroups and Subgroup Lattices; 5. The Center and Commutator Subgroups; 6. Quotient Groups; 7. Direct Products; 8. The Unitary Groups; 9. Composition Series; 10. Introduction to Endomorphisms; 11. The Inner Automorphisms of a Group; 12. The Kernel of an Endomorphism; 13. The Class Equation; 14. Conjugate Subgroups; 15. The Sylow Theorems; Appendix A. Table Generation Menu of *Exploring Small Groups* (*ESG*); Appendix B. Sample Library of *ESG*; Appendix C. Group Library of *ESG*; Appendix D. Group Properties Menu

Exploring Small Groups, the software packaged with this lab manual, is on a $3\frac{1}{2}$" DD PC compatible disk. This is a DOS program that can be run in Windows. The software was developed by Ladnor Geissinger, University of North Carolina at Chapel Hill.

Catalog Code: LABE/JR
112 pp., Paperbound, 1996, ISBN 0-88385-705-7
List: $26.50 MAA Member: $21.00

ORDER FROM:
THE MATHEMATICAL ASSOCIATION OF AMERICA
PO Box 91112, Washington, DC 20090-1112
1-800-331-1622 (301) 617-7800 FAX (301) 206-9789

Membership Code:

____ ____ ____ ____ ____ ____

Name ________________________

Address ________________________

City ________________________

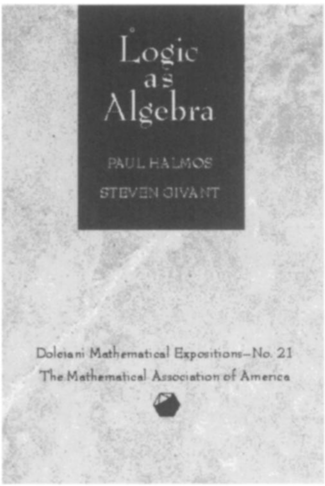
State __________ Zip __________

QTY.	CATALOG CODE	PRICE	AMOUNT
____	LABE/JR	____	____
____	____	____	____
		TOTAL	____

Payment □ Check □ VISA □ MasterCard

Credit Card No. ________________ Expires ___/___

Signature ________________________



Logic as Algebra

PAUL HALMOS

STEVEN GIVANT

Dolciani Mathematical Expositions—No. 21

The Mathematical Association of America

THE MATHEMATICAL ASSOCIATION OF AMERICA

Logic as Algebra

Paul Halmos and Steven Givant

Series: Dolciani Mathematical Expositions

This book is based on the notes of a course in logic given by Paul Halmos. This book retains the spirit and purpose of those notes, which was to show that logic can (and perhaps should) be viewed from an algebraic perspective. When so viewed, many of its principal notions are seen to be old friends, familiar algebraic notions that were "disguised" in logical clothing. Moreover, the connection between the principal theorems of the subject and well-known theorems in algebra becomes clearer. Even the proofs often gain in simplicity.

Propositional logic and monadic predicate calculus—predicate logic with a single quantifier— are the principal topics treated. The connections between logic and algebra are carefully explained. The key notions and the fundamental theorems are elucidated from both a logical and algebraic perspective. The final section gives a unique and illuminating algebraic treatment of the theory of syllogisms—perhaps the oldest branch of logic, and a subject that is neglected in most modern logic texts.

The presentation is aimed at a broad audience—mathematics amateurs, students, teachers, philosophers, linguists, computer scientists, engineers, and professional mathematicians. Whether the reader's goal is a quick glimpse of modern logic or a more serious study of the subject, the book's fresh approach will bring novel and illuminating insights to beginners and professionals alike. All that is required of the reader is an acquaintance with some of the basic notions encountered in a first course in modern algebra. In particular, no prior knowledge of logic is assumed. The book could serve equally well as a fireside companion and as a course text.

Contents: **What is Logic?:** To count or to think; A small alphabet; A small grammar; A small logic; What is truth?; Motivation of the small language; All mathematics. **Propositional Calculus:** Propositional symbols; Propositional abbreviations; Polish notation; Language as an algebra; Concatenation; Theorem schemata; Formal proofs; Entailment; Logical equivalence; Conjunction; Algebraic identities. **Boolean Algebra:** Equivalence classes; Interpretations; Consistency and Boolean algebra; Duality and commutativity; Properties of Boolean algebras; Subtraction; Examples of Boolean algebras. **Boolean Universal Algebra:** Subalgebras; Homomorphisms; Examples of homomorphisms; Free algebras; Kernels and ideals; Maximal ideals; Homomorphism theorem; Consequences; The representation theorem. **Logic via Algebra:** Pre-Boolean algebras; Substitution rule; Boolean logics; Algebra of the propositional calculus; Algebra of proof and consequence. **Lattices and Infinite Operations:** Lattices; Non-distributive lattices; Infinite operations. **Monadic Predicate Calculus:** Propositional functions; Finite functions; Functional monadic algebras; Functional quantifiers; Properties of quantifiers; Monadic algebras; Free monadic algebras; Modal logics; Monadic logics; Syllogisms.

Catalog Code: DOL-21/JR98
152 pp., Paperbound, 1998, ISBN 0-88385-327-2
List: $27.00   MAA Member: $21.95

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm   FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

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

Name ______________________________

Address ______________________________

City ______________ State _____ Zip ________

Phone ______________________________

QTY.	CATALOG CODE	PRICE	AMOUNT
_______	DOL-21/JR98	_______	_______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Shipping & handling ________

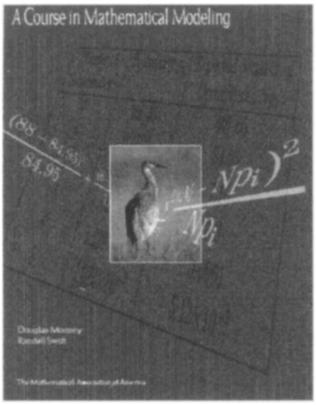
TOTAL ________

Payment □ Check □ VISA □ MasterCard

Credit Card No. ________________ Expires ___/___

Signature ______________________________

THE MATHEMATICAL ASSOCIATION OF AMERICA


A Course in Mathematical Modeling
Douglas Mooney
Randall Swift
The Mathematical Association of America

A Course in Mathematical Modeling

Series: Classroom Resource Materials

Douglas Mooney and Randall Swift

Textbook

This book is intended as a text for a modeling course accessible to students who have mastered a one year course in calculus. It balances a variety of opposing modeling methodologies including theoretical models versus empirical models, analytical models versus simulation, deterministic models versus stochastic models, and discrete models versus continuous models. Most of the examples are drawn from real-world data or from models that have been used in various applied fields. The use of computers in both simulation and in mathematical analysis is an integral part of the presentation.

The authors emphasize the teaching of the modeling process as opposed to merely presenting models. They begin their book with the simple discrete exponential growth model, and successively refine it to include variable growth rates, multiple variables, growth rates fitted to data, and the effects of random factors. The last part of the book moves into continuous-time models. Issues of model validity and purpose are emphasized throughout.

Students taking a course based on this book should have some mathematical maturity, but will need little advanced knowledge. The book presents more advanced topics on an as-needed basis and serves to show how the different topics of undergraduate mathematics can be used together to solve problems. This perspective is valuable as either a road map for beginning students or as a capstone for more advanced students. The course presents elements of discrete dynamical systems, basic probability theory, differential equations, matrix algebra, stochastic processes, curve fitting, statistical testing, and regression analysis. Computer analysis is extensively used in conjunction with these topics.

You can also use this book if you are seeking applications to supplement a course in linear algebra, differential equations, difference equations, probability theory, or statistics.

Catalog Code: MML/JR
400 pp., Paperbound, 1999
ISBN 0-88385-712-X
List: $41.95 MAA Member: $32.95

Visit <www.wku.edu/~swiftrj/Modeling/ modeling.html> where you can visit the authors' website and download data sets, *Mathematica* files, and other modeling resources that execute the models described in the text.

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

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

Name ______________________________

Address ______________________________

City ________________ State _____ Zip ________

Phone ______________________________

QTY.	CATALOG CODE	PRICE	AMOUNT
______	MML/JR	______	______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Shipping & handling ______

TOTAL ______

Payment □ Check □ VISA □ MasterCard

Credit Card No. ____________________ Expires ___/___

Signature ______________________________

THE MATHEMATICAL ASSOCIATION OF AMERICA

New from William Dunham, award-winning author of *Journey through Genius: The Great Theorems of Mathematics,* and *The Mathematical Universe....*

Euler

The Master of Us All

William Dunham

Series: Dolciani Mathematical Expositions

Without question, Leonhard Euler (1707-1783) ranks among history's greatest mathematicians. Across six decades of unmatched productivity, and despite a visual impairment that grew ever worse, he charted the course of mathematics throughout the eighteenth century and beyond. His reputation is captured in Laplace's famous admonition, "Read Euler, read Euler. He is the master of us all."

Written for the mathematically literate reader, this book provides a glimpse of Euler in action. Following an introductory biographical sketch are chapters describing his contributions to eight different topics—number theory, logarithms, infinite series, analytic number theory, complex variables, algebra, geometry, and combinatorics. Each chapter begins with a prologue to establish the historical context and then proceeds to a detailed consideration of one or more Eulerian theorems on the subject at hand. Each chapter concludes with an epilogue surveying subsequent developments or addressing related questions that remain unanswered to this day. At the end of the book is a brief outline of Euler's collected works, the monumental *Opera Omnia*, whose publication has consumed virtually all of the twentieth century.

In all, the book contains three dozen proofs from this remarkable individual. Yet this is merely the tip of the scholarly iceberg, for Euler produced over 30,000 pages of pure and applied mathematics during his lifetime. *Euler: The Master of Us All* samples the work of a mathematician whose influence, industry, and ingenuity are of the very highest order.

Catalog Code: DOL-22/JR
192 pp., Paperbound, ISBN- 0-88385-328-0
List: $29.95 MAA Member: $23.95

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

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

Name ______________________________

Address ______________________________

City ____________________ State _____ Zip ________

Phone ______________________________

Qty.	Catalog Code	Price	Amount
_______	DOL-22/JR	_______	_______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Shipping & handling ________

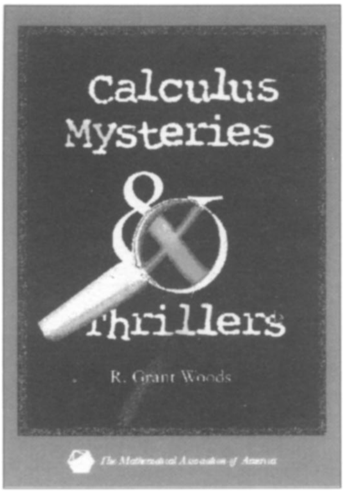
TOTAL ________

Payment □ Check □ VISA □ MasterCard

Credit Card No. ______________________ Expires ___/___

Signature ______________________________

THE MATHEMATICAL ASSOCIATION OF AMERICA



Calculus Mysteries and Thrillers

R. Grant Woods

Series: Classroom Resource Materials

Text
Supplement

Calculus projects you can give to your students . . .

This book consists of eleven mathematics projects based on introductory single-variable calculus, together with some guidance on how to make use of them. Each project is presented as an amusing short story. In many of them a group of undergraduate mathematics students, formed into a consulting company called *Math Iz Us*, is hired to solve mathematical problems brought to them by clients. The problems solved include: helping to prosecute an accused pool shark, defending a driver accused of speeding, assisting a hockey coach in making his star forward a more effective goal scorer, and advising a pirate captain on how to divide a gold-plated goose-egg fairly among his crew.

In each problem, the problem solvers are required to present to their client a detailed written report of their findings. Thus, students must produce and analyze accurate mathematical models of complex, verbally presented "real life" situations, and write a clear technical account of their solution.

Instructors who are looking for problems that are novel, interesting, and several levels more complex than the typical text book "word problem" will find them in this book. It will be of particular value to instructors who wish to combine training in applications of calculus with training in technical writing. The complexity of the problems makes them suitable for use as group projects.

The calculus concepts on which the problems are based include: tangent and normal lines, optimization by use of critical points, inverse trig functions, volumes of solids, surface area integrals, and modeling economic concepts using definite integrals. Although a few ideas from physics and economics are used in the problems, no prior knowledge of these fields is required.

Catalog Code: CTM/JR
144 pp., Paperbound, 1998, ISBN 0-88385-711-1
List: $24.95 MAA Member: $19.95

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

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

Name ____________________

Address ____________________

City ____________ State _____ Zip ________

Phone ____________________

QTY.	CATALOG CODE	PRICE	AMOUNT
______	CTM/JR	______	______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Shipping & handling ________

TOTAL ________

Payment □ Check □ VISA □ MasterCard

Credit Card No. ____________________ Expires ___/___

Signature ____________________

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

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