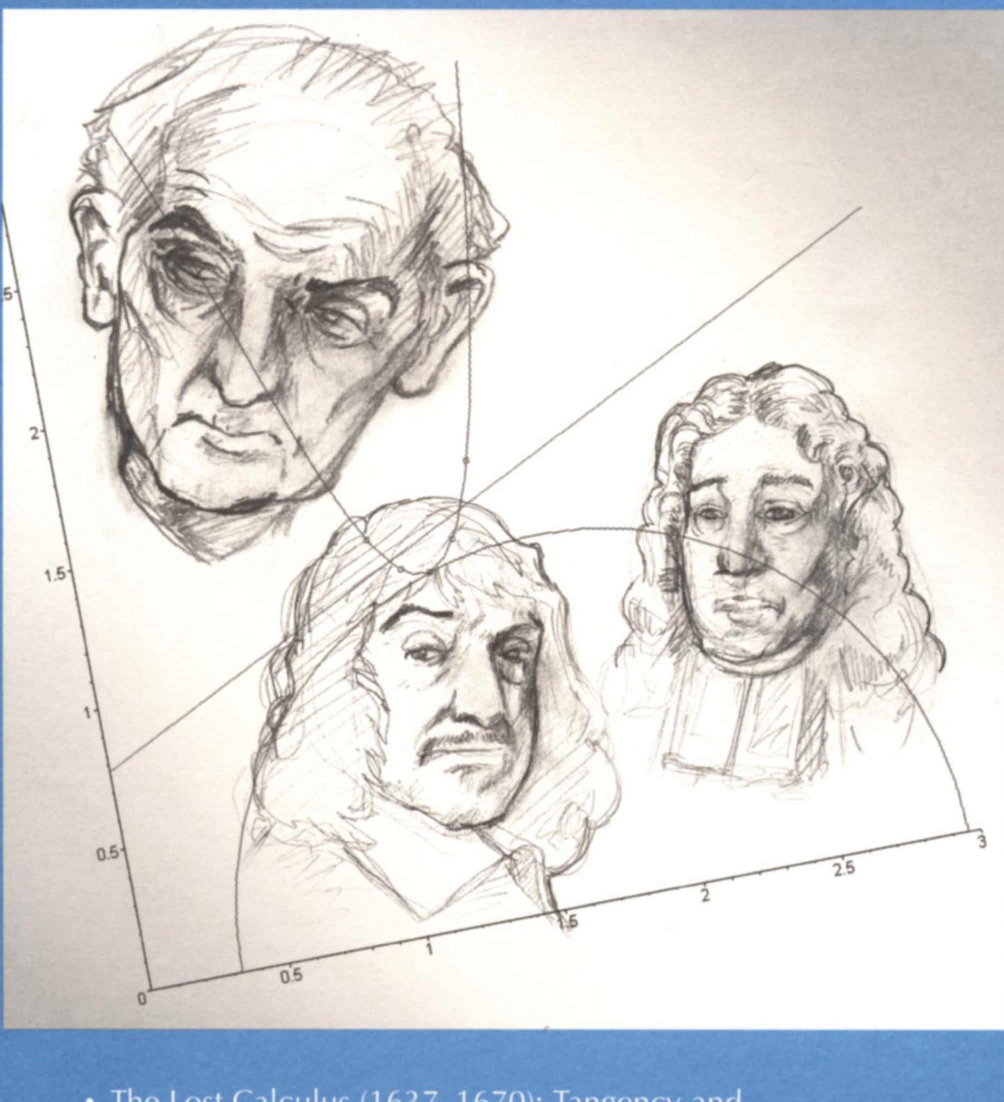
Vol. 78, No. 5, December 2005

MATHEMATICS MAGAZINE



- The Lost Calculus (1637–1670): Tangency and Optimization without Limits
- Basketball, Beta, and Bayes
- The Least-Squares Property of the Lanczos Derivative

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. 74, pp. 75–76, and is available from the Editor or at www.maa.org/pubs/mathmag.html. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, or published by another journal or publisher.

Submit new manuscripts to Allen Schwenk, Editor-Elect, Mathematics Magazine, Department of Mathematics, Western Michigan University, Kalamazoo, MI, 49008. Manuscripts should be laser printed, with wide line spacing, and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should mail three copies and keep one copy. In addition, authors should supply the full five-symbol 2000 Mathematics Subject Classification number, as described in *Mathematical Reviews*.

Cover image: *Off on a Tangent*, by Liam Boylan, based on a concept by Keith Thompson. Two articles in this issue address alternate ways to treat tangency. Featured on the cover are Descartes, Hudde, and Lanczos, along with representations of their contributions to this fruitful field of inquiry.

Liam Boylan graduated from Santa Clara University with a double major in Communications and Studio Art. Between careers at the moment, Liam seeks opportunities to bolster his portfolio. His parents have instructed him to marry rich.

AUTHORS

Jeff Suzuki became a mathematician because his results in laboratory science frequently contradicted the known laws of physics and chemistry. In graduate school he combined his interests in mathematics, history, and physics in a dissertation on the history of the stability problem in celestial mechanics, and has subsequently focused on the mathematics of the 17th and 18th centuries because, in his words, it was "the last time it was possible to know everything." Current research interests include pre-Newtonian methods of solving calculus type problems (hence this article), the early history of analytic number theory, and Renaissance cookery.

Matt Richey has been at St. Olaf College since 1986. He was an undergraduate at Kenyon College and received his Ph.D. in mathematical physics at Dartmouth College in 1985 under the direction of Craig Tracy. His primary interest is computational mathematics, especially randomized algorithms. When not attending to academic pursuits, he is working to convince the baseball strategists of the futility of the sacrifice bunt. His next paper is tentatively entitled "Baseball, Bunting, and Beane (Billy)".

Paul Zorn was born and raised in southern India. He was an undergraduate at Washington University in St. Louis, and earned his Ph.D. in complex analysis at the University of Washington, Seattle, supervised by Professor E. L. Stout. He has been on the faculty at St. Olaf College since 1981. His professional interests include complex analysis, mathematical exposition, and textbook writing. In an earlier millennium he served a term as editor of *Mathematics Magazine*.

Nathanial Burch is a senior at Grand Valley State University, where he is double majoring in mathematics and statistics. The past summers have found Nathanial working on two different funded undergraduate research projects, one at Grand Valley State University and the other at the Worcester Polytechnic Institute REU program. He is a member of the Pi Mu Epsilon and Phi Kappa Phi honor societies. Upon completion of his senior year, Nathanial is planning on attending graduate school in applied mathematics. Interests beyond mathematics include his long-time passions for bowling, fishing, and billiards, along with a newly developed love for traveling.

Paul Fishback received his undergraduate degree from Hamilton College and his Ph.D. from the University of Wisconsin-Madison. Ever since beginning his college teaching career, he strongly desired to investigate the notion of least-squares differentiation, which his college classmate Dan Kopel had researched as an undergraduate. It is by sheer coincidence that he discovered that Dan's work was linked to the Lanczos derivative. In addition to having a broad array of analysis-related interests, he serves as a councillor for Pi Mu Epsilon and as associate editor of the MAA Notes series. His main nonmathematical interest is traditional, Greenland-style kayaking, and he has paddled extensively on Lake Michigan and Lake Superior.

Vol. 78, No. 5, December 2005

MATHEMATICS MAGAZINE

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 *MATHEMATICS MAGAZINE* to an individual member of the Association is $131. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 20% dues discount for the first two years of membership.)

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 Frank Peterson (*FPetersonj@aol.com*), Advertising Manager, the Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

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

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

Printed in the United States of America

ARTICLES

The Lost Calculus (1637–1670): Tangency and Optimization without Limits

JEFF SUZUKI
Brooklyn College
Brooklyn, NY 11210
jeff_suzuki@yahoo.com

If we wished to find the tangent to a given curve or the extremum of a function, we would almost certainly rely on the techniques of a calculus based on the theory of limits, and might even conclude that this is the only way to solve these problems (barring a few special cases, such as the tangent to a circle or the extremum of a parabola). It may come as a surprise, then, to discover that in the years between 1637 and 1670, very general algorithms were developed that could solve virtually every "calculus type" problem concerning algebraic functions. These algorithms were based on the theory of equations and the geometric properties of curves and, given time, might have evolved into a calculus entirely free of the limit concept. However, the work of Newton and Leibniz in the 1670s relegated these techniques to the role of misunderstood historical curiosities.

The foundations of this "lost calculus" were set down by Descartes, but the keys to unlocking its potential can be found in two algorithms developed by the Dutch mathematician Jan Hudde in the years 1657–1658. In modernized form, Hudde's results may be stated as follows: Given any polynomial

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0,$$

1. if $f(x)$ has a root of multiplicity 2 or more at $x = a$, then the polynomial that we know as f' has $f'(a) = 0$, and
2. if $f(x)$ has an extreme value at $x = a$, then $f'(a) = 0$.

These results can, of course, be easily derived through differentiation, so it is tempting to view them as results that point "clearly toward algorithms of the calculus" [**2**, p. 375]. But Hudde obtained them from purely algebraic and geometric considerations that at no point rely on the limit concept. Rather than being heralds of the calculus that was to come, Hudde's results are instead the ultimate expressions of a purely algebraic and geometric approach to solving the tangent and optimization problems.

We examine the evolution of the lost calculus from its beginnings in the work of Descartes and its subsequent development by Hudde, and end with the intriguing possibility that nearly every problem of calculus, including the problems of tangents, optimization, curvature, and quadrature, could have been solved using algorithms entirely free from the limit concept.

Descartes's method of tangents

The road to a limit-free calculus began with Descartes. In *La Géométrie* (1637), Descartes described a method for finding tangents to algebraic curves. Conceptually, Descartes's method is the following: Suppose we wish to find a circle that is tangent to the curve OC at some point C (see FIGURE 1). Consider a circle with center P on some convenient reference axis (we may think of this as the x-axis, though in practice any clearly defined line will work), and suppose this circle passes through C. This circle may pass through another nearby point E on the curve; in this case, the circle is, of course, not tangent to the curve. On the other hand if C is the only point of contact between the circle and curve, then the circle will be tangent to the curve. Thus our goal is simple: Find P so that the circle with center P and radius CP will meet the curve OC only at the point C.

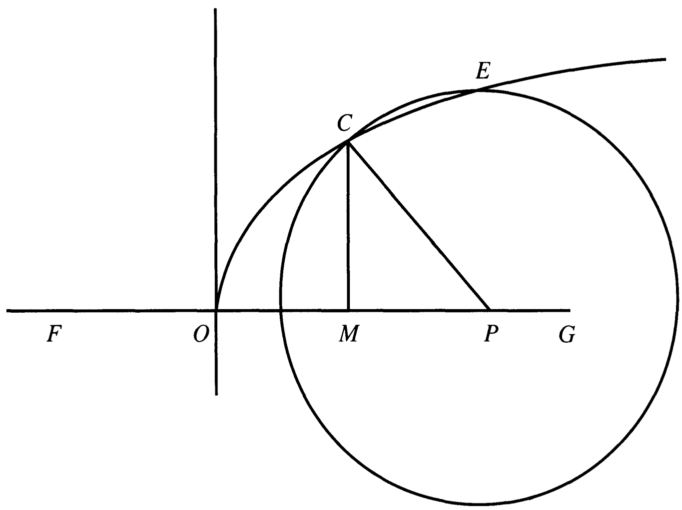


Figure 1 Descartes's method of tangents

Algebraically, any points the circle and curve have in common correspond to a solution to the system of equations representing the curve and circle. If there are two distinct intersections, this system will have two distinct solutions; thus, in order for the circle and curve to be tangent and have just a single point in common, the system of equations must have two equal solutions. In short, the system of equations must have a double root corresponding to the common point C.

Descartes, who wanted his readers to become proficient with his method through practice, never deigned to give simple examples. However, we will present a simple example of Descartes's method in modern form. Suppose OC is the curve $y = \sqrt{x}$, and let C be the point (a^2, a). Imagine a circle passing through the point (a^2, a), with radius r centered on the x-axis at the point $(h, 0)$, with h and r to be determined. Then the circle has equation

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

Expanding and setting the equation equal to zero gives

$$y^2 + x^2 - 2hx + h^2 - r^2 = 0.$$

The points of intersection of the circle and curve correspond to the solutions to the system of equations:

$$y^2 + x^2 - 2hx + h^2 - r^2 = 0 \quad \text{and} \quad y = \sqrt{x}.$$

If we eliminate y using the substitution $y = \sqrt{x}$, we obtain

$$x^2 + (1 - 2h)x + (h^2 - r^2) = 0,$$

which is a quadratic and will in general have two solutions for x. By assumption, the circle and curve have the point (a^2, a) in common; hence $x = a^2$ is a root of this equation. In order for the circle and curve to be tangent, we want $x = a^2$ to be the only root. Thus it is necessary that

$$x^2 + (1 - 2h)x + (h^2 - r^2) = (x - a^2)^2.$$

Expanding the right-hand side and comparing coefficients we find that

$$1 - 2h = -2a^2$$

and thus $h = a^2 + 1/2$. Therefore the circle with center $(a^2 + 1/2, 0)$ will be tangent to the graph of $y = \sqrt{x}$ at the point (a^2, a).

This method of Descartes approached the problem of tangents by locating the center of the tangent circle. Today, we solve the problem by finding the slope of the tangent line. Fortunately there is a simple relationship between the two. From Euclidean geometry, we know that the radius through a point C on a circle will be perpendicular to the tangent line of the circle through C. In this case the radius PC will lie on a line with a slope $-2a$; hence the tangent line through C will have slope $1/(2a)$. This is, of course, what we would obtain using the derivative, but here we used only the algebraic properties of equations and the geometrical properties of curves.

The method in *La Géométrie* is elegant, and works very well for all quadratic forms. Unfortunately, it rapidly becomes unwieldy for all but the simplest curves. For example, suppose we wish to find the tangent to the curve $y = x^3$. As before, let the center of our circle be at $(h, 0)$; we want the system

$$x^2 - 2hx + h^2 + y^2 - r^2 = 0 \quad \text{and} \quad y = x^3$$

to have a double root at the point of tangency (a, a^3). Substituting x^3 for y gives

$$x^6 + x^2 - 2hx + h^2 - r^2 = 0. \tag{1}$$

If we wish to find the tangent at the point (a, a^3), this equation should have a double root at $x = a$; since the left-hand side is a 6th degree monic polynomial, it must factor as the product of $(x - a)^2$ and a fourth degree monic polynomial:

$$x^6 + x^2 - 2hx + h^2 - r^2 = (x - a)^2(x^4 + Bx^3 + Cx^2 + Dx + F).$$

Expanding the right-hand side gives us

$$\begin{aligned} &x^6 + x^2 - 2hx + h^2 - r^2 \\ &\qquad = x^6 + (B - 2a)x^5 + (a^2 - 2aB + C)x^4 + (a^2B - 2aC + D)x^3 \\ &\qquad\quad + (a^2C - 2aD + F)x^2 + (a^2D - 2aF)x + a^2F. \end{aligned}$$

Comparing coefficients gives us the system

$$B - 2a = 0 \tag{2}$$

$$a^2 - 2aB + C = 0 \tag{3}$$

$$a^2B - 2aC + D = 0 \tag{4}$$

$$a^2C - 2aD + F = 1 \tag{5}$$

$$a^2D - 2aF = -2h \tag{6}$$

$$a^2F = h^2 - r^2. \tag{7}$$

From Equation 2 we have $B = 2a$. Substituting this into (3) we have

$$a^2 - 2a(2a) + C = 0;$$

hence $C = 3a^2$. Substituting the values for B and C into (4) gives us

$$a^2(2a) - 2a(3a^2) + D = 0;$$

hence $D = 4a^3$. Substituting these values into (5) gives

$$a^2(3a^2) - 2a(4a^3) + F = 1;$$

hence $F = 1 + 5a^4$. Substituting into (6) gives

$$a^2(4a^3) - 2a(1 + 5a^4) = -2h;$$

hence $h = a + 3a^5$ and the center of the tangent circle will be at $(a + 3a^5, 0)$. As before, the perpendicular to the curve will have slope $-a^3/(3a^5) = -1/(3a^2)$, and thus the slope of the line tangent to the curve $y = x^3$ at $x = a$ will be $3a^2$.

It is clear from this example that the real difficulty in applying Descartes's method is this: If $y = f(x)$, where $f(x)$ is an nth degree polynomial, then finding the tangent to the curve at the point where $x = a$ requires us to find the coefficients of $(x - a)^2$ multiplied by an arbitrary polynomial of degree $2n - 2$. The problem is not so much difficult as it is tedious, and any means of simplifying it would significantly improve its utility.

Descartes discovered one simplification shortly after the publication of *La Géométrie*. He described his modified method in a 1638 letter to Claude Hardy [**5**, vol. VII, p. 61ff]. Descartes's second method of tangents still relies on the system of equations having a double root corresponding to the point of tangency, but Descartes simplified the procedure by replacing the circle with a line and used the slope idea implicitly (as the ratio between the sides of similar triangles).

In modern terms, we describe Descartes's second method as follows: The equation of a line that touches the curve $f(x, y) = 0$ at (a, b) is $y = m(x - a) + b$, where m denotes a parameter to be determined. In order for the line to be tangent, the system of equations

$$f(x, y) = 0 \quad \text{and} \quad y = m(x - a) + b$$

must have a double root at $x = a$ (alternatively a double root at $y = b$).

For example, if we wish to find the tangent to $y = x^3$ at (a, a^3), the system of equations

$$y = x^3 \quad \text{and} \quad y = m(x - a) + a^3$$

can be reduced to

$$x^3 - mx + (ma - a^3) = 0$$

by substituting the first expression for y into the second equation. In order for the line to be tangent at $x = a$, it is necessary that $x = a$ be a double root, so $(x - a)^2$ is a factor of this polynomial; if we call the other factor $(x - r)$, we can write

$$\begin{aligned} x^3 - mx + (ma - a^3) &= (x-a)^2(x-r) \\ &= x^3 - (r+2a)x^2 + (a^2+2ar)x - a^2 r. \end{aligned}$$

Comparing coefficients gives us the system

$$r + 2a = 0, \quad a^2 + 2ar = -m, \quad \text{and} \quad -a^2 r = ma - a^3.$$

Solving this system gives us $m = 3a^2$. This is, of course, the same answer we would obtain by differentiating $y = x^3$, but obtained entirely without the use of limits.

Either of the two methods of Descartes will serve to find the tangent to any algebraic curve, even curves defined implicitly (since, as Descartes pointed out, an expression for y can be found from the equation of the circle or the tangent line and substituted into the equation of the curve). For example, during a dispute with Fermat over their respective method of tangents, Descartes challenged Fermat and his followers to find the tangent to a curve now known as the folium of Descartes [**5**, vol. VII, p. 11], a curve whose equation we would write as $x^3 + y^3 = pxy$.

The reader may be interested in applying Descartes's method to the folium. To find the line tangent to the folium at the point (x_0, y_0), we want the system

$$x^3 + y^3 = pxy \quad \text{and} \quad y = m(x - x_0) + y_0$$

to have a double root $x = x_0$. It should be pointed out that, contrary to Descartes's expectations, Fermat's method *could* be applied to the folium; Descartes subsequently challenged Fermat to find the point on the folium where the tangent makes a 45-degree angle with the axis (and again Fermat responded successfully).

Hudde's first letter: polynomial operations

The key to Descartes's methods is knowing when the system of equations that determine the intersection(s) of the two curves (whether a line and the curve, or a circle and the curve) has a double root, which corresponds to a point of tangency. An efficient algorithm for detecting double roots of polynomials would vastly enhance the usability of Descartes's method. Such a method was discovered by the Dutch mathematician Jan Hudde.

Hudde studied law at the University of Leiden, but while there he joined a group of Dutch mathematicians gathered by Franciscus van Schooten. At the time van Schooten, who had already published one translation of Descartes's *La Géométrie* from French into Latin, was preparing a second, more extensive edition. This edition, published in two volumes in 1659 and 1661, included not only a translation of *La Géométrie*, but explanations, elaborations, and extensions of Descartes's work by the members of the Leiden group, including van Schooten, Florimond de Beaune, Jan de Witt, Henrik van Heuraet, and Hudde.

Hudde (along with Jan de Witt) would soon leave mathematics for politics, and eventually became a high official of the city of Amsterdam. When Louis XIV invaded The Netherlands in 1672, Hudde helped direct Dutch defenses [**2, 6**]; for this, Hudde became a national hero. De Witt was less fortunate: his actions were deemed partially responsible for the ineptitude and unpreparedness of the Dutch army in the early stage of the war, and he and his brother were killed by a mob on August 20, 1672.

Hudde's return to politics may have saved The Netherlands, but mathematics lost one of its rising stars. Leibniz in particular was impressed with Hudde's work, and when Johann Bernoulli proposed the brachistochrone problem, Leibniz lamented:

> If Huygens lived and was healthy, the man would rest, except to solve your problem. Now there is no one to expect a quick solution from, except for the Marquis de l'Hôpital, your brother [Jacob Bernoulli], and Newton, and to this list we might add Hudde, the Mayor of Amsterdam, except that some time ago he put aside these pursuits [**9**, vol. II, p. 370].

As Leibniz's forecast was correct regarding the other three, one wonders what would have happened had Hudde not put aside mathematics for politics.

Hudde's work in the 1659 edition of Descartes consists of two letters. The first, "On the Reduction of Equations," was addressed to van Schooten and dated the "Ides of July, 1657" (July 15, 1657). In the usage of the time, to "reduce" an equation meant to factor it, usually as the first step in finding all its roots. Thus the letter begins with a sequence of rules (what we would call algorithms) that can be used to find potential factors of a polynomial. These factors have one of two types: those corresponding to a root of multiplicity 1, or those corresponding to a root of multiplicity greater than 1. Since Descartes's method of tangents (and Hudde's method of finding extreme values) relies on finding multiple roots, this has particular importance.

Key to Hudde's method of finding roots of multiplicity greater than 1 is the ability to find the greatest common divisor (GCD) of two polynomials. How can this be done? One way is to factor the two polynomials and see what factors they have in common. However this is impractical for any but the most trivial polynomials (and in any case requires knowing the roots we are attempting to find). A better way is to use the Euclidean algorithm for polynomials. For example, suppose we wish to find the GCD of $f(x) = x^3 - 4x^2 + 10x - 7$ and $g(x) = x^2 - 2x + 1$. To apply the Euclidean algorithm we would divide $f(x)$ by $g(x)$ to obtain a quotient (in this case, $x - 2$) and a remainder (in this case, $5x - 5$); we can express this division as

$$x^3 - 4x^2 + 10x - 7 = (x^2 - 2x + 1)(x - 2) + (5x - 5).$$

Next, we divide the old divisor, $x^2 - 2x + 1$, by the remainder $5x - 5$, to obtain a new quotient and remainder:

$$x^2 - 2x + 1 = (5x - 5)\left(\frac{1}{5}x - \frac{1}{5}\right) + 0.$$

The last nonzero remainder (in this case, $5x - 5$) corresponds to the GCD; in general, it will be a constant multiple of it.

While this is the way the Euclidean algorithm for polynomials is generally taught today, Hudde presented a clever variation worth examining. Since we are only interested in the remainder when the polynomials are divided, we can, instead of performing the division, find the remainder modulo the divisor. In particular, Hudde's steps treated the divisor as being "equal to nothing"; he set the two polynomials equal to zero and solved for the highest power term in each. In our example, we would have

$$x^2 = 2x - 1$$
$$x^3 = 4x^2 - 10x + 7.$$

The first gives us an expression for x^2 that can be used to eliminate the x^2 and higher degree terms of the other factor. Substituting and solving for the highest power remaining, we write the following sequence of steps:

$$\begin{aligned} x^3 &= 4x^2 - 10x + 7 \\ x(x^2) &= 4(2x - 1) - 10x + 7 \\ x(2x - 1) &= 4(2x - 1) - 10x + 7 \\ 2x^2 - x &= 8x - 4 - 10x + 7 \\ 2(2x - 1) - x &= -2x + 3 \\ 3x - 2 &= -2x + 3 \\ 5x &= 5 \\ x &= 1. \end{aligned}$$

Thus the second equation, $x^3 = 4x^2 - 10x + 7$, has been reduced to $x = 1$; we note that our result is equivalent to showing

$$x^3 - 4x^2 + 10x - 7 = x - 1 \bmod x^2 - 2x + 1.$$

Next, we can use $x = 1$ to eliminate all terms of the first or higher degree terms of the other equation:

$$\begin{aligned} x^2 &= 2x - 1 \\ 1^2 &= 2(1) - 1 \\ 1 &= 1 \end{aligned}$$

Since this is an identity, then the GCD is the factor corresponding to the last substitution: here, $x = 1$ corresponds to the factor $x - 1$.

The value of finding the GCD is made apparent in Hudde's tenth rule, which concerns reducing equations with two or more equal roots:

> If the proposed equation has two equal roots, multiply by whatever arithmetic progression you wish: that is, [multiply] the first term of the equation by the first term of the progression; the second term of the equation by the second term of the progression, and so forth; and set the product which results equal to 0. Then with the two equations you have found, use the previously explained method to find their greatest common divisor; and divide the original equation by the quantity as many times as possible [**4**, p. 433–4].

Hudde implies but does not state that the GCD will contain all the repeated factors; this is the first appearance of what we might call Hudde's Theorem:

HUDDE'S THEOREM. *Let $f(x) = \sum_{k=0}^{n} a_k x^k$, and let $\{b_k\}_{k=0}^{n}$ be any arithmetic progression. If $x = r$ is a root of $f(x)$ with multiplicity 2 or greater, then $x = r$ will be a root of $g(x) = \sum_{k=0}^{n} b_k a_k x^k$.*

We will refer to the polynomial $g(x)$ derived in this way from $f(x)$ as a *Hudde polynomial* (note that it is not unique). As an example, Hudde seeks to find the roots of $x^3 - 4x^2 + 5x - 2$. Using the arithmetic sequence 3, 2, 1, 0, we produce the Hudde polynomial $3x^3 - 8x^2 + 5x$ using a tabular array like this:

$$\begin{array}{cccc} x^3 & -\,4x^2 & +\,5x & -\,2 \\ 3 & 2 & 1 & 0 \\ \hline 3x^3 & -\,8x^2 & +\,5x & \end{array}$$

The greatest common divisor of $x^3 - 4x^2 + 5x - 2$ and $3x^3 - 8x^2 + 5x$ is $x - 1$. Hence, if the original has a repeated factor, it can only be $x - 1$. Attempting to factor it out, we find $x^3 - 4x^2 + 5x - 2 = (x-1)^2(x-2)$. Thus the roots are 1, 1, and 2.

Hudde points out that *any* arithmetic sequence will work; indeed, he uses the same polynomial but a different arithmetic sequence:

$$\begin{array}{cccc} x^3 & -\,4x^2 & +\,5x & -\,2 \\ 1 & 0 & -1 & -2 \\ \hline x^3 & & -\,5x & +\,4 \end{array}$$

As before, the GCD of $x^3 - 5x + 4$ and $x^3 - 4x^2 + 5x - 2$ is $x - 1$.

Hudde further notes that the procedure can be repeated if there is a triple root, repeated twice if there is a quadruple root, and so on; in modern terms, we would say that if $f(x)$ has a root $x = a$ of multiplicity n, then a Hudde polynomial generated from $f(x)$ will have a root $x = a$ of multiplicity $n - 1$.

The value of the method is obvious: If the original polynomial has missing terms, the arithmetic sequence can be constructed to take advantage of these missing terms. In his example for an equation with three or more equal roots, Hudde takes advantage of this ability to choose the arithmetic sequence: Given the equation $x^4 - 6x^2 + 8x - 3 = 0$, Hudde applies his procedure to the polynomial twice, using an arithmetic sequence beginning with 0:

$$\begin{array}{ccccc} x^4 & *\,- & 6x^2 & +\,8x & -\,3 \\ 0 & 1 & 2 & 3 & 4 \\ \hline & & -\,12x^2 & +\,24x & -\,12 \\ & & 0 & 1 & 2 \\ \hline & & & 24x & -\,24 \end{array}$$

(we follow Hudde and Descartes in the use of "$*$" to represent a missing term).

The GCD of the last polynomial, $24x - 24$ and the original polynomial $x^4 - 6x^2 + 8x - 3$ is $x - 1$; dividing the original by $x - 1$ repeatedly yields a factorization and consequently the roots: 1, 1, 1, and -3.

Hudde points out that this method can be used to solve the Cartesian tangent problem [**4**, p. 436], and solves one of Descartes's problems using his method. Let us apply Hudde's method to our earlier problem of finding the tangent to $y = x^3$. Recall that in this case we wished to find the center $(h, 0)$ of a circle that passed through the point (a, a^3). The corresponding system of equations

$$y = x^3 \quad \text{and} \quad (x-h)^2 + y^2 = r^2$$

could be reduced by substituting x^3 for y in the second equation; this results in

$$x^6 + x^2 - 2hx + h^2 - r^2 = 0.$$

In order for the circle to be tangent to the curve at (a, a^3), this equation must have a double root at $x = a$. Hence the corresponding Hudde polynomial will have a root at $x = a$. We can construct a Hudde polynomial by multiplying through by an arithmetic sequence ending in zero:

$$\begin{array}{cccccccc} x^6 & * & * & * & + & x^2 - 2hx & & h^2 - r^2 \\ 6 & 5 & 4 & 3 & & 2 & 1 & 0 \\ \hline 6x^6 & & & & & + 2x^2 - 2hx & & \end{array}$$

If this Hudde polynomial has a root at $x = a$, then h must satisfy

$$6a^6 + 2a^2 - 2ha = 0.$$

Hence $h = a + 3a^5$, and so the center of the tangent circle will be located at $(a + 3a^5, 0)$.

We can also apply Hudde's method to Descartes's second method of tangents (Hudde himself seemed unaware of this improved algorithm). By finding an expression for the slope m of the tangent line to a curve at a point, Hudde's methods would give us a tool equivalent to the derivative. In the case of $y = x^3$, we would want the system

$$y = x^3 \quad \text{and} \quad y = m(x - a) + a^3$$

to have a double root at $x = a$. Substituting $y = x^3$ into the second equation and setting it equal to zero gives us

$$x^3 - mx + am - a^3 = 0.$$

By assumption, $x = a$ is a double root; hence $x = a$ will be a root of any Hudde polynomial formed from this equation. For simplicity, we will form the Hudde polynomial by multiplying the kth degree term by k:

$$\begin{array}{cccc} x^3 & * & - mx & + (am - a^3) \\ 3 & 2 & 1 & 0 \\ \hline 3x^3 & & - mx & \end{array}$$

If $x = a$ is a root of the Hudde polynomial, then m must satisfy

$$3a^3 - ma = 0.$$

Hence $m = 3a^2$, which is precisely what the derivative of $y = x^3$ would give us. We leave it to the reader to show that all the standard rules for finding tangents to graphs of algebraic functions can be derived using Hudde's method.

Hudde's second letter: extreme values

In his first letter, Hudde also mentioned that his method could be used to find extrema, though details only appear in the second, much shorter letter dated "6 Calends of February 1658" (February 6, 1658). This letter has been translated from the original Latin into Dutch [**7**], though I am not aware of a translation into any other language.

The letter opens with a restatement of Hudde's Theorem, which he then proves. The proof is purely algebraic, rigorous by both contemporary and modern standards: hence, Hudde's methods, all based on Hudde's Theorem, neither make nor require any appeal to limits, infinitesimals, or any other ideas of calculus.

Hudde's proof, slightly modified for purpose of clarity, is the following: Suppose a polynomial $P(x)$ is the product of the third-degree polynomial $x^3 + px^2 + qx + r$ and a second-degree polynomial with a double root $x^2 - 2yx + y^2$ (whence $x = y$ is the double root). Hence the roots of $P(x)$ satisfy

$$\begin{aligned} P(x) = &\left(x^2 - 2yx + y^2\right) x^3 \\ &+ \left(x^2 - 2yx + y^2\right) px^2 \\ &+ \left(x^2 - 2yx + y^2\right) qx \\ &+ \left(x^2 - 2yx + y^2\right) r \quad = 0, \end{aligned}$$

where for convenience we designate the polynomial $\left(x^2 - 2yx + y^2\right)$ as the "coefficients" of the terms of the cubic polynomial.

Note that the x^2, $-2yx$, and y^2 terms of the coefficients correspond to terms of descending degree in $P(x)$; hence when a Hudde polynomial is formed from it, the coefficients will be multiplied by successive terms in the arithmetic sequence a, $a + b$, $a + 2b$, to become

$$ax^2 - (a + b)2yx + (a + 2b)y^2.$$

If $x = y$, this coefficient will be

$$ay^2 - (a + b)2y^2 + (a + 2b)y^2,$$

which is identically zero. Hence all coefficients of the Hudde polynomial will be zero when $x = y$, and thus $x = y$ will be a root of the Hudde polynomial. This proves Hudde's theorem (at least for fifth-degree polynomials; the extension of the proof to polynomials of arbitrary degree should be clear).

Geometrically, the application of Hudde's rule to finding the extreme value of a polynomial function is clear: Suppose $f(x)$ has an extremum at $x = a$, with $f(a) = Z$. Then $f(x) - Z$ will have a double root at $x = a$, and the corresponding Hudde polynomial will have a root of $x = a$.

It would seem that Hudde's method requires knowledge of the extreme value Z in order to find the extreme value. But if $f(x)$ is a polynomial function, the arithmetic sequence can be chosen so the constant term (and thus Z) is multiplied by 0 and eliminated. For example, consider the problem of finding the extreme values of $x^3 - 10x^2 - 7x + 346$. Suppose $x^3 - 10x^2 - 7x + 346 = Z$ is the extreme value, which occurs at $x = a$; then $x^3 - 10x^2 - 7x + 346 - Z$ will have a double root at $x = a$. By Hudde's Theorem, any corresponding Hudde polynomial will have a root at $x = a$. If we multiply by an arithmetic sequence ending in zero, we can eliminate the Z:

$$\begin{array}{ccccc} x^3 & -\,10x^2 & -\,7x & +\,346 - Z \\ 3 & 2 & 1 & 0 \\ \hline 3x^3 & -\,20x^2 & -\,7x & \end{array}$$

(Modern readers will recognize this Hudde polynomial as $x \cdot f'(x)$.) Hudde gives no details, but presumably one would find the location of the extrema by setting the Hudde polynomial $3x^3 - 20x^2 - 7x$ equal to zero (giving an equation we will designate the Hudde equation). By assumption, $x = a$ is a double root of the original polynomial, so by Hudde's theorem, $x = a$ is a solution to $3x^3 - 20x^2 - 7x = 0$. The solutions are $x = 0$, $x = -1/3$, and $x = 7$. By assumption, at least one of the roots would have to correspond to a double root of the original for the appropriate value of Z; as with the corresponding calculus procedure, we must verify which (if any) correspond to an actual extremum. In this case, $x = -1/3$ corresponds to a local maximum, $x = 7$ to a local minimum, and $x = 0$ is extraneous.

Rational functions

Hudde also used his theorem to find extrema of rational functions, and we might compare Hudde's method with our own. Suppose we wish to find the extreme values of

$$f(x) = \frac{x^2 - 2x + 7}{x^2 + 4}.$$

Using calculus, we would find the critical points by solving $f'(x) = 0$; since

$$f'(x) = \frac{(x^2 + 4)(2x - 2) - (x^2 - 2x + 7)(2x)}{(x^2 + 4)^2},$$

then $f'(x) = 0$ if $(x^2 + 4)(2x - 2) - (x^2 - 2x + 7)(2x) = 0$; hence the solutions to this equation are the critical points.

In a like manner, Hudde used his theorem to obtain an equation whose roots correspond to the critical points of the rational function. As before, if $f(x)$ is a rational function with an extreme value Z at $x = a$, then $f(x) = Z$ will have a double root at $x = a$. To find the location of the extreme value, Hudde presents a rather more complex rule (though in fairness, it is not significantly more difficult than the quotient rule for differentiation): First, we are free to drop any constant terms (they can be subsumed into the extreme value Z). Next, break the denominator into individual terms and multiply each term of the denominator by a Hudde polynomial formed from the numerator polynomial using an arithmetic sequence whose terms are the difference between the power of the term in the numerator and the power of the term from the denominator. If the sum is set equal to zero, then a double root of the original rational expression will correspond to a root of this equation.

Using Hudde's method, we would first break the denominator of $f(x)$ into its component terms x^2 and 4. Then each term would be multiplied by a Hudde polynomial made from the numerator using an arithmetic sequence whose terms are the differences between the power of the numerator term and the power of the denominator term. Thus x^2 will be multiplied by $(2-2)\cdot x^2 - (1-2)\cdot 2x + (0-2)\cdot 7 = 2x - 14$, while 4 will be multiplied by $(2-0)\cdot x^2 - (1-0)\cdot 2x + (0-0)\cdot 7 = 2x^2 - 2x$; the Hudde polynomial will then be

$$x^2(2x - 14) + 4(2x^2 - 2x).$$

If $x = a$ corresponds to an extreme value Z, then $x = a$ will be a root of this polynomial. The reader may easily verify that the roots are $x = 0$, $x = 4$, and $x = -1$; a graph shows that $x = 0$ is extraneous, and the relative maximum occurs at $x = -1$ and the relative minimum at $x = 4$.

The method of Hudde for rational functions seems very complex, but Hudde shows how it derives naturally from the previous work. As an example, Hudde sought to find the extreme values of

$$\frac{ba^2x + a^2x^2 - bx^3 - x^4}{ba^2 + x^3} - a + x.$$

First, the constant term $-a$ can be dropped, and the expression can be rewritten as a single quotient:

$$\frac{2ba^2x + a^2x^2 - bx^3}{ba^2 + x^3}.$$

Suppose this has an extremum Z at $x = c$; we may write

$$\frac{2ba^2x + a^2x^2 - bx^3}{ba^2 + x^3} = Z \tag{8}$$

and note that, as before, $x = c$ corresponds to a double root of this equation. Multiply to convert this into a polynomial equation:

$$2ba^2x + a^2x^2 - bx^3 = Zba^2 + Zx^3.$$

In previous problems, we multiplied by an arithmetic sequence ending in zero to eliminate the term corresponding to the (unknown) extreme value Z. But here the extreme value Z appears in both the constant and third-degree terms, so how can we pick an arithmetic sequence that will eliminate it? The answer is remarkably simple: we can eliminate one of the terms including Z as before by choosing our arithmetic sequence appropriately. The remaining Zs can be eliminated using (8).

In this case, we can form a Hudde polynomial by multiplying the kth-degree term by k:

$$1 \cdot 2ba^2x + 2 \cdot a^2x^2 - 3 \cdot bx^3 = 3 \cdot Zx^3.$$

Solving this equation for Z gives us

$$Z = \frac{1 \cdot 2ba^2x + 2 \cdot a^2x^2 - 3 \cdot bx^3}{3 \cdot x^3}. \tag{9}$$

Substituting in the expression for Z from (8) yields

$$\frac{2ba^2x + a^2x^2 - bx^3}{ba^2 + x^3} = \frac{1 \cdot 2ba^2x + 2 \cdot a^2x^2 - 3 \cdot bx^3}{3 \cdot x^3}.$$

Cross-multiplying and collecting all the terms on one side gives us the equation

$$(1 \cdot 2ba^2x + 2 \cdot a^2x^2 - 3 \cdot bx^3)ba^2 + (-2 \cdot 2ba^2x + (-1) \cdot a^2x^2 - 0 \cdot bx^3)x^3 = 0,$$

which must have $x = c$ as a root.

We will prove Hudde's rule for the case of a rational function consisting of the quotient of two quadratics. Suppose

$$\frac{a + bx + cx^2}{d + ex + fx^2} = Z, \tag{10}$$

where Z is a local extremum. Clearing denominators we obtain

$$Z(d + ex + fx^2) - a - bx - cx^2 = 0.$$

One of the many possibilities for the Hudde equation is

$$Z(ex + 2fx^2) - bx - 2cx^2 = 0.$$

Solve this for Z:

$$Z = \frac{bx + 2cx^2}{ex + 2fx^2}.$$

Substituting this value of Z into (10) gives

$$\frac{a + bx + cx^2}{d + ex + fx^2} = \frac{bx + 2cx^2}{ex + 2fx^2}.$$

Cross-multiplying and rearranging the terms gives

$$\begin{aligned}&(0\cdot a+1\cdot bx+2\cdot cx^2)d\\&\quad+(-1\cdot a+0\cdot bx+1\cdot cx^2)ex\\&\quad+(-2\cdot a+-1\cdot bx+0\cdot cx^2)fx^2=0,\end{aligned}$$

which gives a solvable form of the Hudde equation. In general:

HUDDE'S RULE FOR QUOTIENTS. Let $f(x)=g(x)/h(x)$, where

$$g(x)=\sum_{k=0}^{n}a_kx^k \quad\text{and}\quad h(x)=\sum_{j=0}^{m}b_jx^j,$$

and suppose $f(x)$ has an extremum Z at $x=a$. Then $x=a$ is a double root to

$$\sum_{j=0}^{m}\left(b_jx^j\sum_{k=0}^{n}a_k(k-j)x^k\right).$$

Constrained extrema

At the end of his second letter Hudde applied his method to a constrained extrema; this allows the method to be extended to functions defined implicitly, which means that all algebraic functions can be handled using his methods. For a simple example, suppose we wish to maximize the objective function xy subject to the constraint $x^3+y^3=8xy$. Begin by assuming $Z=xy$ is the maximum. Solving the objective function for y we have $y=Z/x$, and substituting this into the constraint equation gives

$$x^3+\frac{Z^3}{x^3}=8Z.$$

Multiplying by x^3 and rearranging this gives

$$x^6-8Zx^3+Z^3=0.$$

The corresponding Hudde equation might be

$$6x^6-24Zx^3=0.$$

Solving for Z gives $Z=x^3/4$. Since $Z=xy$, we can equate the two expressions for Z to find $y=x^2/4$. Substituting this last into the constraint equation, we get an equation in x alone:

$$x^3+\frac{x^6}{64}=2x^3$$

Hence $x=4$. Since $y=x^2/4$, we also have $y=4$.

Why does Hudde's procedure work? Consider the problem of maximizing an objective function $g(x,y)$ subject to the constraint $f(x,y)=0$. The equation $g(x,y)=Z$ corresponds to a family of curves, and for any specific value of Z, the curve might or might not intersect the graph of $f(x,y)=0$. The intersections, if they exist, correspond to points where the objective function has value Z. Provided f and g are sufficiently smooth, then the level curves $f(x, y) = 0$ and $g(x, y) = Z$, where Z is a maximum or minimum of $g(x, y)$, must be mutually tangent at points corresponding to extrema. Hence the corresponding system of equations must have a double root, and Hudde's procedure is applicable.

The lost calculus

Hudde's work shows that any problem involving the derivatives of algebraic functions, even those defined implicitly, could be handled using only algebra and geometry. Other developments suggested that limit-free calculus could go much farther, even to the point of being able to solve all the traditional problems of calculus (at least for algebraic functions). We will describe these developments briefly, as they are interesting enough to warrant separate treatment.

In addition to solving the tangent and optimization problems, the derivative is also used to find points of inflection. This problem may also be solved algebraically by noting that the points of inflection correspond to points where the system of equations for the tangent line and curve has a triple root; this was first pointed out by Claude Rabuel in his 1730 edition of Descartes [**3**]. We leave the application of Hudde's procedures to finding inflection points as an exercise for the reader.

Meanwhile another approach to finding tangents, developed by Apollonius of Perga and subsequently revived by John Wallis in his *Treatise on Conic Sections* (1655), would lay the groundwork for a useful link between the derivative and the integral. In modern terms, the method used by Wallis and Apollonius is the following: Suppose we wish to find the tangent to a curve $y = f(x)$ at the point $(a, f(a))$. The tangent line $y = T(x)$ may be defined as the line resting on one side of the curve; hence either $f(x) > T(x)$ for all $x \neq a$, or $f(x) < T(x)$ for all $x \neq a$. (This is generally true for curves that do not change concavity; if the curve does change concavity, we must restrict our attention to intervals where the concavity is either positive or negative.)

For example, if we wish to find the tangent to $y = x^2$ at (a, a^2), then we want the line $y = m(x - a) + a^2$ to always be below the curve; hence it is necessary that

$$m(x - a) + a^2 \leq x^2$$

for all x, with equality occurring only for $x = a$. Rearranging gives

$$m(x - a) \leq x^2 - a^2$$

$$m(x - a) \leq (x - a)(x + a).$$

First, consider the interval $x > a$; then $x - a > 0$, and we may divide both sides of the inequality to obtain the inequality $m \leq x + a$ for all $x > a$; hence $m \leq 2a$ is sufficient to guarantee the line is below the curve for $x > a$. Next, on the interval $x < a$, it is necessary that $m \geq x + a$; hence $m \geq 2a$ is sufficient to guarantee the line is below the curve for $x < a$. Thus if $m = 2a$, the line will be below the curve for all $x \neq a$; therefore, the line will lie on one side of the curve and be tangent at $x = a$.

A few years after Hudde's work Isaac Barrow proved a version of the Fundamental Theorem of Calculus using the same type of double-inequality argument used by Wallis. Barrow's proof appears in his *Geometrical Lectures* (1670, but based on lectures given in 1664–1666); we present a slightly modernized form.

Consider the curve $y = f(x)$, assumed positive and increasing, and an auxiliary curve $y = F(x)$ with the property that $F(a)$ is the area under $y = f(x)$ and above the x-axis over the interval $0 \leq x \leq a$. Now consider the line that passes through

$(a, F(a))$ and has slope $f(a)$; let the equation of this line be $T(x) = f(a)(x - a) + F(a)$. We seek to prove that this line will be tangent to the graph of $y = F(x)$.

First, take any b where $0 \leq b \leq a$, and note that $F(a) - F(b)$ is the area under $y = f(x)$ over the interval $b \leq x \leq a$. Since (by assumption) $f(x)$ is positive and increasing, this area is smaller than $f(a)(a - b)$; hence $F(a) - F(b) < f(a)(a - b)$; rearranging, we have $F(b) > f(a)(b - a) + F(a) = T(b)$. Thus for all b in $0 \leq x \leq a$, the line $T(x)$ lies below the curve $y = F(x)$. In a similar manner if $b > a$, then $F(b) - F(a)$ is the area below the curve over the interval $a \leq x \leq b$, and by assumption this area is greater than $f(a)(b - a)$; thus $F(b) - F(a) > f(a)(b - a)$ or $F(b) > f(a)(b - a) + F(a) = T(b)$, and the line $T(x)$ is again below the curve $y = F(x)$ for all $x > a$. Thus the line $T(x)$ passing through $(a, F(a))$ with slope $f(a)$ will be tangent to the curve $y = F(x)$. $T(x) = f(a)(x - a) + F(a)$ will be tangent to the curve $y = F(x)$. If we take the geometric interpretation of the integral as the area under a curve and the derivative as the slope of the tangent line, then Barrow has proven:

THE FUNDAMENTAL THEOREM OF CALCULUS. (BARROW'S VERSION) *Let $f(x)$ be positive and increasing function on $I = [0, b]$, and let $F(x) = \int_0^x f(t)\,dt$ for all x in I. Then $F'(x) = f(x)$ for all x in I.*

Together, the work of Descartes, Hudde, Wallis, and Barrow was on the verge of creating a calculus of algebraic functions that at no point required the use of limits or infinitesimal quantities. The main advantage of the limit-based calculus of Newton and Leibniz introduced in the 1670s is that it is capable of handling transcendental functions. Thus despite the lack of a theory of limits and concern over the use of infinitesimals, the calculus of Newton and Leibniz quickly supplanted the calculus of Descartes and Hudde, and the "lost calculus" vanished from the mathematical scene.

REFERENCES

1. M. E. Baron, *The Origins of the Infinitesimal Calculus*, Dover, Mineola, NY, 1969.
2. C. Boyer and U. C. Merzbach, *A History of Mathematics*, John Wiley and Sons, New York, 1991.
3. D. E. Smith and M. Latham, *The Geometry of Rene Descartes*, Dover, Mineola, NY, 1925.
4. R. Descartes, *Geometria*, 2nd ed., Louis and Daniel Elsevier, Amsterdam, 1659.
5. R. Descartes, *Oeuvres de Descartes*, Victor Cousin, Paris, 1824.
6. C. C. Gillespie, *Dictionary of Scientific Biography*, Scribner, New York, 1970–1980.
7. A. W. Grootendorst, Johann Hudde's *Epistola secunda de maximis et minimis*, *Nieuw Arch. Wisk. (4)*, 5:3 (1987) 303–334.
8. V. Katz, *A History of Mathematics*, Addison-Wesley, Boston, 1998.
9. G. W. Leibniz, *Mathematische Schriften*, ed. C. I. Gerhardt, 1855, Olms, Hildsheim, 1971.
10. I. Newton, *Mathematical Works. Assembled with an Introduction by Derek T. Whiteside*, Johnson Reprint Corporation, New York, 1964–7.
11. D. Struik, *A Concise History of Mathematics*, Dover, Mineola, NY, 1987.
12. D. Struik, *A Source Book in Mathematics 1200–1800*, Harvard University Press, Cambridge, MA, 1969.
13. J. Suzuki, *A History of Mathematics*, Prentice-Hall, Upper Saddle River, NJ, 2002.

Basketball, Beta, and Bayes

MATTHEW RICHEY
PAUL ZORN
St. Olaf College
Northfield, Minnesota 55057
richeym@stolaf.edu
zorn@stolaf.edu

Problem B1 on the 2002 Putnam competition [**1**] reads as follows:

> PUTNAM PROBLEM (PP): Shanille O'Keal shoots free throws on a basketball court. She hits the first, misses the second, and thereafter the probability that she hits the next shot is equal to the proportion of the shots she has hit so far. What is the probability that she hits exactly 50 out of her first 100 shots?

The answer is $1/99$, as the reader may enjoy showing. Indeed, the probability that Shanille hits k of her next 98 shots is $1/99$ for *all* k with $0 \le k \le 98$; we will prove this later as a corollary to a more general result.

The fact that the number of later hits is uniformly distributed on $\{0, 1, \dots, 98\}$ may seem unexpected, but it reflects the fact that the starting conditions—one hit and one miss—convey very little information. In Bayesian jargon, as we will see, these starting conditions would be called *noninformative*, not because they convey no information whatever but because they lead to a uniform distribution on the outcome space.

Let's generalize the problem slightly. Suppose Shanille begins with a hits and b misses, and then takes n additional shots. We will consider two associated random variables:

$$S_n := \text{the number of successes among } n \text{ attempts;}$$
$$\theta_n := \frac{S_n + a}{a + b + n}.$$

We can think of θ_n as an "updated belief" in Shanille's shooting ability based on all accumulated data. In the Putnam problem (PP) we have $a = b = 1$ and we seek $\mathrm{P}(S_{98} = 49)$, the probability of exactly 49 hits among the next 98 shots. (We will write $\mathrm{P}(X = k)$ for the probability that the discrete random variable X has value k.)

> GENERALIZED PUTNAM PROBLEM (GPP): Shanille shoots free throws. To begin, she hits a and misses b shots; thereafter, she hits with probability equal to the proportion of hits so far. Determine the probability distribution of S_n. Equivalently, describe the distribution of θ_n.

As noted above, S_n is uniformly distributed on $\{0, 1, \dots, n\}$ for $a = b = 1$. We will soon see that S_n is *not* uniformly distributed for any other choice of a and b.

The GPP is a *probability* problem—at each stage the success probability θ is fixed—but its ingredients suggest the *Bayesian* approach to *statistics*:

- an initial belief (a *prior distribution*, in Bayesian jargon) about Shanille's shooting accuracy: $\theta = a/(a + b) = \text{hits}/(\text{hits} + \text{misses})$;
- data: the outcome of one or more shots;
- an updated (*posterior*) belief, based on the data, about Shanille's accuracy.

After solving the GPP we will propose and solve a Bayesian variant of the basketball problem, in which the GPP can be seen as "embedded." En route, we introduce Bayesian statistics in general and the *beta–binomial* model in particular. The Bayesian perspective can help explain some surprises in the GPP's solution.

Bayesian statistics: a primer

To put what follows in context, we note some differences between probability and statistics. Both disciplines deal with *parameters* (such as θ, the success probability in some experiment) and *data* or *random variables* (such as S_n in the GPP). A probabilist usually takes parameters as known, and studies properties of the data or random variables, such as their distribution and expected values. Statisticians study the inverse problem: Beginning from data, they try to describe parameters.

The difference between the Bayesian statistician and the (more common) *Frequentist* statistician centers mainly on how each views the roles of parameters and data. A Frequentist views parameters as *fixed but unknown* quantities, and the data as *random.* Inferences, such as those concerning confidence intervals for parameters, are obtained through thought experiments starting with "Imagine all possible data produced by the parameter" or "If we sampled arbitrarily often ...."

Consider, for instance, a Frequentist method, the one-sample t-test for estimating a population mean θ. Suppose that $X_1, X_2, \dots, X_n$ is an independent, identically distributed sequence of random variables and let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. The distribution of the X_i may not be known exactly, but some assumptions are made, such as that the distribution is approximately normal and that the X_i have finite mean and variance.

Given specific data $x_1, x_2, \dots, x_n$ and their sample mean $\bar{x}$, one obtains a specific 95% confidence interval $(\bar{x} - m, \bar{x} + m)$, with $\bar{x}$ a point estimate for θ. (The value m is the margin of error and is determined from the variability observed in the data.) Notice that, from the Frequentist perspective, 95% is *not* the probability that θ lies in the calculated interval. On the contrary, θ is fixed while the confidence interval is random. The 95% probability concerns "coverage": If one samples often, each time calculating a confidence interval, then about 95% of these intervals will contain θ.

A Bayesian, by contrast, sees the data as fixed, but expresses *belief* about a parameter θ as a probability distribution—subject to change as additional data arise. A Bayesian would concede that in some simple cases θ has a "true" value. If, say, θ is a coin's probability of landing "heads," then (according to the law of large numbers) one could approximate θ by flipping the coin many times. But in less repeatable cases, such as whether it will rain tomorrow, a Bayesian would rather model our *belief* (or uncertainty) about this likelihood.

Consider, for instance, how a Bayesian makes an inference about a population parameter, θ, contained in a parameter space Ω. The Bayesian starts with a *prior distribution* $\pi(\theta)$, which expresses an initial belief about the relative likelihood of possible values of θ. The data $\mathbf{X} = X_1, X_2, \dots, X_n$ and their *joint distribution* $f(\mathbf{X} = \mathbf{x} \mid \theta)$ are known. To obtain the distribution of θ conditioned on $\mathbf{X}$ (called the *posterior distribution* and denoted by $f(\theta \mid \mathbf{X} = \mathbf{x})$), a Bayesian invokes *Bayes's rule*:

$$f(\theta \mid \mathbf{X} = \mathbf{x}) = \frac{f(\mathbf{X} = \mathbf{x} \mid \theta)\pi(\theta)}{\int_{\Omega} f(\mathbf{X} = \mathbf{x} \mid \hat{\theta})\pi(\hat{\theta})\, d\hat{\theta}}, \tag{1}$$

where Ω is the parameter space for θ. (The integral is a sum if θ is a discrete random variable.) In simple cases (1) can be evaluated in closed form; other cases require numerical methods, such as Markov Chain Monte Carlo (MCMC) techniques. A variety of sources discuss Bayesian methods in general [**3, 4, 5, 6**] and MCMC techniques in particular [**2, 4, 5**].

Knowledge of $f(\theta \mid \mathbf{X} = \mathbf{x})$ amounts, from a statistical perspective, to knowing "everything" about θ. For example, a Bayesian might use the expectation $\mathrm{E}(\theta \mid \mathbf{X} = x)$ as a point estimate of θ. A 95% confidence interval for θ could be any interval (θ_1, θ_2) for which

$$\int_{\theta_1}^{\theta_2} f(\theta \mid \mathbf{X} = \mathbf{x})\, d\theta = 0.95. \tag{2}$$

Among all such intervals, one might choose the one symmetric about the expected value $\mathrm{E}(\theta \mid \mathbf{X} = \mathbf{x})$, or, alternatively, the narrowest interval for which (2) holds. In any event, (2) expresses the Bayesian's belief that θ lies in the confidence interval with 95% probability. Thus, Bayesians and Frequentists agree that a 95% confidence interval depends on the data, but a Bayesian expresses the dependence explicitly, using the posterior distribution.

Bayesians v. Frequentists Where do Bayesians and Frequentists disagree? Following is a somewhat caricatured discussion; in practice, many statisticians adopt aspects of both approaches.

Frequentists see the Bayesian notion of a prior distribution as too subjective. Bayesians counter that Frequentists make equally subjective assumptions, such as that a given distribution is normal. Frequentists claim that Bayesians can, by choosing a suitable prior, obtain any desired result. Not so, Bayesians reply: Sufficient data always "overwhelm the prior"; besides, ill-chosen priors are soon revealed as such. Frequentists appreciate the computational tractability of their methods, and see Bayesian posterior distributions as unduly complex. Not any more, say Bayesians—modern computers and algorithms make posterior distributions entirely manageable.

Bayesian inference the beta–binomial way

We illustrate Bayesian inference using the *beta–binomial* model (which we apply later to a Bayesian-flavored basketball problem). Suppose X is a *Bernoulli* random variable with values 0 or 1 and $\mathrm{P}(X = 1) = \theta$; we write this as $X \sim \mathrm{Bernoulli}(\theta)$. For fixed n, let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables and set $Y = X_1 + X_2 + \cdots + X_n$. Then Y is the number of successes (1s) in n Bernoulli trials, a *Binomial* random variable, and we write $Y \sim \mathrm{Binomial}(n, \theta)$.

As a prior distribution on θ Bayesians often choose a $\mathrm{Beta}(a, b)$ distribution on $[0, 1]$. This distribution is defined for arbitrary positive a and b, and has density

$$f_{a,b}(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, \theta^{a-1}(1-\theta)^{b-1};$$

we write $\theta \sim \mathrm{Beta}(a, b)$. (The *gamma function* is defined for $x > 0$ by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt.$$

Among its familiar properties are $\Gamma(x) = (x-1)\Gamma(x-1)$ and $\Gamma(1) = 1$. In particular, $\Gamma(n) = (n-1)!$ for positive integers n. It is an interesting exercise to show directly that $\int_0^1 f_{a,b}(\theta)\, d\theta = 1$.)

FIGURE 1 shows Beta(a, b) densities $f_{a,b}(\theta)$ for several choices of a and b. Notice how the parameters control the shape of the distribution. Straightforward calculations give the expectation and variance for $\theta \sim$ Beta(a, b):

$$\mathrm{E}(\theta) = \frac{a}{a+b}; \qquad \mathrm{Var}(\theta) = \frac{ab}{(a+b)^2(a+b+1)}. \tag{3}$$

These formulas suggest that a and b can be thought of as imagined prior hits and misses among $n = a + b$ attempts. Indeed, suppose that a, b are actual hits and misses from a Binomial$(a + b, \theta)$ distribution. Then the Frequentist's (unbiased) estimate of θ,

$$\hat{\theta} = \frac{a}{a+b},$$

is precisely the Bayesian's expected value of θ. Similarly,

$$\mathrm{Var}\left(\hat{\theta}\right) = \frac{\hat{\theta}(1-\hat{\theta})}{a+b} = \frac{ab}{(a+b)^3},$$

which is $(a + b + 1)/(a + b)$ times the Bayesian's Var(θ).

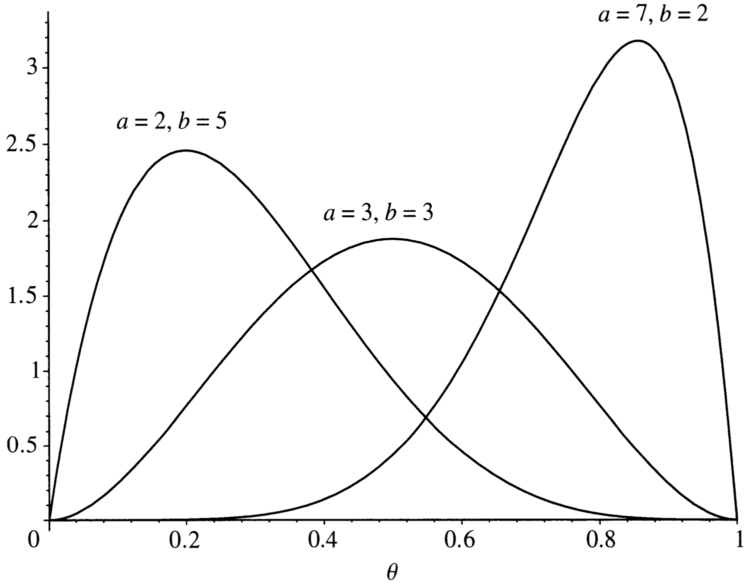


Figure 1 Plots of $f_{a,b}(\theta)$ for different values of a, b

Notice also that one can arrange *any* positive mean and variance by choosing a and b judiciously in (3). For example, setting $a = 7$ and $b = 2$ in (3) reflects Bayesian belief in a mean near 7/9; see FIGURE 1. Setting $a = b = 1$ produces the *uniform* distribution on [0, 1], known here as the *noninformative prior* because it reflects little or no prior belief about θ.

The following well-known result links the beta and the binomial distributions in the Bayesian setting.

PROPOSITION 1. *Suppose that $\theta \sim$ Beta(a, b) and $X \sim$ Binomial(n, θ), so the prior distribution has density function*

$$\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, \theta^{a-1}(1-\theta)^{b-1}$$

and

$$\mathrm{P}(X = k \mid \theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

Then θ *has posterior distribution* $\mathrm{Beta}(a+k, b+n-k)$. *That is,*

$$f(\theta \mid X = k) = \frac{\Gamma(a+b+n)}{\Gamma(a+k)\Gamma(b+n-k)} \theta^{a+k-1} (1-\theta)^{b+n-k-1}.$$

Note Proposition 1 and equation (3) imply that the posterior expectation $\mathrm{E}(\theta)$ satisfies

$$\mathrm{E}(\theta) = \frac{a+k}{a+b+n} = \frac{a+b}{a+b+n} \cdot \frac{a}{a+b} + \frac{n}{a+b+n} \cdot \frac{k}{n}.$$

Thus, $\mathrm{E}(\theta)$ is a convex combination of the prior mean $a/(a+b)$ and the sample mean k/n; the Bayesian method favors the prior for small samples but tends toward the Frequentist figure as the sample size increases.

Proof. Bayes's rule (1) gives

$$f(\theta \mid X = k) = \frac{\binom{n}{k}\theta^k (1-\theta)^{n-k} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}}{\int_0^1 \binom{n}{k}\hat{\theta}^k (1-\hat{\theta})^{n-k} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \hat{\theta}^{a-1}(1-\hat{\theta})^{b-1}\, d\hat{\theta}}.$$

Canceling common factors and collecting powers of θ and $(1-\theta)$ gives

$$f(\theta \mid X = k) = C\,\theta^{a+k-1}(1-\theta)^{b+n-k-1}$$

for some constant C. Since $f(\theta \mid X = k)$ is a probability density, it must be $\mathrm{Beta}(a+k, b+n-k)$. ■

The beta-binomial model's main ingredients resemble those of the GPP:

- a prior distribution: $\theta \sim \mathrm{Beta}(a, b)$, with $\mathrm{E}(\theta) = a/(a+b)$;
- data: $X \sim \mathrm{Binomial}(n, \theta)$;
- a posterior distribution: $\theta \sim \mathrm{Beta}(a+k, b+n-k)$ with

$$\mathrm{E}_{\mathrm{post}}(\theta) = \frac{a+k}{a+b+n}.$$

The analogy continues. For the GPP, the initial success probability has the fixed value $\theta_0 = a/(a+b)$; for the beta-binomial, θ_0 is random, but $\mathrm{E}(\theta_0) = a/(a+b)$. After n shots with k successes, the GPP gives $\theta_n = (a+k)/(a+b+n)$, which is also the expected value of θ_n after a beta-binomial update. A similar theme will be seen in what follows.

Repeated updates Repeated beta–binomial updates turn out to be equivalent to a single beta–binomial update. More precisely, let n_1 and n_2 be positive integers and a and b fixed positive numbers, with $\theta \sim \mathrm{Beta}(a, b)$ and $k_1 \sim \mathrm{Binomial}(n_1, \theta)$. Updating θ once gives $\theta \sim \mathrm{Beta}(a+k_1, b+n_1-k_1)$. If $k_2 \sim \mathrm{Binomial}(n_2, \theta_1)$, then updating θ again gives

$$\theta \sim \mathrm{Beta}\left(a + (k_1+k_2),\ b + (n_1+n_2) - (k_1+k_2)\right),$$

which shows that θ can be obtained from *one* beta–binomial update with $k_1 + k_2$ successes in $n_1 + n_2$ trials. This result plays well with Bayesian philosophy: The data and the model—not an arbitrary subdivision into two parts—determine our posterior belief.

If $n = 1$, the binomial random variable reduces to the Bernoulli case; the result might be called a *beta–Bernoulli model.* The preceding observation implies that updating a Beta(a, b) prior distribution with a sequence of beta–Bernoulli updates is equivalent to a single beta–binomial update.

Bayesian prediction Our goal so far has been to combine a prior distribution with data to predict future values of a parameter θ. A natural next step is to use our new knowledge of θ to predict future values of the random variable X. If θ were fixed we would obtain the traditional binomial distribution:

$$\mathrm{P}(X = k) = \binom{n}{k}\theta^k(1-\theta)^{n-k}.$$

Because our θ is random, we average over all values of θ, weighted by its posterior density. Not surprisingly, the randomness of θ leads to greater variability in X. With $\theta \sim$ Beta(a, b), we have

$$\begin{aligned}
\mathrm{P}(X = k) &= \int_0^1 \mathrm{P}(X = k \mid \theta) f_{a,b}(\theta)\, d\theta \\
&= \int_0^1 \binom{n}{k}\theta^k(1-\theta)^{n-k}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\, d\theta \\
&= \binom{n}{k}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 \theta^{a+k-1}(1-\theta)^{b+n-k-1}\, d\theta \\
&= \binom{n}{k}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+k)\Gamma(b+n-k)}{\Gamma(a+b+n)}. \qquad (4)
\end{aligned}$$

This result is called the *marginal distribution* of X, or, in Bayesian parlance, the *predictive posterior distribution.*

The distribution (4), known as the *diffuse binomial,* is close kin to the ordinary binomial distribution. FIGURE 2 suggests how: Both distributions have the same general

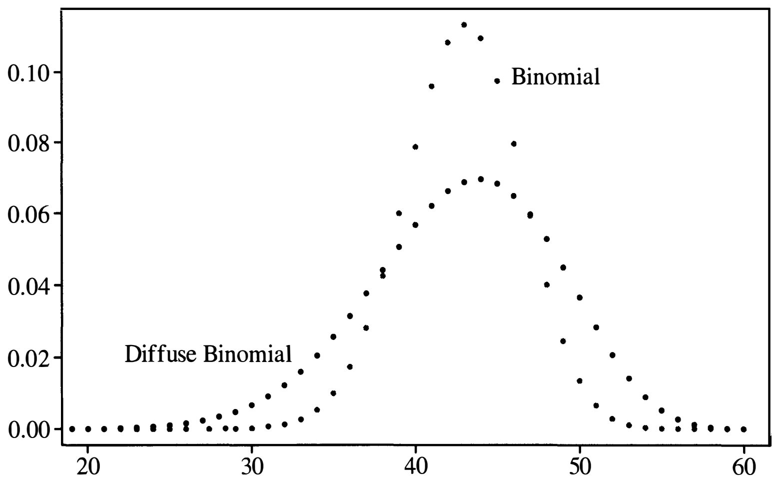


Figure 2 Distributions of $\mathrm{P}(X = k)$ for the diffuse binomial with $n = 60$, $a = 25$, $b = 10$; and for Binomial$(60, \theta)$ with $\theta = 25/35$

shape but the diffuse version has larger "tails." Equation (4) has more than passing interest—it will reappear when we solve the GPP.

Like the updating process, the predictive posterior distribution does not depend on a partitioning of n. To see this, suppose that $\theta_1 \sim \text{Beta}(a, b)$, $X_1 \sim \text{Binomial}(n_1, \theta_1)$, $\theta_2 \sim \text{Beta}(a + k_1, b + n_1 - X_1)$, and $X_2 \sim \text{Binomial}(n_2, \theta_2)$.

If $X = X_1 + X_2$, $n = n_1 + n_2$, and $k = k_1 + k_2$, then we have

$$\begin{aligned}
\text{P}(X = k) &= \sum_{k_1=0}^{n_1} \sum_{k_2=k-k_1}^{n_2} P\,(X_1 = k_1 \text{ and } X_2 = k_2) \\
&= \sum_{k_1=0}^{n_1} \sum_{k_2=k-k_1}^{n_2} P(X_2 = k_2 \mid X_1 = k_1)\,\text{P}(X_1 = k_1).
\end{aligned}$$

Substituting (4) in both factors above gives $\text{P}(X = k) =$

$$\begin{aligned}
&\sum_{k_1=0}^{n_1} \sum_{k_2=k-k_1}^{n_2} \binom{n_2}{k_2} \frac{\Gamma(a+b+n_1)}{\Gamma(a+k_1)\Gamma(b+n_1-k_1)} \\
&\quad \times \frac{\Gamma(a+k_1+k_2)\Gamma(b+(n_1+n_2)-(k_1-k_2))}{\Gamma(a+b+(n_1+n_2))} \\
&\quad \times \binom{n_1}{k_1} \frac{\Gamma(a+b)}{\Gamma(b)\Gamma(a+b)} \frac{\Gamma(a+k_1)\Gamma(b+n_1-k_1)}{\Gamma(a+b+n_1)} \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+k)\Gamma(b+n-k)}{\Gamma(a+b+n)} \sum_{k_1=0}^{n_1} \sum_{k_2=k-k_1}^{n_2} \binom{n_2}{k_2}\binom{n_1}{k_1} \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+k)\Gamma(b+n-k)}{\Gamma(a+b+n)} \binom{n}{k}.
\end{aligned}$$

Thus, an imagined partitioning of the attempts does not—and to a Bayesian should not—change the predicted outcome.

No free lunch One might also ask about the marginal posterior distribution of θ, found by averaging over all possible outcomes of X. To a Bayesian this exercise is fruitless: Imagining *all* possible outcomes should not change the initial belief about θ. The following calculation explains this for θ a continuous and X a discrete random variable. (The same result holds for any combination of discrete and continuous random variables.)

Consider a continuous random variable θ, with prior density $\pi(\theta)$ and values in Ω, and suppose that a discrete random variable X, with values x in a set S, has conditional density $\text{P}(X = x \mid \theta)$. Then, for each x, the conditional posterior distribution for θ has density

$$f_{\text{post}}(\theta \mid X = x) = \frac{\text{P}(X = x \mid \theta)\,\pi(\theta)}{\int_\Omega \text{P}(X = x \mid \hat{\theta})\pi(\hat{\theta})\,d\hat{\theta}}.$$

Now we can find the marginal posterior distribution:

$$\begin{aligned}\hat{f}_{\text{post}}(\theta) &= \sum_{x\in S} f_{\text{post}}(\theta \mid X = x)\,\mathrm{P}(X = x)\\ &= \sum_{x\in S} \frac{\mathrm{P}(X = x \mid \theta)\pi(\theta)}{\int_\Omega \mathrm{P}(X = x \mid \hat{\theta})\pi(\hat{\theta})\,d\hat{\theta}}\,\mathrm{P}(X = x)\\ &= \sum_{x\in S} \frac{\mathrm{P}(X = x \mid \theta)\pi(\theta)}{\mathrm{P}(X = x)}\,\mathrm{P}(X = x) = \pi(\theta)\sum_{x\in S} P(X = x \mid \theta) = \pi(\theta).\end{aligned}$$

That the prior and the marginal posterior distributions for θ are identical illustrates a Bayesian "no free lunch" principle: Real data, not thought experiments, are needed to update a prior distribution.

Bayesian basketball Yet again the tireless Shanille shoots free throws and we seek to model our belief about her success probability θ. As cautious Bayesians we choose for θ the "noninformative prior"—the uniform distribution on $[0, 1]$. Lacking further information, we see *every* subinterval of $[0, 1]$ of width $\Delta\theta$ as equally likely to contain the "true" θ. A more informed Bayesian fan might choose an informative prior, say, Beta(4, 4). FIGURE 3 shows both choices.

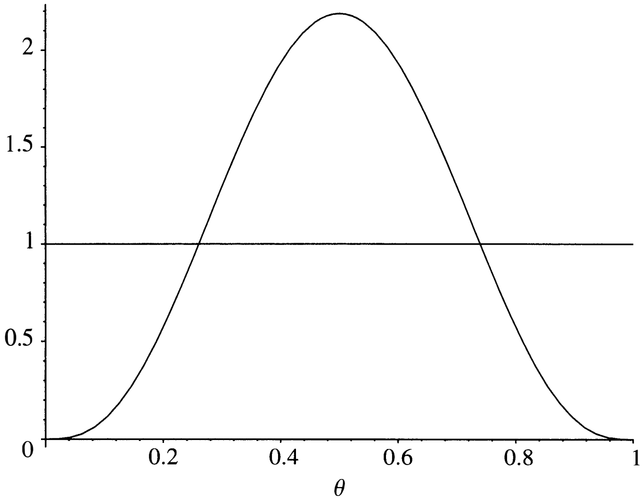


Figure 3 The Beta(1, 1) (uniform, noninformative) and Beta(4, 4) priors for θ, Shanille's success probability

Here, unlike in the GPP, we never know Shanille's exact success probability. (If we knew θ exactly we wouldn't bother to estimate it!) If Shanille shoots n free throws and hits k, then, using the beta–binomial model and invoking Proposition 1, we obtain the following posterior distributions for θ:

$$\text{for prior Beta}(1,1)\text{:}\quad \text{Beta}(1+k, 1+n-k),\ \text{with } \mathrm{E}_{\text{post}}(\theta) = \frac{k+1}{n+2}$$

$$\text{for prior Beta}(4,4)\text{:}\quad \text{Beta}(4+k, 4+n-k),\ \text{with } \mathrm{E}_{\text{post}}(\theta) = \frac{k+4}{n+8}.$$

In each case a 95% confidence interval (u, v) for θ can be obtained by finding (perhaps numerically) any numbers u and v for which

$$\frac{\Gamma(n+a+b)}{\Gamma(k+a)\Gamma(n-k+b)}\int_u^v \theta^{a-1+k}(1-\theta)^{b-1+n-k}\,d\theta = 0.95.$$

The following table compares the Frequentist (in this case binomial) and two different Bayesian point estimates and symmetric 95% confidence intervals for θ using different values for n and k.

TABLE 1: Frequentist and Bayesian estimates for θ

Model	(k, n)	Point estimate	95% confidence	Interval width
Frequentist	(2, 3)	$2/3 \approx .667$	(.125, .982)	.857
	(30, 45)	$30/45 \approx .667$	(.509, .796)	.287
	(60, 90)	$60/90 \approx .667$	(.559, .760)	.201
Bayesian, $a = 1, b = 1$	(2, 3)	$3/5 = .600$	(.235, .964)	.729
	(30, 45)	$31/47 \approx .660$	(.527, .793)	.266
	(60, 90)	$61/82 \approx .663$	(.567, .759)	.191
Bayesian, $a = 4, b = 4$	(2, 3)	$6/11 \approx .545$	(.270, .820)	.550
	(30, 45)	$34/54 \approx .642$	(.514, .769)	.254
	(60, 90)	$64/98 \approx .653$	(.560, .747)	.187

Observe:

- The Frequentist point estimate for θ is always 2/3, the sample proportion of successes. Bayesian point estimates, by contrast, are weighted averages of the sample proportions and the prior mean, 1/2.
- For fixed k and n, widths of 95% confidence intervals decrease as we move from the Frequentist through the noninformative Bayesian to the informative Bayesian perspective.
- The Frequentist and Bayesian estimates come together as n increases. "Data overwhelm the prior," a Bayesian might say.
- For the informative prior Beta(4, 4), FIGURE 4 shows the densities for θ becoming narrower and moving toward the sample mean as n increases.

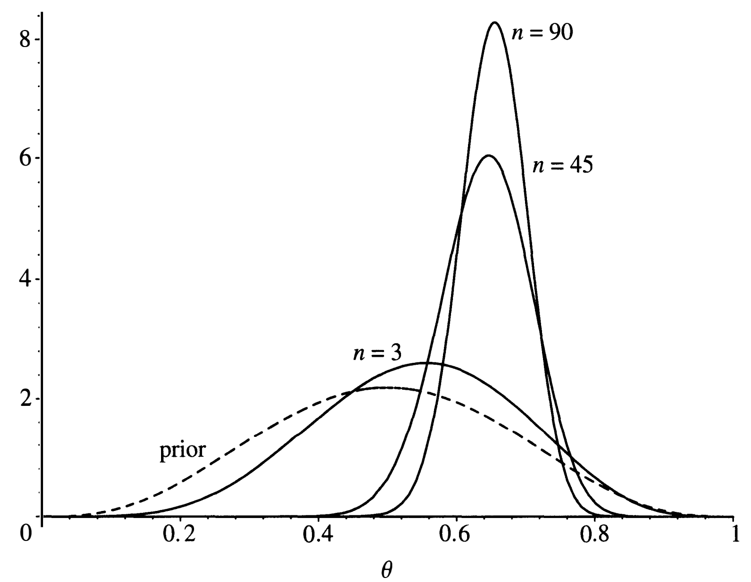


Figure 4 An informative prior (Beta(4, 4), dashed) and several posterior densities for θ

Solving the GPP

Leaving (for now) the realm of statistics, we now return to probability and the GPP. We asserted above without proof that in the original PP (with $a = b = 1$), the random variable S_n is uniformly distributed:

$$P(S_n = k) = \frac{1}{n+1} \quad \text{for } k = 0, 1, \ldots, n.$$

Instead of proving this directly, we calculate $P(S_n = k)$ for arbitrary positive integers a and b. This takes a little work, but as a small reward we see the beta–binomial predictive posterior distribution (4) crop up again. We see, too, that (5) reduces to the uniform distribution if, but *only* if, $a = b = 1$.

PROPOSITION 2. *With notation as in the GPP, we have*

$$P(S_n = k) = \binom{n}{k} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+k)\Gamma(b+n-k)}{\Gamma(a+b+n)}. \tag{5}$$

Proof. There are $\binom{n}{k}$ ways to achieve k successes in n attempts. The key observation is that *all* sequences of k successes and $n-k$ failures are equally probable. To see why, consider any point at which s successes and f failures have occurred. The probability of a success followed by a failure is thus

$$\frac{s}{s+f}\,\frac{f}{s+f+1},$$

while the probability of a failure followed by a success is

$$\frac{f}{s+f}\,\frac{s}{s+f+1}.$$

Because these two quantities are equal we can rearrange *any* sequence of successes and failures so that (say) all k successes come first. The probability of this special arrangement is

$$\begin{aligned}&\frac{a(a+1)\cdots(a+k-1)\; b(b+1)\cdots(b+n-k-1)}{(a+b)(a+b+1)\cdots(a+b+n-1)}\\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\,\frac{\Gamma(a+k)\Gamma(b+n-k)}{\Gamma(a+b+n)};\end{aligned}$$

where the equality follows by rewriting the right side in terms of factorials. Finally, the probability of any k successes among n attempts is $\binom{n}{k}$ times the preceding quantity. ∎

Back to the Beta Now we can start to explain why formulas (4) and (5) are identical. FIGURE 5 shows probability distributions of θ_{100} (Shanille's success rate on her 100th shot) for several choices of a and b.

The plots in FIGURE 5 closely resemble the corresponding beta density functions $f_{a,b}$ in FIGURE 1; they differ mainly by a vertical scale factor of $a+b+n$. We can describe this graphical similarity in probabilistic language. Let I be any interval contained in $[0, 1]$ and $\theta \sim \text{Beta}(a, b)$. We will show that

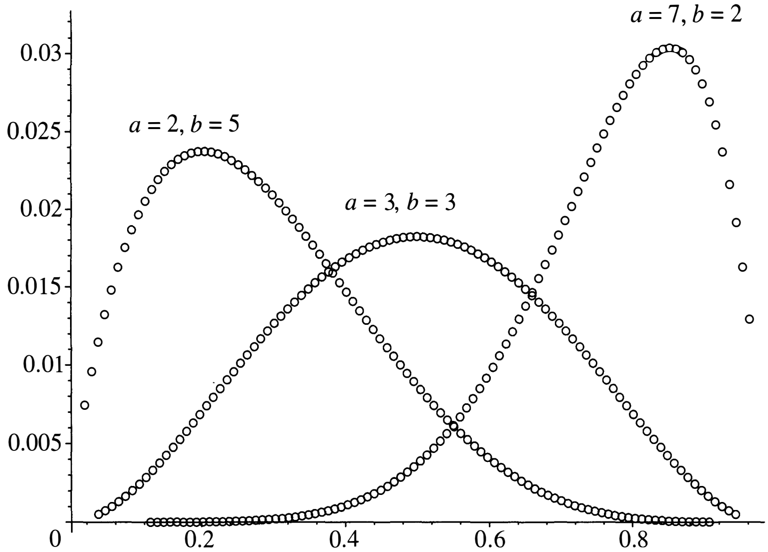


Figure 5 Density for plots of θ_{100} for different values of a and b; compare to FIGURE 1

$$\left|\mathrm{P}(\theta_n \in I) - \mathrm{P}(\theta \in I)\right| = O(1/n), \tag{6}$$

which means that if we start with a hits and b misses, then for large n the *discrete* probability distribution for θ_n is approximated by the *continuous* Beta(a, b) distribution.

To prove equation (6) is a routine but slightly messy exercise in O-arithmetic. First, let

$$t_{n,k} = \frac{a+k}{a+b+n} \quad \text{and} \quad P_{n,k} = \mathrm{P}(\theta_n = t_{n,k})$$

for $k = 0, 1, 2, \ldots, n$. Equation (6) will follow if we show that

$$f_{a,b}(t_{n,k}) = (a+b+n)\, P_{n,k} + O(1/n) \tag{7}$$

uniformly in k. Because $f_{a,b}(t_{n,k})$ and $P_{n,k}$ have the common factor

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)},$$

to prove (7) it will suffice to show that

$$D_{n,k} := (a+b+n)\binom{n}{k}\frac{\Gamma(a+k)\Gamma(b+n-k)}{\Gamma(a+b+n)} = t_{n,k}^{a-1}\,(1-t_{n,k})^{b-1} + O(1/n),$$

uniformly in k. Writing out $D_{n,k}$ in terms of factorials gives

$$D_{n,k} = (a+b+n) \times \frac{(a+k-1)(a+k-2)\cdots(k+1)\,(b+n-k-1)(b+n-k-2)\cdots(n-k+1)}{(a+b+n-1)(a+b+n-2)\cdots(n+1)}.$$

Note that both numerator and denominator have $a+b-1$ factors. Dividing *all* factors by $N = a+b+n$ and substituting $t_{n,k} = (a+k)/N$ and $1 - t_{n,k} = (b+n-k)/N$ now gives

$$\frac{\left(t_{n,k}-\frac{1}{N}\right)\left(t_{n,k}-\frac{2}{N}\right)\cdots\left(t_{n,k}-\frac{a-1}{N}\right)\left(1-t_{n,k}-\frac{1}{N}\right)\left(1-t_{n,k}-\frac{2}{N}\right)\cdots\left(1-t_{n,k}-\frac{b-1}{N}\right)}{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{a+b-1}{N}\right)\cdots}$$

$$=\frac{\left(t_{n,k}+O(1/n)\right)^{a-1}\left(1-t_{n,k}+O(1/n)\right)^{b-1}}{(1+O(1/n))^{a+b-1}}$$

$$=t_{n,k}^{a-1}\left(1-t_{n,k}\right)^{b-1}+O(1/n)=D_{n,k},$$

as desired. (The assiduous reader is invited to verify the O-arithmetic in the last step.)

Next we use (7) to prove (6). To begin, write $I=[\alpha,\beta]$, fix n, and suppose that I contains the $t_{n,k}$ ranging from t_{n,k_0} to t_{n,k_M}, where $0\le k_0<k_M\le n$. Now the continuity of $f_{a,b}$ on [0, 1] implies that

$$\mathrm{P}(\theta\in I)=\int_\alpha^\beta f_{a,b}(t)\,dt=\int_{t_{n,k_0}}^{t_{n,k_M}} f_{a,b}(t)\,dt+O(1/n).$$

The last integral can, in turn, be approximated by a convenient approximating sum with step size $1/(a+b+n)$:

$$\int_{t_{n,k_0}}^{t_{n,k_M}} f_{a,b}(t)\,dt=\frac{f_{a,b}(t_{n,k_0})+f_{a,b}(t_{n,k_0+1})+\cdots+f_{a,b}(t_{n,k_M})}{a+b+n}+O(1/n).$$

(The $O(1/n)$ assertion holds because the integrand $f_{a,b}(t)$ has bounded first derivative.) Now (7) implies that

$$\frac{f_{a,b}(t_{n,k})}{a+b+n}=P_{n,k}+O\left(1/n^2\right)$$

for all k, which in turn gives

$$\begin{aligned}\frac{f_{a,b}(t_{n,k_0})+\cdots+f_{a,b}(t_{n,k_M})}{a+b+n}&\\ &=P_{n,k_0}+P_{n,k_0+1}+\cdots+P_{n,k_M}+M\cdot O\left(1/n^2\right)+O(1/n)\\ &=P_{n,k_0}+P_{n,k_0+1}+\cdots+P_{n,k_M}+O(1/n),\end{aligned}$$

and (6) follows.

Great expectations Next we find expected values of S_n and θ_n; perhaps surprisingly, we can do so without recourse to their explicit distributions, given in (2). First we write $S_n=S_{n-1}+X_n$, where X_n is 1 if the nth shot hits and 0 otherwise. Then we have

$$\begin{aligned}\mathrm{P}(X_n=1)&=\sum_{k=0}^{n-1}\mathrm{P}(X_n=1\mid S_{n-1}=k)\mathrm{P}(S_{n-1}=k)\\ &=\sum_{k=0}^{n-1}\frac{a+k}{a+b+n-1}\mathrm{P}(S_{n-1}=k)\\ &=\frac{a\sum\mathrm{P}(S_{n-1}=k)+\sum k\,\mathrm{P}(S_{n-1}=k)}{a+b+n-1}=\frac{a+\mathrm{P}(S_{n-1})}{a+b+n-1}.\end{aligned}\tag{8}$$

Thus,

$$\begin{aligned} P(S_n) &= P(S_{n-1}) + P(X_n) = P(S_{n-1}) + P(X_n = 1) \\ &= P(S_{n-1}) + \frac{a + P(S_{n-1})}{a+b+n-1} = \frac{a}{a+b+n-1} + P(S_{n-1})\,\frac{a+b+n}{a+b+n-1}. \end{aligned}$$

Now clearly $E(S_0) = 0$, and it follows by induction that

$$P(S_n) = \frac{na}{a+b}$$

for $n \geq 0$. The expected value of θ_n is readily found—and seen to be constant:

$$P(\theta_n) = \frac{a + P(S_n)}{a+b+n} = \frac{a + \frac{na}{a+b}}{a+b+n} = \frac{a}{a+b}.$$

In particular, (8) shows that

$$P(X_n = 1) = P(\theta_{n-1}) = \frac{a}{a+b}.$$

That $P(X_n = 1)$ is constant recalls the earlier Bayesian "no free lunch" result, in which averaging over all possible outcomes did not produce new knowledge. Here, too, Shanille's initial success probability—$a/(a+b)$—remains unaltered as we imagine future outcomes.

Another way to think of this is to imagine a large number, M, of Shanille-clones, each starting with a hits and b misses and hence an initial $a/(a+b)$ success probability. Thereafter, each Shanille updates her own probability by the GPP rule. At each stage the distribution of all M Shanilles' success probabilities is as described in Proposition 2, but the group's *average* success probability remains at $a/(a+b)$, and its expected number of hits is $Ma/(a+b)$. After many stages, the distribution of the individual success probabilities strongly resembles a Beta(a, b) distribution.

Bayesian basketball Finally, we consider the *Bayesian Beta–Binomial Basketball Putnam Problem* (BBBPP):

> Shanille O'Keal, now a converted Bayesian, shoots free throws. Starting with a Beta distribution, at each stage she draws a value of θ from the distribution and, with success probability, shoots a basket and then updates her distribution by the beta-binomial method. Describe the marginal distributions of θ_n and of S_n, her success probability and total number of successes after n shots.

The BBBPP extends the GPP in several senses. First, the GPP starting points, a hits and b misses, mirror the initial Beta parameters a and b. Second, while the GPP Shanille is certain of her success probability at each stage, the BBBPP Shanille has less certainty—but she could always return to her GPP ways by replacing the random draw of θ_n with its expected value at each stage.

In the BBBPP setting S_n and θ_n have a new relationship: S_n is still a discrete random variable, but θ_n is now a continuous random variable that describes Shanille's skill. Now S_n is conditioned on θ_k, for $0 \leq k < n$, and θ_n is conditioned on S_n. Each θ_k is obtained by a beta-Bernoulli update of θ_{k-1}.

We might seem to have traded the relatively simple discrete GPP, with its single discrete random variable S_n, for a more complex BBBPP, with two intertwined random variables, one discrete and the other continuous. The bargain is better than it might seem—properties of the beta–binomial and of the Bayesian scheme turn out to simplify our solution. For instance, we can replace the sequence $\theta_0, \theta_1, \ldots, \theta_n$ of beta-Bernoulli updates with a *single* beta–binomial, starting with θ_0 and producing θ_n.

The marginal distribution of θ_n follows from the "no free lunch" principle: The marginal posterior and the prior distribution are identical. Thus, for all n, the success probability θ_n has marginal posterior distribution Beta(a, b), and $\mathrm{E}(\theta_n) = a/(a+b)$. In the GPP, by comparison, the distributions of θ_n *approximate* the Beta(a, b) distribution for large n, and $\mathrm{E}(\theta_n) = a/(a+b)$ for all n.

Consider, in particular, the noninformative (uniform) Beta$(1, 1)$ prior distribution. Averaging over all possible outcomes assures that, at the nth stage, all values of θ_n *remain* equally likely. This brings the original PP to mind: All values θ_n for the updated success probability are equally likely at each stage.

Now we consider S_n. For the GPP the distribution of S_n, found in Proposition 2, is identical to the predictive posterior of a Beta(a, b) prior, shown in (4). The same result holds for the BBBPP, but here the connection is more natural. A sequence of n Bernoulli updates is equivalent, as shown earlier, to a single binomial update with n trials. The probability we seek, $\mathrm{P}(S_n = k)$, was found in (4):

$$\mathrm{P}(S_n = k) = \binom{n}{k} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+k)\Gamma(b+n-k)}{\Gamma(a+b+n)}.$$

It follows that $\mathrm{E}(S_n) = na/(a+b)$ for all n, as in the GPP.

Conclusion

Here we end our tour of Shanille's basketball adventures and our detour through Bayesian statistics. We explored the GPP, a problem in probability, by "embedding" it in a Bayesian context. The embedding amounts, essentially, to replacing a fixed θ with a random variable with *expectation* θ. Almost every property of the fixed-θ case is reflected in a property of the variable-θ setting. The embedding helped reveal some interesting properties of the GPP and links to Bayesian principles, such as the "no free lunch" property.

Finally, a confession: Even the BBBPP smacks more of probability than of Bayesian statistics at its purest. To a fully committed Bayesian, what changes over successive free throw attempts is not really Shanille's success probability, as the BBBPP implies. Shanille's skill remains unchanged throughout—only our belief is knowable, and subject to updating. For Bayesians, it's all about belief.

REFERENCES

1. 63rd Annual William Lowell Putnam Mathematical Competition, this MAGAZINE **76** (2003), 76–80.
2. Siddhartha Chib, Edward Greenberg, Understanding the Metropolis-Hastings algorithm, *The American Statistician* **49** (1995), 327–335.
3. Peter Congdon, *Bayesian Statistical Modeling*, John Wiley and Sons, New York, NY, 2001.
4. Bradley Carlin and Thomas Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, CRC Press, Boca Raton, FL, 2000.
5. Andrew Gelman, John Carlin, Hal Stern, and Donald Rubin, *Bayesian Data Analysis*, CRC Press, Boca Raton, FL, 1995.
6. Robert V. Hogg, Allen T. Craig, *Introduction to Mathematical Statistics*, Prentice Hall, Upper Saddle River, NJ, 1995.
7. Sheldon Ross, *A First Course in Probability*, Prentice Hall, Upper Saddle River, NJ, 2002.

The Least-Squares Property of the Lanczos Derivative

NATHANIAL BURCH
PAUL E. FISHBACK
Grand Valley State University
Allendale, MI 49401
fishbacp@gvsu.edu

RUSSELL GORDON
Whitman College
Walla Walla, WA 99362
gordon@whitman.edu

In memory of Daniel Kopel.

Everyone knows that the derivative of a function at a point measures the slope, or instantaneous rate of change, of the function at that point; this follows from the usual difference-quotient definition. However, various alternate formulations of the derivative have proved valuable. Many of these are discussed in Bruckner's text [**1**].

The *Lanczos derivative* is one alternate definition not mentioned by Bruckner, which has been a focus of recent attention [**3, 5, 10**]. Attributed to the Hungarian-born mathematician Cornelius Lanczos (pronounced Lan'tsosh) [**7**], this derivative is actually defined in terms of a definite integral! The Lanczos derivative of a function f at x is given by

$$f_L'(x) = \lim_{h \to 0^+} \frac{3}{2h^3} \int_{-h}^{h} f(x+t)t\,dt, \tag{1}$$

provided the limit exists.

Groetsch [**3**] showed that (1) is a proper extension of the usual derivative. For instance, f need only possess a left derivative $f_-'(x)$ and a right derivative $f_+'(x)$ at x in order to be Lanczos differentiable at x. In fact, when these two derivatives exist,

$$f_L'(x) = \frac{f_-'(x) + f_+'(x)}{2}. \tag{2}$$

For example, if $f(x) = |x|$, then $f_L'(0) = 0$. Groetsch also investigated the Lanczos derivative from the perspective of numerical analysis. Shen [**10**] related the Lanczos derivative to an expectation operator in statistics, while Hicks and Liebrock [**5**] considered some computational aspects of the Lanczos derivative that are related to astrophysics.

Although Lanczos's original motivation arose from the need to differentiate an empirically defined function—a standard question in numerical analysis—we develop it through a surprising connection to linear regression. We sample the function uniformly in a small neighborhood of a point and use the method of least squares to find the slope of the line that best fits that data. Broadly speaking, the Lanczos derivative is the limit of this slope as the number of data points increases to infinity and the width of the sampling interval shrinks to zero.

We press this idea further to show that the Lanczos derivative can be viewed as a limiting covariance between two random variables relating the input and output of the function f. Finally, we develop higher-order Lanczos derivatives using higher-order regression, approximating f by polynomials of various degrees.

The Lanczos derivative via least squares

To connect the Lanczos derivative to linear regression, fix x and assume f is defined in some open interval about x. Thus, for $h > 0$ and n sufficiently large, $f(x+t)$, when viewed as a function of t, is defined at the $2n+1$ inputs

$$t_k = \frac{k}{n}h, \quad \text{where } -n \le k \le n. \tag{3}$$

Let $y(t) = m(h,n)t + b(h,n)$ denote the best linear approximation to $f(x+t)$ near $t = 0$, in the sense of least squares, using the data points $(t_k, f(x+t_k))$.

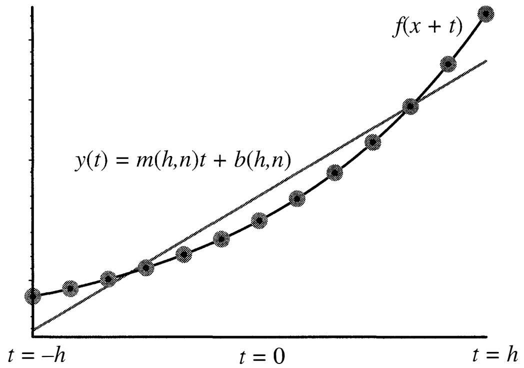


Figure 1 The least-squares regression line using the data points $(t_k, f(x+t_k))$

The standard recipes for the slope and intercept of the best linear approximation simplify to

$$m(h,n) = \frac{3\sum_{k=1}^{n}\left(f\left(x+\frac{k}{n}h\right) - f\left(x-\frac{k}{n}h\right)\right)k}{(2n^2+3n+1)h} \tag{4}$$

$$= \frac{3n^2\sum_{k=-n}^{n} f\left(x+\frac{k}{n}h\right)\left(\frac{k}{n}h\right)\frac{h}{n}}{(2n^2+3n+1)h^3} \tag{5}$$

and

$$b(h,n) = \frac{n\sum_{k=-n}^{n} f\left(x+\frac{k}{n}h\right)\frac{h}{n}}{(2n+1)h}. \tag{6}$$

Thus, if f is a function represented in tabular form, whose $2n+1$ inputs are spaced h/n units apart, then (4) gives an approximation to $f'(x)$. Furthermore, if $f'(x)$ exists, then an application of L'Hôpital's Rule followed by algebraic simplification reveals that $\lim_{h\to 0^+} m(h,n) = f'(x)$ for each fixed value of n. However, if instead of fixing n and letting h tend to 0, we let n tend to infinity first, (5) and (6) become

$$m(h) = \frac{3}{2h^3}\int_{-h}^{h} f(x+t)t\,dt \quad \text{and} \quad b(h) = \frac{1}{2h}\int_{-h}^{h} f(x+t)\,dt. \tag{7}$$

We thus see that $f'_L(x) = \lim_{h\to 0^+} m(h)$. In other words, the Lanczos derivative of f at x is the limiting slope of the least-squares lines that approximate f in a neighborhood of x.

Before proceeding, we pause to consider the relationship between the Lanczos derivative and the symmetric derivative. The symmetric derivative is defined by

$$f'_s(x) = \lim_{h\to 0^+} \frac{f(x+h) - f(x-h)}{2h},$$

provided that the limit exists. Note that the difference quotient that appears in this definition is simply $m(h, 1)$. It is not difficult to prove that $f'_s(x) = (f'_-(x) + f'_+(x))/2$, assuming that these one-sided derivatives exist. The converse is false as indicated by the function S defined by $S(x) = x\sin(1/x)$ for $x \neq 0$ and $S(0) = 0$. This function has a symmetric derivative at 0 but does not have one-sided derivatives at 0.

For a continuous function f, it is easy to prove that f has a Lanczos derivative at a point x whenever it has a symmetric derivative at x. Using L'Hôpital's Rule and the Fundamental Theorem of Calculus, we find that

$$\begin{aligned} f'_L(x) &= \lim_{h\to 0^+} \frac{\int_{-h}^{h} f(x+t)t\,dt}{2h^3/3} \\ &= \lim_{h\to 0^+} \frac{f(x+h)h - f(x-h)(-h)(-1)}{2h^2} = f'_s(x). \end{aligned}$$

We will see in the next section that a continuous function may have a Lanczos derivative at x even when it fails to have a symmetric derivative at x.

Interestingly, a second route to (1) avoids any discretization. This method utilizes the observation that the Lanczos derivative is in fact a symmetric form of the *least-squares derivative* developed by Kopel and Schramm [**6**]. Let m and b denote the quantities that minimize the mean square error

$$E(m, b) = \int_{-h}^{h} \left(f(x+t) - (mt+b)\right)^2 dt.$$

Setting to zero the partial derivatives of E with respect to m and b, integrating where possible, and solving the resulting system of equations for m and b in terms of f, x, and h also yields (7).

As we shall see, both methods for deriving the Lanczos derivative give insight into its meaning. The discrete method yields a statistical interpretation of the Lanczos derivative. The integral method provides insights into connections between the Lanczos derivative and Fourier analysis.

The calculus of the Lanczos derivative To what extent do familiar results from calculus remain valid when the usual derivative is replaced by the Lanczos derivative? Since integration is a linear operation, the Lanczos derivative enjoys the same linearity properties as the usual derivative. But because the Lanczos derivative allows us to differentiate many functions not differentiable in the usual sense, we should not be surprised to discover that other properties of the usual derivative are not shared by its Lanczos derivative counterpart.

For example, by considering the absolute value function on the interval $[-1, 2]$, or any closed interval containing the origin in its interior but not at the center of the interval, we observe that the Mean Value Theorem fails to hold when the usual derivative is replaced by the Lanczos derivative.

Even simpler calculus results fail to hold for the Lanczos derivative. Because the formula for the Lanczos derivative involves a Riemann integral, which is independent of the value of the function f at any particular point, it is trivial to construct examples of pairs of functions that fail to satisfy the product or chain rule. A more difficult assignment asks us to construct a counterexample involving only continuous functions. To accomplish this task, we introduce the *oscillating root function*:

$$f(t) = \begin{cases} (-1)^n \cdot \sqrt{t} \cdot \phi(t) & \text{if } t \in \left(\dfrac{1}{n+1}, \dfrac{1}{n}\right] \text{ for some } n \in \mathbb{N}, \\ 0 & \text{otherwise.} \end{cases}$$

In this definition, ϕ denotes a nonnegative, continuous, piecewise linear function that is equal to one most of the time, which we introduce to make f continuous. More specifically, we assume ϕ is bounded by one, equal to zero outside [0, 1], satisfies $\phi(1/n) = 0$ for all n in $\mathbb{N}$ and has the property that $\int_{1/(n+1)}^{1/n} (1 - \phi(t))\, dt$ is very small. Assuming this integral is less than $1/n^5$ will suffice for our purposes.

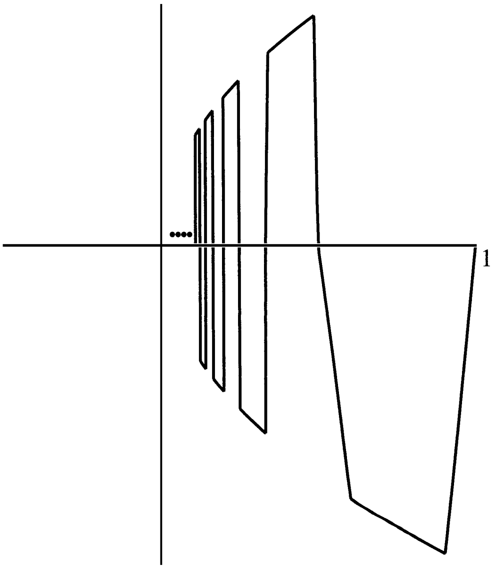


Figure 2 The oscillating root function

Using properties of alternating series, we can show that $f_L'(0)$ exists, and thus $2f(0)f_L'(0)$ is zero. However, $(f^2)_L'(0) = 1/2$. The details may be found posted at the MAGAZINE website. This example then establishes that neither the product nor the chain rules hold in general for the Lanczos derivative, even under the assumption of continuity. Furthermore, the fact that $(f^2)_L'(0)$ exists but the symmetric derivative of f^2 at the origin does not, illustrates how the Lanczos derivative differs from this other derivative extension.

The uniform random variable, f_L', and correlation

Based upon the results of the preceding section, we might be discouraged that the Lanczos derivative has little more to offer us. However, further investigating reveals that this is not the case and establishes deeper connections between the Lanczos derivative and probability and statistics.

Suppose X denotes a continuous random variable that is uniformly distributed over the interval $[-1, 1]$. Then the probability density function for X is

$$p(t) = \begin{cases} 1/2 & \text{if } t \in [-1, 1], \\ 0 & \text{otherwise.} \end{cases}$$

Its expected value, or mean, is $E(X) = \int_{\mathbb{R}} tp(t)\,dt = 0$, and its variance, or measure of spread, is given by

$$\sigma_X^2 = E(X^2) - \big(E(X)\big)^2 = \int_{\mathbb{R}} t^2 p(t)\,dt - 0 = \frac{1}{3}.$$

We observe that for fixed x and h, the formulas hX and $f(x + hX)$ also define random variables. The first has expected value $E(hX) = 0$ and variance $\sigma_{hX}^2 = h^2/3$, whereas the second has expected value

$$E\big(f(x + hX)\big) = \int_{\mathbb{R}} f(x + ht)p(t)\,dt = \frac{1}{2}\int_{-1}^{1} f(x + ht)\,dt$$

and variance

$$\begin{aligned}
\sigma_{f(x+hX)}^2 &= E\big(f(x + hX)^2\big) - E\big(f(x + hX)\big)^2 \\
&= \frac{1}{2}\int_{-1}^{1} f(x + ht)^2\,dt - \left(\frac{1}{2}\int_{-1}^{1} f(x + ht)\,dt\right)^2 \\
&= \frac{1}{2h}\int_{-h}^{h} f(x + t)^2\,dt - \left(\frac{1}{2h}\int_{-h}^{h} f(x + t)\,dt\right)^2.
\end{aligned}$$

The Cauchy-Schwarz inequality guarantees that $\sigma_{f(x+hX)}^2$ is strictly positive if we assume that f is not constant on $[x - h, x + h]$ [**8**].

To any pair of random variables, one may associate the covariance, a general measure of the correlation between the two random variables that describes the extent to which a change in the one variable affects the other. For the random variables hX and $f(x + hX)$, this covariance may be computed using their means as follows [**8**]:

$$\begin{aligned}
\text{Cov}\,\big(f(x + hX), hX\big) &= E\Big(\Big(f(x + hX) - E\big(f(x + hX)\big)\Big)\big(hX - E(hX)\big)\Big) \\
&= \frac{1}{2}\int_{-1}^{1} \Big(f(x + ht) - E\big(f(x + hX)\big)\Big)ht\,dt \\
&= \frac{1}{2h}\int_{-h}^{h} f(x + t)t\,dt.
\end{aligned}$$

Thus

$$m(h) = \frac{\text{Cov}(f(x + hX), hX)}{\sigma_{hX}^2}, \tag{8}$$

a probabilistic interpretation of $m(h)$ that we will establish is analogous to the difference quotient used to define the usual derivative.

Indeed, if we had instead assumed that X was a discrete Bernoulli random variable with probability density function

$$P(X = 1) = P(X = 0) = \frac{1}{2},$$

then elementary calculations yield $\sigma^2_{hX} = h^2/4$. In this case, we find that

$$\frac{\operatorname{Cov}(f(x+hX), hX))}{\sigma^2_{hX}} = \frac{f(x+h) - f(x)}{h},$$

the usual difference quotient.

Intimately connected to the covariance and important for problems related to linear regression is the correlation coefficient of hX and $f(x+hX)$. This correlation coefficient, which lies between -1 and 1, measures the degree to which two random variables satisfy a linear relation. The extreme cases are usually called *perfect negative* and *perfect positive* linear correlation. This correlation coefficient may be computed directly from the covariance and variances and is given by

$$\begin{aligned} r_x(h) &= \frac{\operatorname{Cov}(f(x+hX), hX)}{\sqrt{\sigma^2_{hX}\sigma^2_{f(x+hX)}}} = m(h)\sqrt{\frac{\sigma^2_{hX}}{\sigma^2_{f(x+hX)}}} \\ &= \frac{m(h)}{\sqrt{\frac{3}{h^2}\left(\frac{1}{2h}\int_{-h}^{h} f(x+t)^2\,dt - \left(\frac{1}{2h}\int_{-h}^{h} f(x+t)\,dt\right)^2\right)}}. \end{aligned} \tag{9}$$

Of particular interest is what happens to this correlation coefficient $r_x(h)$ as $h \to 0^+$. Given the fact that the usual derivative leads to a tangent line that is useful for estimation purposes, the following result should not come as a surprise. It states that functions with nonzero derivatives correspond to situations of perfect *local correlation*, that is, perfect *local linearity*, and that the absence of such correlation at an input x indicates the presence of a critical value there.

THEOREM 1. *Suppose that f is differentiable at x and $f'(x) \neq 0$. Then*

$$\lim_{h\to 0^+} r_x(h) = f'(x)/\left|f'(x)\right|.$$

Proof. By a simple change of variable, it suffices to assume that $x = 0$ and $f(0) = 0$. In light of (9) and the fact that $\lim_{h\to 0^+} m(h) = f'(0)$, the conclusion follows if we prove that

$$\lim_{h\to 0^+} \frac{3}{2h^3}\int_{-h}^{h} f(t)^2\,dt = f'(0)^2 \quad\text{and}\quad \lim_{h\to 0^+} \frac{1}{h^2}\int_{-h}^{h} f(t)\,dt = 0. \tag{10}$$

Since $f'(0)$ exists, the function z defined by

$$z(t) = \begin{cases} f(t)/t - f'(0) & \text{if } t \neq 0, \\ 0 & \text{if } t = 0, \end{cases}$$

is continuous at $t = 0$. Define $M_h = \sup\{|z(t)| : -h \le t \le h\}$ for each sufficiently small $h > 0$. These numbers are guaranteed to exist since z is locally bounded at the origin. In addition, the continuity of z at the origin reveals that $\lim_{h\to 0^+} M_h = 0$. We will establish the first limit in (10); the second limit is easier. Note that

$$\begin{aligned} \left|\frac{3}{2h^3}\int_{-h}^{h} f(t)^2\,dt - f'(0)^2\right| &= \left|\frac{3}{2h^3}\int_{-h}^{h} \left(tf'(0) + tz(t)\right)^2 dt - f'(0)^2\right| \\ &= \left|\frac{3}{2h^3}\int_{-h}^{h} \left(2f'(0)t^2 z(t) + t^2 z(t)^2\right) dt\right| \end{aligned}$$

$$\leq \frac{3}{2h^3}\left(\int_{-h}^{h} \left|2f'(0)t^2 z(t)\right| dt + \int_{-h}^{h} \left|t^2 z(t)^2\right| dt\right)$$

$$\leq \frac{3|f'(0)|M_h}{h^3}\int_{-h}^{h} t^2\,dt + \frac{3M_h^2}{2h^3}\int_{-h}^{h} t^2\,dt$$

$$\leq 2|f'(0)|M_h + M_h^2.$$

Since this last quantity approaches 0 as h does, the proof is complete. ■

When $f'(x)$ is zero or does not exist, a wide variety of other outcomes can occur. For example if f is any even function, then $\lim_{h\to 0^+} r_0(h) = 0$. Also, if $f(x) = x^n$, where n is a positive odd integer, $\lim_{h\to 0^+} r_0(h) = \sqrt{6n+3}/(n+2)$. This quantity decays to zero as n grows and equals 1 only if $n = 1$. This is not surprising, if one recalls from statistics that the correlation coefficient of constant data is zero by definition and considers how the graph of $f(x) = x^n$ becomes flatter in a small neighborhood of the origin as n increases.

Higher-order Lanczos derivatives

One might think that constructing the second-order Lanczos derivative would involve taking the Lanczos derivative of f_L'. However, such an approach would not remain faithful to the least-squares property of the Lanczos derivative. Instead, we construct the second-order Lanczos derivative by performing a quadratic regression using the data points from (3). Let us call the closest-fitting quadratic polynomial

$$y(t) = q(h,n)t^2 + m(h,n)t + b(h,n).$$

Then, by analogy, $f_L^{(2)}(x)$ should have something to do with the second derivative of $y(t)$. Specifically, we expect $f_L^{(2)}(x)$ to be $\lim_{h\to 0^+} 2q(h)$, where $q(h) = \lim_{n\to\infty} q(h,n)$.

This approach is equivalent to determining $q = q(h)$ by minimizing the function of three variables,

$$\int_{-h}^{h} \left(f(x+t) - \left(qt^2 + mt + b\right)\right)^2 dt, \tag{11}$$

which leads to the formula

$$q(h) = \frac{45}{8h^5}\int_{-h}^{h} f(x+t)\left(t^2 - \frac{h^2}{3}\right) dt.$$

We therefore define

$$f_L^{(2)}(x) = \lim_{h\to 0^+} \frac{45}{4h^5}\int_{-h}^{h} f(x+t)\left(t^2 - \frac{h^2}{3}\right) dt.$$

The example $f(x) = \operatorname{sgn}(x)\sqrt{|x|}$ at $x = 0$ illustrates that this limit can exist independently of whether $f_L'(x)$ is defined. We shall subsequently show that $f_L^{(2)}(x)$ is a proper extension of the usual second derivative.

A closer look at (11) suggests that to construct the nth-degree Lanczos derivative, we need to determine the nth-degree polynomial Q_n that minimizes

$$\int_{-h}^{h} \left(f(x+t) - Q_n(t)\right)^2 dt.$$

After a change of variable and renaming the polynomial, this task is identical to finding the nth-degree polynomial Q_n that minimizes

$$\int_{-1}^{1} \left(f(x+ht) - Q_n(t)\right)^2 dt. \tag{12}$$

Some geometric imagery is helpful here: The set of nth-degree polynomials is a linear subspace of the vector space of square-integrable functions on the interval $[-1, 1]$; minimizing (12) amounts to finding the polynomial in this subspace that is closest to $f(x+ht)$, and this is done by a method of projection. The celebrated *Legendre polynomials* provide one convenient means of accomplishing this task. They are given by

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

with

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

This latter characterization of P_n is sometimes referred to as Rodrigues' Formula [**11**]. We minimize (12) by attempting to express $f(x+ht)$ in terms of these Legendre polynomials.

A well-known result from the theory of orthogonal polynomials states: If f square-integrable on the interval $[x-h, x+h]$ for some fixed $x \in \mathbb{R}$ and $h > 0$, then the nth-degree polynomial Q_n that best approximates $f(x+ht)$ on $[-1, 1]$ is given by

$$Q_n(t) = \sum_{k=0}^{n} a_k(h) P_k(t),$$

where

$$a_k(h) = \frac{2k+1}{2}\int_{-1}^{1} f(x+ht) P_k(t)\, dt.$$

(This quantity is called the *kth Legendre coefficient* [**2, 9, 11**].) By "best approximates," we mean that the integral (12) is minimized.

Observe that the minimizing polynomial $Q_1(t)$ is given by

$$Q_1(t) = \frac{1}{2}\int_{-1}^{1} f(x+ht)\, dt + \left(\frac{3}{2}\int_{-1}^{1} f(x+ht)\, t\, dt\right) t.$$

After a change of variables, we recognize the first integral in this sum as $b(h)$. The coefficient of t given by the second integral, after the same change of variable and division by h, yields $m(h)$.

The coefficient of t^2 in $Q_2(t)$ arises from $a_2(h)P_2(t) = a_2(h)\frac{1}{2}(3t^2 - 1)$ and simplifies to

$$\frac{15}{4}\int_{-1}^{1} f(x+ht)\, P_2(t)\, dt.$$

After performing a change in variable and dividing by h^2, we see that the resulting expression coincides with $q(h)$.

This result motivates a possible approach for constructing even higher-order derivative extensions. If $f(x+ht)$ possesses an nth-degree Taylor polynomial about x in the variable t, then the coefficient of t^n is $h^n f^{(n)}(x)/n!$. On the other hand, the corresponding coefficient in the Legendre expansion $Q_n(t)$ is given by the leading coefficient of $a_n(h)P_n(t)$. Because the leading coefficient of P_n is $(2n)!/(2^n(n!)^2)$ and because both Q_n and the Taylor polynomial approximate f near x when this latter polynomial exists, we are led to define the nth-order Lanczos derivative as

$$\begin{aligned} f_L^{(n)}(x) &= \lim_{h\to 0^+} \frac{n!}{h^n} \cdot \text{Coefficient of } t^n \text{ in } Q_n(t) \\ &= \lim_{h\to 0^+} \frac{n!}{h^n} \left(\frac{2n+1}{2} \int_{-1}^{1} f(x+ht)P_n(t)\,dt \right) \cdot \frac{(2n)!}{2^n(n!)^2} \\ &= \lim_{h\to 0^+} \frac{(2n+1)!}{2^{n+1}h^n n!} \int_{-1}^{1} f(x+ht)P_n(t)\,dt. \end{aligned} \tag{13}$$

Evaluating the polynomials P_1 and P_2 and performing a change of variables in the integral, we see that (13) generalizes our previously-obtained first- and second-order Lanczos derivatives.

With this formula in hand, we can now establish a higher-order version of (2), thus demonstrating that $f_L^{(n)}$ generalizes the usual nth-order derivative.

THEOREM 2. *Suppose that f is $(n-1)$-times continuously differentiable in some neighborhood of x and that $f^{(n)}$ exists and is continuous in a neighborhood of x, except possibly at x itself, where the limits of both the left and right nth derivatives, $f_-^{(n)}(x) = \lim_{h\to 0^+} f^{(n)}(x-h)$ and $f_+^{(n)}(x) = \lim_{h\to 0^+} f^{(n)}(x+h)$ exist. Then the nth-order Lanczos derivative exists at x and*

$$f_L^{(n)}(x) = \frac{f_-^{(n)}(x) + f_+^{(n)}(x)}{2}.$$

In particular, if $f^{(n)}(x)$ exists, then $f_L^{(n)}(x) = f^{(n)}(x)$.

Proof. As a first step in the proof, we apply Rodrigues' formula:

$$\lim_{h\to 0^+} \frac{1}{h^n} \int_{-1}^{1} f(x+ht)P_n(t)\,dt = \lim_{h\to 0^+} \frac{1}{2^n n!} \cdot \frac{1}{h^n} \int_{-1}^{1} f(x+ht) \frac{d^n}{dt^n}(t^2-1)^n\,dt.$$

The smoothness conditions on f justify integrating by parts n times to reduce this limit to

$$\lim_{h\to 0^+} \frac{1}{2^n n!} \int_{-1}^{1} f^{(n)}(x+ht)(1-t^2)^n\,dt.$$

We now use the left and right continuity of $f^{(n)}$ on the respective intervals $[x-\delta, x]$ and $[x, x+\delta]$ for some $\delta > 0$ to see that

$$\begin{aligned} &\lim_{h\to 0^+} \int_{-1}^{1} f^{(n)}(x+ht)(1-t^2)^n\,dt \\ &= \lim_{h\to 0^+} \left(\int_{-1}^{0} f^{(n)}(x+ht)(1-t^2)^n\,dt + \int_{0}^{1} f^{(n)}(x+ht)(1-t^2)^n\,dt \right) \\ &= f_-^{(n)}(x) \int_{-1}^{0} (1-t^2)^n\,dt + f_+^{(n)}(x) \int_{0}^{1} (1-t^2)^n\,dt \\ &= \left(f_-^{(n)}(x) + f_+^{(n)}(x)\right) \int_{0}^{1} (1-t^2)^n\,dt. \end{aligned}$$

The reduction formula for sine yields

$$\int_0^1 (1-t^2)^n\,dt = \int_0^{\pi/2} \sin^{2n+1}\theta\,d\theta = \frac{4^n(n!)^2}{(2n+1)!}.$$

Putting all of this information together, we find that

$$\begin{aligned} f_L^{(n)}(x) &= \lim_{h\to 0^+} \frac{(2n+1)!}{2^{n+1}n!}\cdot\frac{1}{h^n}\int_{-1}^{1} f(x+ht)P_n(t)\,dt \\ &= \frac{(2n+1)!}{2^{n+1}n!}\cdot\frac{1}{2^n n!}\cdot\left(f_-^{(n)}(x)+f_+^{(n)}(x)\right)\cdot\frac{4^n(n!)^2}{(2n+1)!} \\ &= \frac{f_-^{(n)}(x)+f_+^{(n)}(x)}{2}, \end{aligned}$$

the value given in the statement of the theorem. ∎

The central result of Theorem 2 has been noted before. Indeed, after proving it ourselves, we discovered a paper by Haslam-Jones [**4**], who considers a large collection of generalized derivatives that involve integrals. Although he does not explicitly mention the Lanczos derivative, his results imply that the nth-order Lanczos derivative $f_L^{(n)}(x)$ extends the usual nth-order derivative. In fact, his work also establishes the weaker condition that $f_L^{(n)}(x)$ exists whenever f possesses what is referred to as an nth-order *Peano derivative* at x. The advantage of our approach is that it gives a least-squares interpretation to the higher-order Lanczos derivatives.

Discussion The fact that the the Lanczos derivative arises from the solution to a simple regression problem leads to both probabilistic and Fourier interpretations of this quantity. Namely (1) may be rewritten using either (8) or (13) as

$$\begin{aligned} f_L'(x) &= \lim_{h\to 0^+} \frac{\operatorname{Cov}(f(x+hX),hX)}{\sigma_{hX}^2} \\ &= \lim_{h\to 0^+} \frac{3}{2h}\int_{-1}^{1} f(x+ht)P_1(t)\,dt. \end{aligned}$$

These two interpretations may leave the reader wondering about the connection between the uniform random variable X and Legendre polynomials. To those familiar with Legendre polynomials, such a result should not come as a surprise. Namely, one means of constructing the sequence of Legendre polynomials $\{P_n\}$ is to start with the standard basis $\{1, t, t^2, \ldots\}$ and perform the Gram-Schmidt process on this family using a *weight function* w and the inner product

$$\langle f, g\rangle = \int_{\mathbb{R}} f(t)g(t)w(t)\,dt.$$

Different choices of w give rise to different families of orthogonal polynomials, the Legendre polynomials arising in the simple case of a constant weight [**11**]

$$w(t) = \begin{cases} 1 & \text{if } t\in[-1,1], \\ 0 & \text{otherwise.} \end{cases}$$

We note that, when appropriately scaled, w is nothing more than the probability density function of X.

Acknowledgment. The first two authors were supported by the Grand Valley State University Student Summer Scholars Program.

REFERENCES

1. A. M. Bruckner, *Differentiation of Real Functions*, The American Mathematical Society, Providence, 1994.
2. R. Burden and J. Faires, *Numerical Analysis*, Brooks/Cole, Pacific Grove, 2001.
3. C. Groetsch, Lanczos' generalized derivative, *Amer. Math. Monthly* **105** (1998), 320–326.
4. U. S. Haslam-Jones, On a generalized derivative, *Quart. J. Math, Oxford Ser.* **2** (1953), 190–197.
5. D. L. Hicks and L. M. Liebrock, Lanczos' generalized derivative: insights and applications, *Appl. Math. and Comp.* **112** (2000), 63–73.
6. D. Kopel and M. Schramm, A new extension of the derivative, *Amer. Math. Monthly* **97** (1990), 230–233.
7. C. Lanczos, *Applied Analysis*, Prentice Hall, Englewood Cliffs, 1960.
8. S. Ross, *A First Course in Probability*, Prentice Hall, Englewood Cliffs, 2002.
9. G. Sansone, *Orthogonal Functions*, Interscience Publishers, New York, 1959.
10. J. Shen, On the generalized "Lanczos' generalized derivative," *Amer. Math. Monthly* **106** (1999), 766–768
11. G. Szegö, *Orthogonal Polynomials*, The American Mathematical Society, Providence, 1967.
12. E. T. Whittaker and G. N. Watson, *A Course in Modern Analysis*, Cambridge University Press, London, 1963.

Poem: Triangles

They are widely admired, and for good reason.
Like handsome shop windows or smartly dressed museum guides,
　they invite our warm inspection, and never disappoint.
They are tranquil on the page.
Elegant arrows pointing out toward unknown empty skies,
　the immaculate yard, protected by the three and perfect lines,
　perhaps the magic property of a wizard or a king,
　a stately yellow princess wearing alabaster rings.
A charming place to pause and rest.
They would not mind.
But look around, and ponder now the stubborn questions.
What goes on here?

Quite a bit, or so I'm told, by Euler and his friends,
　who spend their lives investigating geometric trends.
Bisectors perpendicular converging at a point,
　the altitudes will follow suit, a second point, and then,
　the medians all do the same, a third point like the rest.
These three locations, clean and true, fall on a single line,
　a perfect twig upon the lawn, immune to breath of air,
　undisturbed by petty jealousies, or by diamonds to compare.

This is true for every one of them. All triangles. All of them, I say.
And to know this is astonishing, miraculous, and fine.
It never fails to make me shed a tear, this Euler line.
My eyes are wide with admiration; it can make me feel this way.
So very very happy.
And so very very small.

—— Greg Tuleja
74 High Street
Southampton, MA 01073

NOTES

Honey, Where Should We Sit?

JOHN A. FROHLIGER
BRIAN HAHN
St. Norbert College
De Pere, WI 54115
john.frohliger@snc.edu

There are times when, in their haste to solve a particular problem, students (and their instructors) miss an opportunity to notice some interesting mathematics. For example, when calculus students are introduced to the derivatives of inverse trigonometric functions, they frequently run across a classic problem that goes something like this:

There is a 6-foot tall picture on a wall, 2 feet above your eye level. How far away should you sit (on the level floor) in order to maximize the vertical viewing angle θ? (See FIGURE 1.)

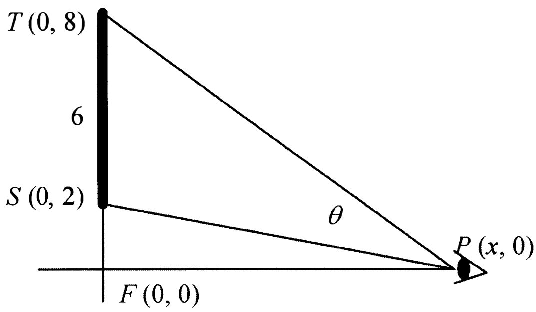


Figure 1 Find where θ is a maximum

This problem can be solved using the standard calculus technique for maximization. First, on the coordinate plane, we could set the top and bottom of the picture at $T(0, 8)$ and $S(0, 2)$, respectively. Then it is easy to show that if your eye is at a point $P(x, 0)$ on the positive x-axis, the viewing angle would be $\theta = \tan^{-1}(8/x) - \tan^{-1}(2/x)$. From the derivative,

$$\frac{d\theta}{dx} = \frac{6(16 - x^2)}{(x^2 + 8^2)(x^2 + 2^2)},$$

you can easily show that the only critical number for $x > 0$ occurs at $x = 4$. Finally, (the part that many students like to skip) the first or second derivative test can provide arguments that θ must be an absolute maximum at $P(4, 0)$.

At this point, many calculus students declare that the greatest viewing angle occurs 4 feet from the wall, express some relief and gratitude for having solved the problem, and move on to the next assignment. In doing so, unfortunately, they miss some fascinating geometry. Notice that, if we let F represent the origin, then at the point P of maximum θ, $PF/FS = 2 = TF/PF$ (FIGURE 2). This makes $\triangle PFS$ and

$\triangle TFP$ *similar* right triangles. Thus, the viewing angle is largest at the point P where $\angle FPS \cong \angle FTP$!

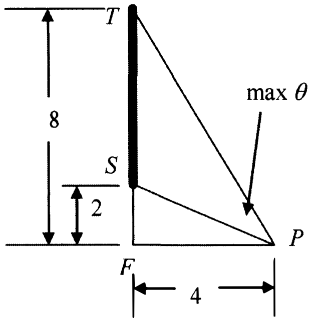


Figure 2 Similar triangles $\triangle PFS$ and $\triangle TFP$

So now a mathematician starts to wonder: is this result just a coincidence (if there is such a thing as a mathematical coincidence)? What if we change the y-coordinates of S and T? How about if, instead of being level, the floor were slanted and P were on a line $y = mx$? (Stewart gives a numerical approach to a variation of this problem [**1**, p. 478].)

Curiously enough, even in these cases the answer is that the viewing angle is a maximum where $\angle FPS \cong \angle FTP$. (This could be a good assignment for a bright student.) In fact, we can generalize even further and consider the case where the floor is curved rather than straight. The result is the following:

THEOREM. *Let $S(0, a)$ and $T(0, b)$ be points on the y-axis with $a < b$, and let $y = f(x)$ be a continuous function on* $[0, \infty)$ *and, without loss of generality, $f(0) < a$. Then there is point $P(x, f(x))$, $x > 0$, on the graph of f such that the measure of $\angle TPS$ is a maximum. Furthermore, if f is differentiable at P, then $\angle FPS \cong \angle FTP$, where F is the point where the tangent to $f(x)$ at P intersects the y-axis* (FIGURE 3).

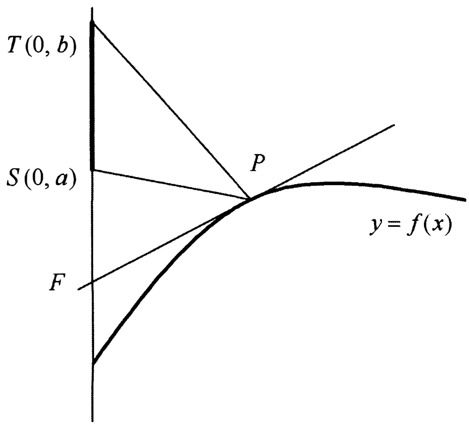


Figure 3 The generalized case

Note: In the original problem P is on the x-axis $y = 0$, and in the variation P is on the line $y = mx$. Both times, the point F is given as the origin. This notation is consistent with our generalized property since, in those cases, the tangent line to the graph of $y = f(x)$, which is simply the graph itself, intersects the y-axis at $(0, 0)$. Also, when we refer to a maximum θ, or θ being maximized, we shall implicitly restrict ourselves to the domain $(0, \infty)$.

Proof. The property that $\angle FPS \cong \angle FTP$ at maximum θ can be proved using standard calculus. Suppose f is differentiable at the maximum angle. We will assume for the time being that a greatest θ exists. It is straightforward to show that, if point P has coordinates $(x,\ f(x))$, then $\angle TPS$ has measure

$$\theta = \tan^{-1}\left(\frac{b - f(x)}{x}\right) + \tan^{-1}\left(\frac{f(x) - a}{x}\right).$$

Differentiating and simplifying, we can see that

$$\frac{d\theta}{dx} = (a-b)\frac{\left[x^2 + (xf'(x))^2\right] - [a - (f(x) - xf'(x))][b - (f(x) - xf'(x))]}{[x^2 + (b - f(x))^2][x^2 + (a - f(x))^2]}.$$

Since the denominator involves products of sums of perfect squares, and since $f(0)$ is neither a nor b, we can see that the denominator is never zero; hence, $d\theta/dx$ is never undefined. It follows that at the maximum, the derivative must be zero. At this point then,

$$x^2 + (xf'(x))^2 = [a - (f(x) - xf'(x))]\,[b - (f(x) - xf'(x))]. \qquad (1)$$

All we need to do is interpret this in terms of lengths. The slope of the tangent to $f(x)$ at P is $f'(x)$. If we follow the tangent line back to the y-axis, we see that F has coordinates $(0,\ f(x) - xf'(x))$, as in FIGURE 4.

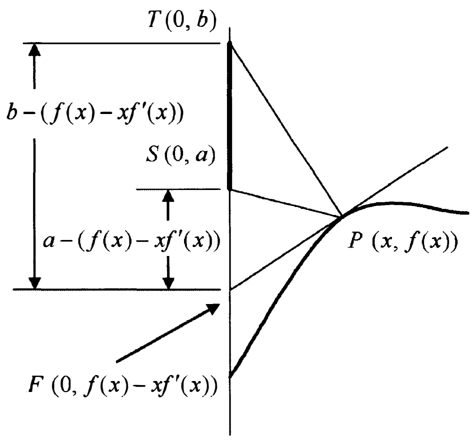


Figure 4 Where the tangent hits the y-axis

From (1), we see that $PF^2 = SF \cdot TF$, that is,

$$\frac{PF}{SF} = \frac{TF}{PF}.$$

Since they share a common angle and have two pairs of proportional sides, it follows that $\triangle SFP$ and $\triangle PFT$ are similar triangles. Therefore, we can conclude that $\angle FPS \cong \angle FTP$ when P is chosen to make $\angle TPS$ largest. ■

Geometric approach Now we turn to some more general questions: Assuming f is continuous, not necessarily differentiable, on $[0, \infty)$, are we guaranteed that there is a point P where the viewing angle is greatest? If there is such a point P, is it necessarily unique or might the maximum angle occur at more than one point on the graph? We can answer these questions by taking a different approach to the problem. Let's leave calculus and its potentially messy computations and turn instead to geometry (with just a pinch of topology).

Recall that, in a circle, the measure of an inscribed angle is one-half that of the intercepted arc [**3**]. A corollary of this property is that every inscribed angle that intercepts the same arc has the same measure. Conversely, given fixed points T and S and an angle θ, the set of all points Q on one side of $\overline{ST}$ satisfying $m(\angle SQT) = \theta$ is a portion of a circle passing through S and T.

Now let's return to our problem. Again, we let S and T represent the top and bottom of our picture. For a fixed positive measure c, consider the set of points Q on the right half-plane such that $m(\angle SQT) = c$. From our discussion above, we can easily see that this level curve is the right-hand portion of a circle passing through S and T (FIGURE 5).

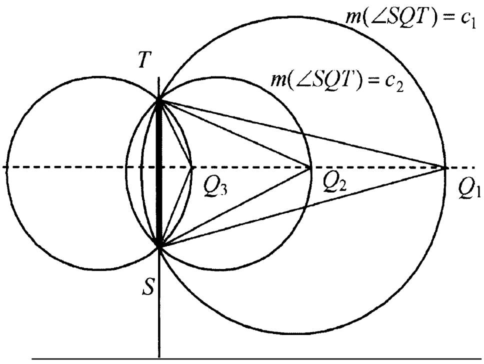


Figure 5 Level curves of constant angles

Moreover, the smaller the value of c, the farther the center of the circle is to the right. For instance, if Q_1, Q_2, and Q_3 are placed on the perpendicular bisector of $\overline{ST}$ as shown in FIGURE 5, it is easy to see that $m(\angle SQ_1T) < m(\angle SQ_2T) < m(\angle SQ_3T)$. Also notice that the regions bounded by $\overline{ST}$ and these circular curves are nested: If $0 < c_1 < c_2$, then the region bounded by $\overline{ST}$ and the curve $m(\angle SQT) = c_2$ is contained in the region bounded by $\overline{ST}$ and $m(\angle SQT) = c_1$.

Now we can answer the questions we posed earlier. Must there be a point P along the graph of $y = f(x)$ at which $m(\angle SPT)$ is a maximum? If so, where is P? The answer to the second question is that P occurs where $y = f(x)$ intersects the circular arc $m(\angle SQT) = c$ for the largest value of c, that is, the leftmost curve $m(\angle SQT) = c$ (FIGURE 6). It is probably obvious that there must be such a point; however, to be safe, we could turn to a little topology. (If this result is obvious, feel free to skip the next paragraph.)

Let G represent the graph of $y = f(x)$. For each positive c, let D_c be the closed bounded region in the right closed half-plane bounded by $\overline{ST}$ and the arc $m(\angle SQT) = c$. Then define G_c to be the intersection of G with D_c. Now consider the nonempty collection $A = \{G_c : G_c \neq \emptyset\}$ of nonempty intersections of G with the sets D_c. The continuity of f implies that G is closed; hence, each G_c is compact. Furthermore, since the D_cs are nested, it follows that the G_cs satisfy the finite intersection

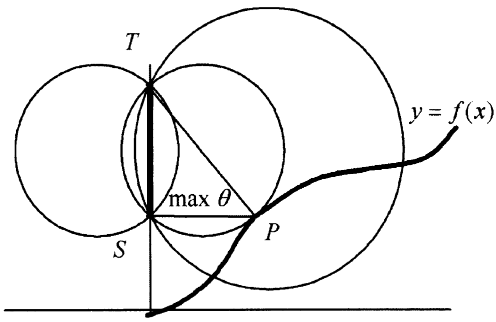


Figure 6 Where θ is maximized

property [2]. Therefore, $\bigcap_{G_c \in A} G_c \neq \emptyset$ and $m(\angle SPT)$ is a maximum at any point P in $\bigcap_{G_c \in A} G_c$.

We can see that this result is consistent with our earlier findings about similar triangles. If the tangent to the circle at P intersects the y-axis at F (FIGURE 7) then, since $\angle SPF$ and $\angle PTF$ intercept the same arc, they are congruent. Consequently, $\triangle SFP$and$\triangle PFT$ are similar triangles and $PF/SF = TF/PF$, as before.

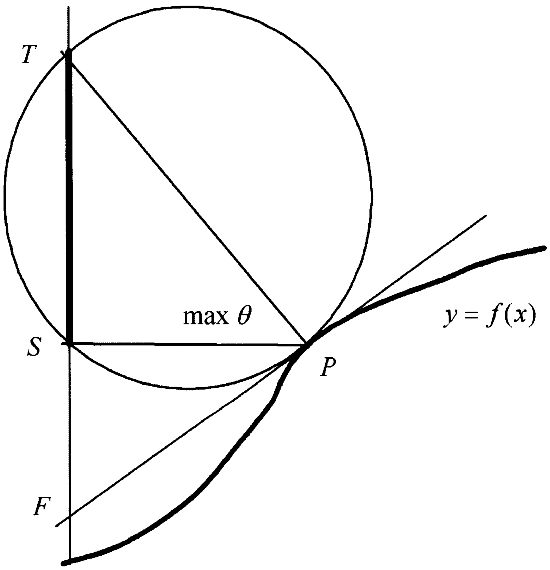


Figure 7 Similar triangles in the general case

This geometric approach allowed us to see, without ugly computations, that there must be a point P on G such that the viewing angle, $m(\angle SPT)$, is maximized. Furthermore, an easy construction allows us to show that, depending upon G, this point of greatest angle may occur at more than one point (FIGURE 8a). In fact, if G moves along a section of one such circular arc, there would be an infinite number of such points (FIGURE 8b).

We now address one final question: How do we construct such a point P? As we showed earlier, sometimes you can find P using possibly cumbersome calculus computations. In the special cases where the graph G is a line, however, we can use the geometry of the situation to physically construct the point of maximum angle using a compass and straightedge. In these situations the smallest circle through S and T that

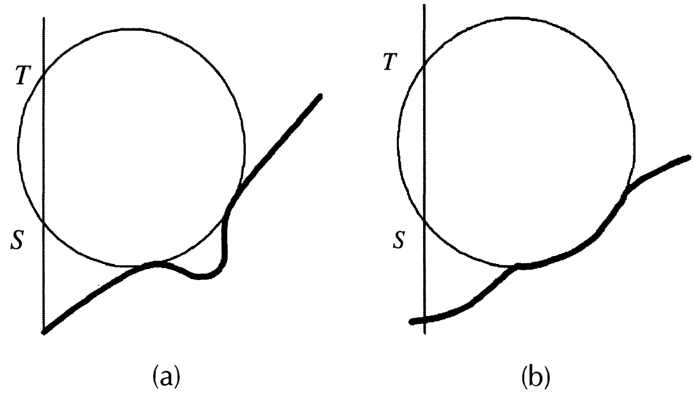


Figure 8 Cases where θ is maximized at multiple points

intersects G must be tangent to G at that point. Thus, all we need to do is find this tangent circle and determine the point P of tangency.

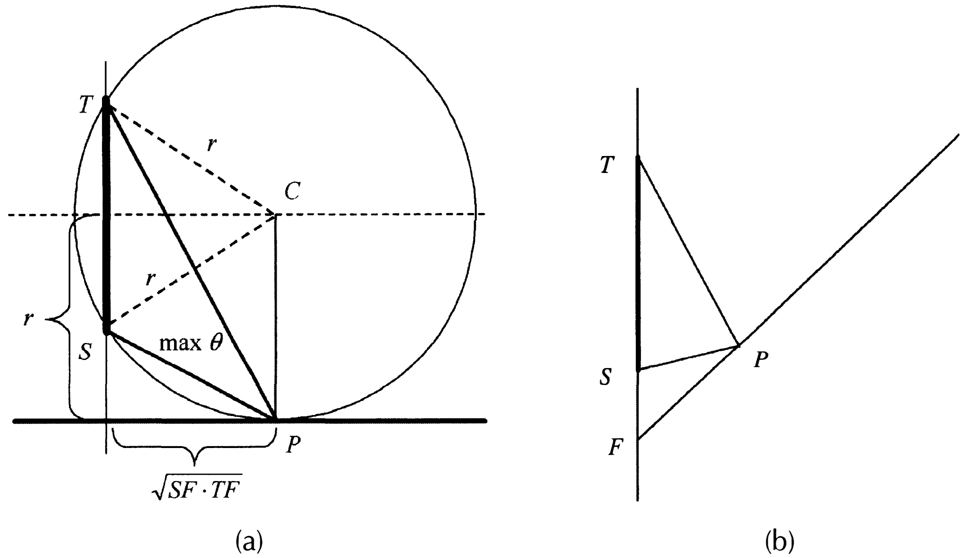


Figure 9 (a) Constructing the point of maximum θ (b) Constructing the slant line solution

This task is especially easy if G is a horizontal line (FIGURE 9a). In this situation, the one we started with, the smallest circle through S and T that intersects G must be tangent to G at that point. Thus, all we need to do is find this tangent circle and determine the point P of tangency. First we find the distance r from the perpendicular bisector of $\overline{ST}$ to G. Next we locate the point C on the right side of this perpendicular bisector that is r units from both S and T. The maximum angle then occurs at the foot P of the perpendicular from C to G. Notice that, from our previous discussion, $PF/SF = TF/PF$; hence, $PF = \sqrt{SF \cdot TF}$, so PF is the geometric mean of SF and TF.

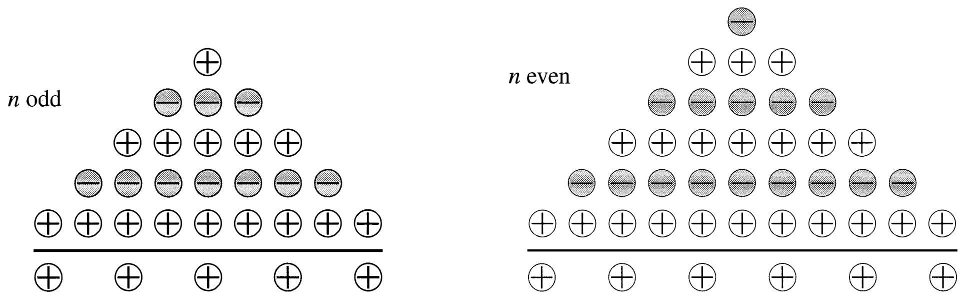
Now that we've constructed the solution for a horizontal line, the solution for the slant line situation becomes easy. At the point of greatest angle measure, we still have the similar triangles, so the distance from P to F is still $PF = \sqrt{SF \cdot TF}$. We constructed this distance in the horizontal line case. All we need to is to construct a circle with center F and radius $\sqrt{SF \cdot TF}$. The desired point P is the intersection of this circle and the slant line (FIGURE 9b).

REFERENCES

1. James Stewart, *Calculus*, 4th ed., Brooks/Cole, Pacific Grove, CA, 1999.
2. James R. Munkres, *Topology, A First Course*, Prentice-Hall, Englewood Cliffs, NJ, 1975.
3. James R. Smart, *Modern Geometries*, 5th ed., Brooks/Cole, Pacific Grove, CA, 1998.

Proof Without Words: Alternating Sums of Odd Numbers

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



—ARTHUR T. BENJAMIN
HARVEY MUDD COLLEGE
CLAREMONT, CA 91711

A Short Proof of Chebychev's Upper Bound

Kimberly Robertson
William Staton
University of Mississippi
University, MS 38677
mmstaton@olemiss.edu

Examining $\pi(n)$, the number of primes less than or equal to n, is surely one of the most fascinating projects in the long history of mathematics. In 1852, Chebychev **[3]** proved that there are constants A and B so that, for all natural numbers $n > 1$,

$$\frac{An}{\ln(n)} < \pi(n) < \frac{Bn}{\ln(n)}.$$

Later, in 1896, with arguments of analysis, the Prime Number Theorem was proved, showing that for n sufficiently large, A and B may be taken arbitrarily close to 1. Establishing the Prime Number Theorem is difficult and the proof is often omitted from texts in elementary number theory. Chebychev's arguments, by contrast, were elementary and wonderfully clever, using properties of the middle binomial coefficients $\binom{2n}{n}$. Our purpose here is to provide what we believe is a brief and elegant approach to Chebychev's upper bound, a proof accessible in its brevity to even the beginning number theory student. We were initially motivated by Bollobas' lovely English presentation [**2**] of Paul Erdős' proof of Bertrand's Postulate.

We begin with three simple lemmas, in which n always denotes a positive integer.

LEMMA. *For all* n, $n^{\pi(2n)-\pi(n)} \le \binom{2n}{n}$.

Proof. There are exactly $\pi(2n) - \pi(n)$ primes between n and $2n$. Each appears precisely once in the numerator of the factorial expression for $\binom{2n}{n}$ and none appears at all in the denominator. So each is a factor of $\binom{2n}{n}$, and of course each is bigger than n. ∎

LEMMA. *For all* $n \ge 8$, $\pi(2n) \le n - 2$.

Proof. By induction: $\pi(16) = 6 \le 8 - 2$. If $\pi(2k) \le k - 2$, then $\pi(2k+2) \le (k-2) + 1$ since $2k+2$ is certainly not prime. So $\pi(2k+2) \le (k+1) - 2$ and the lemma follows. ∎

LEMMA. *For all n, we have* $\binom{2n}{n} \le \frac{1}{2}(4^n)$.

Proof. Begin a proof by induction, by noting that, for $n = 1$, $\binom{2}{1} = 2 = \frac{1}{2}(4^1)$. Dividing $\binom{2n+2}{n+1}$ by $\binom{2n}{n}$ yields $\frac{4n+2}{n+1} < 4$. Hence, if $\binom{2n}{n} < \frac{1}{2}(4^n)$, then

$$\binom{2n+2}{n+1} = \frac{4n+2}{n+1}\binom{2n}{n} < 4\left(\frac{1}{2}\right)(4^n) = \frac{1}{2}(4^{n+1})$$

and the lemma follows. ∎

We will prove our theorem by a variant of induction, moving from k to both $2k$ and $2k-1$. That this scheme is adequate is easily seen by noting that the truth of the statement for all k up to 2^r then implies the truth of the statement for all k up to 2^{r+1}.

THEOREM. *For all* $n \ge 1$, $n^{\pi(n)} < 8^n$.

Proof. For $n \le 8$ this is clear since both the bases and the exponents compare in the desired way. Furthermore, since $16^{\pi(16)} = 16^6 = 8^8 < 8^9$, the statement holds for $9 \le n \le 16$. Now suppose $n \ge 8$ and $n^{\pi(n)} < 8^n$. Then, applying our various lemmas, we have $(2n-1)^{\pi(2n-1)} < (2n)^{\pi(2n)} = 2^{\pi(2n)}n^{\pi(2n)} = 2^{\pi(2n)}n^{\pi(2n)-\pi(n)}n^{\pi(n)} < 2^{n-2}\binom{2n}{n}8^n < (2^{n-2})(4^n/2)(8^n) = 8^{2n-1} < 8^{2n}$. Hence the inequality holds for $2n-1$ and $2n$ and the theorem follows. ∎

COROLLARY. (CHEBYCHEV, 1852) *There is a constant B so that for every positive integer* $n > 1$,

$$\pi(n) \le \frac{Bn}{\ln(n)}.$$

Proof. $n^{\pi(n)} < 8^n$. Taking natural logarithms yields $\pi(n)\ln(n) < n\ln(8)$ and the inequality is established with $B = \ln(8)$. ∎

We see the potential value of this proof as twofold. First, it appears cleaner and shorter than what is found in most texts. And our constant, ln(8), is modest compared to Sierpinski's 4 [**7**], Apostol's 6 [**1**], or the 32 ln(2) offered in earlier editions of Niven and Zuckerman [**6**]. LeVeque [**5**], Hardy and Wright [**4**], and the latest edition of Niven and Zuckerman [**6**] give no particular constant, merely proving that one exists. Chebychev [3] achieved a much smaller constant than ours, but with considerably more effort. We hope that our short proof will be found to have pedagogical value.

REFERENCES

1. Tom M. Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag, New York, 1976.
2. Bela Bollobas, Erdős first paper: A proof of Bertrand's Postulate, pre-print, 1998.
3. P.L. Chebychev, *Memoire sur les nombres premiers*, *J. Math. Pures et Appl.* **17**, (1852), 366–390.
4. G.H. Hardy and E.M. Wright, *An Introduction to the Theory of Numbers*, 5th Edition, Clarendon Press, Oxford, 1979.
5. William J. LeVeque, *Fundamentals of Number Theory*, Dover Publications, New York, 1977.
6. I. Niven and H.S. Zuckerman, *An Introduction to the Theory of Numbers*, 3rd, 4th, 5th Editions, Wiley, New York, 1972, 1980, 1991.
7. Waclaw Sierpinski, *Elementary Theory of Numbers*, PWN, Warsaw, 1964.

Recounting the Odds of an Even Derangement

ARTHUR T. BENJAMIN
Harvey Mudd College
Claremont, CA 91711
benjamin@hmc.edu

CURTIS T. BENNETT
Loyola Marymount University
Los Angeles, CA 90045
cbennett@lmu.edu

FLORENCE NEWBERGER
California State University
Long Beach, CA 90840-1001
fnewberg@csulb.edu

Odd as it may sound, when n exams are randomly returned to n students, the probability that no student receives his or her own exam is almost exactly $1/e$ (approximately 0.368), for all $n \geq 4$. We call a permutation with no fixed points, a *derangement*, and we let $D(n)$ denote the number of derangements of n elements. For $n \geq 1$, it can be shown that $D(n) = \sum_{k=0}^{n}(-1)^k n!/k!$, and hence the *odds* that a random permutation of n elements has no fixed points is $D(n)/n!$, which is within $1/(n+1)!$ of $1/e$ [**1**].

Permutations come in two varieties: even and odd. A permutation is even if it can be achieved by making an even number of swaps; otherwise it is odd. Thus, one might *even* be interested to know that if we let $E(n)$ and $O(n)$ respectively denote the number of even and odd derangements of n elements, then (oddly enough),

$$E(n) = \frac{D(n) + (n-1)(-1)^{n-1}}{2}$$

and

$$O(n) = \frac{D(n) - (n-1)(-1)^{n-1}}{2}.$$

The above formulas are an immediate consequence of the equation $E(n) + O(n) = D(n)$, which is obvious, and the following theorem, which is the focus of this note.

THEOREM. *For* $n \geq 1$,

$$E(n) - O(n) = (-1)^{n-1}(n-1). \tag{1}$$

Proof 1: Determining a Determinant The fastest way to derive equation (1), as is done in [**3**], is to compute a determinant. Recall that an n-by-n matrix $A = [a_{ij}]_{i,j=1}^{n}$ has determinant

$$\det(A) = \sum_{\pi \in S_n} a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)} \operatorname{sgn}(\pi), \tag{2}$$

where S_n is the set of all permutations of $\{1, \ldots, n\}$, $\operatorname{sgn}(\pi) = 1$ when π is even, and $\operatorname{sgn}(\pi) = -1$ when π is odd. Let A_n denote the n-by-n matrix whose nondiagonal entries are $a_{ij} = 1$ (for $i \neq j$), with zeroes on the diagonal. For example, when $n = 4$,

$$A_4 = J_4 - I_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

By (2), every permutation that is not a derangement will contribute 0 to the sum (since it uses at least one of the diagonal entries), every even derangement will contribute 1 to the sum, and every odd derangement will contribute -1 to the sum. Consequently, $\det(A_n) = E(n) - O(n)$. To see that $\det(A_n) = (-1)^{n-1}(n-1)$, observe that $A_n = J_n - I_n$, where J_n is the matrix of all ones and I_n is the identity matrix. Since J_n has rank one, zero is an eigenvalue of J_n, with multiplicity $n-1$, and its other eigenvalue is n (with an eigenvector of all 1s). Apply $J_n - I_n$ to the eigenvectors of J_n to find the eigenvalues of A_n: -1 with multiplicity $n-1$ and $n-1$ with multiplicity 1. Multiplying the eigenvalues gives us $\det(A_n) = (-1)^{n-1}(n-1)$, as desired. ∎

A 1996 Note in the MAGAZINE [**2**] gave *even odder* ways to determine the determinant of A_n.

Although the proof by determinants is quick, the form of (1) suggests that there should also exist an *almost* one-to-one correspondence between the set of even derangements and the set of odd derangements.

Proof 2: Involving an Involution Let D_n denote the set of derangements of $\{1, \ldots, n\}$, and let X_n be a set of $n-1$ *exceptional* derangements (that we specify later), each with sign $(-1)^{n-1}$. We exhibit a *sign reversing involution* on $D_n - X_n$. That is, letting $T_n = D_n - X_n$, we find an invertible function $f : T_n \to T_n$ such that for π in T_n, π and $f(\pi)$ have opposite signs, and $f(f(\pi)) = \pi$. In other words, except for the $n-1$ exceptional derangements, every even derangement "holds hands" with an odd derangement, and vice versa. From this, it immediately follows that $|E_n| - |O_n| = (-1)^{n-1}(n-1)$.

Before describing f, we establish some notation. We express each π in D_n as the product of k disjoint cycles $C_1, \ldots, C_k$ with respective lengths $m_1, \ldots, m_k$ for some

$k \geq 1$. We follow the convention that each cycle begins with its smallest element, and the cycles are listed from left to right in increasing order of the first element. In particular, $C_1 = (1\ a_2\ \cdots\ a_{m_1})$ and, if $k \geq 2$, C_2 begins with the smallest element that does not appear in C_1. Since π is a derangement on n elements, we must have $m_i \geq 2$ for all i, and $\sum_{i=1}^{k} m_i = n$. Finally, since a cycle of length m has sign $(-1)^{m-1}$, it follows that π has sign $(-1)^{\sum_{i=1}^{k}(m_i-1)} = (-1)^{n-k}$.

Let π be a derangement in D_n with first cycle $C_1 = (1\ a_2\ \cdots\ a_m)$ for some $m \geq 2$. We say that π has *extraction point* $e \geq 2$ if e is the smallest number in the set $\{2, \ldots, n\} - \{a_2\}$ for which C_1 does *not* end with the numbers of $\{2, \ldots, e\} - \{a_2\}$ written in decreasing order. Note that π will have extraction point $e = 2$ unless the number 2 appears as the second term or last term of C_1. We illustrate this definition with some pairs of examples from D_9. Notice that in each pair below, the number of cycles of π and π' differ by one, and the extraction point e occurs in the first cycle of π and is the leading element of the second cycle of π'.

$\pi = (1\,9\,7\,2\,8)(3\,6)(4\,5)$	and	$\pi' = (1\,9\,7)(2\,8)(3\,6)(4\,5)$	have $e = 2$.
$\pi = (1\,2\,9\,7\,3\,8\,5)(4\,6)$	and	$\pi' = (1\,2\,9\,7)(3\,8\,5)(4\,6)$	have $e = 3$.
$\pi = (1\,9\,7\,3\,8\,5\,2)(4\,6)$	and	$\pi' = (1\,9\,7\,2)(3\,8\,5)(4\,6)$	have $e = 3$.
$\pi = (1\,9\,4\,8\,5\,3\,2)(6\,7)$	and	$\pi' = (1\,9\,3\,2)(4\,8\,5)(6\,7)$	have $e = 4$.
$\pi = (1\,4\,9\,5\,8\,3\,2)(6\,7)$	and	$\pi' = (1\,4\,9\,3\,2)(5\,8)(6\,7)$	have $e = 5$.
$\pi = (1\,3\,8\,6\,9\,7\,5\,4\,2)$	and	$\pi' = (1\,3\,8\,5\,4\,2)(6\,9\,7)$	have $e = 6$.

Observe that every derangement π in D_n contains an extraction point unless π consists of a single cycle of the form $\pi = (1\ a_2\ Z)$, where Z is the ordered set $\{2, 3, \ldots, n-1, n\} - \{a_2\}$, written in decreasing order. For example, the 9-element derangement $(1\ 5\ 9\ 8\ 7\ 6\ 4\ 3\ 2)$ has no extraction point. Since a_2 can be any element of $\{2, \ldots, n\}$, there are exactly $n-1$ derangements of this type, all of which have sign $(-1)^{n-1}$. We let X_n denote the set of derangements of this form. Our problem reduces to finding a sign reversing involution f over $T_n = D_n - X_n$.

Suppose π in T_n has extraction point e. Then the first cycle C_1 of π ends with the (possibly empty) ordered subset Z consisting of the elements of $\{2, \ldots, e-1\} - \{a_2\}$ written in decreasing order. Our sign reversing involution $f : T_n \to T_n$ can then be succinctly described as follows:

$$(1\ a_2\ X\ e\ Y\ Z)\sigma \overset{f}{\longleftrightarrow} (1\ a_2\ X\ Z)(e\ Y)\sigma, \tag{3}$$

where X and Y are ordered subsets, Y is nonempty, and σ is the rest of the derangement π.

Notice that since the number of cycles of π and $f(\pi)$ differ by one, they must be of opposite signs. The derangements on the left side of (3) are those for which the extraction point e is in the first cycle. In this case, Y must be nonempty, since otherwise "$e\ Z$" would be a longer decreasing sequence and e would not be the extraction point. The derangements on the right side of (3) are those for which the extraction point e is not in the first cycle (and must therefore be the leading element of the second cycle). In this case, Y is nonempty since π is a derangement. Thus for any derangement π, the derangement $f(\pi)$ is also written in standard form, with the same extraction point e and with the same associated ordered subset Z. Another way to see that π and $f(\pi)$ have opposite signs is to notice that $f(\pi) = (xy)\pi$ (multiplying from left to right), where x is the last element of X ($x = a_2$ when X is empty), and y is the last element of Y. Either way, $f(f(\pi)) = \pi$, and f is a well-defined, sign-reversing involution, as desired. ∎

In summary, we have shown combinatorially that for all values of n, there are almost as many even derangements as odd derangements of n elements. Or to put it another way, when randomly choosing a derangement with at least five elements, the *odds* of having an even derangement are nearly *even.*

Acknowledgment. We are indebted to Don Rawlings for bringing this problem to our attention and we thank Magnhild Lien, Will Murray, and the referees for many helpful ideas.

REFERENCES

1. R. A. Brualdi, *Introductory Combinatorics*, 3rd ed., Prentice-Hall, New Jersey, 1999.
2. C. A. McCarthy and A. T. Benjamin, Determinants of the tournaments, this MAGAZINE, **69** (1996), 133–135.
3. C. D. Olds, Odd and even derangements, Solution E907, *Amer. Math. Monthly*, **57** (1950), 687–688.

Volumes of Generalized Unit Balls

XIANFU WANG
UBC Okanagan
3333 University Way, Kelowna
B.C., Canada, V1V 1V7
shawn.wang@ubc.ca

Diamonds, cylinders, squares, stars, and balls. These geometric figures are familiar to undergraduate students, but what could they possibly have in common? One answer is: They are generalized balls. The standard Euclidean ball can be distorted into a variety of strange-shaped balls by linear and nonlinear transformations. The purpose of this note is to give a unified formula for computing the volumes of generalized unit balls in n-dimensional spaces.

A generalized unit ball in $\mathbb{R}^n$ is described by the set

$$\mathbb{B}_{p_1 p_2 \cdots p_n} = \{\mathbf{x} = (x_1, \ldots, x_n) : |x_1|^{p_1} + \cdots + |x_n|^{p_n} \leq 1\}, \tag{1}$$

where $p_1 > 0, p_2 > 0, \ldots, p_n > 0$.

When the numbers $p_1, \ldots, p_n$ are all greater than or equal to 1, the unit ball $\mathbb{B}_{p_1 \cdots p_n}$ is convex. Since $|x|^p$ is not concave on $[-1, 1]$ for $0 < p < 1$, $\mathbb{B}_{p_1 \cdots p_n}$ is not necessarily convex anymore when $n > 1$. When $p_1 = p_2 = \cdots = p_n = p \geq 1$, we obtain the usual l_p ball. The l_2 ball is denoted by $\mathbb{B}$. By choosing different numbers p_i, we can alter the appearance of the generalized balls greatly, as shown in FIGURE 1 with examples in $\mathbb{R}^3$.

Motivated by an article by Folland [**5**], I derived a unified formula for calculating the volume of these balls. Although the volume formulas for the standard Euclidean ball $\mathbb{B}$ and simplex have been known for a long time [**4**, pp. 208, 220], the unified formula is (relatively) new. It is surprising that no matter how strange the balls look, the volume of any ball can be computed by a single formula, as follows:

THEOREM. *Assume $p_1, \ldots, p_n > 0$. The volume of the unit ball $\mathbb{B}_{p_1 p_2 \cdots p_n}$ in $\mathbb{R}^n$ is equal to*

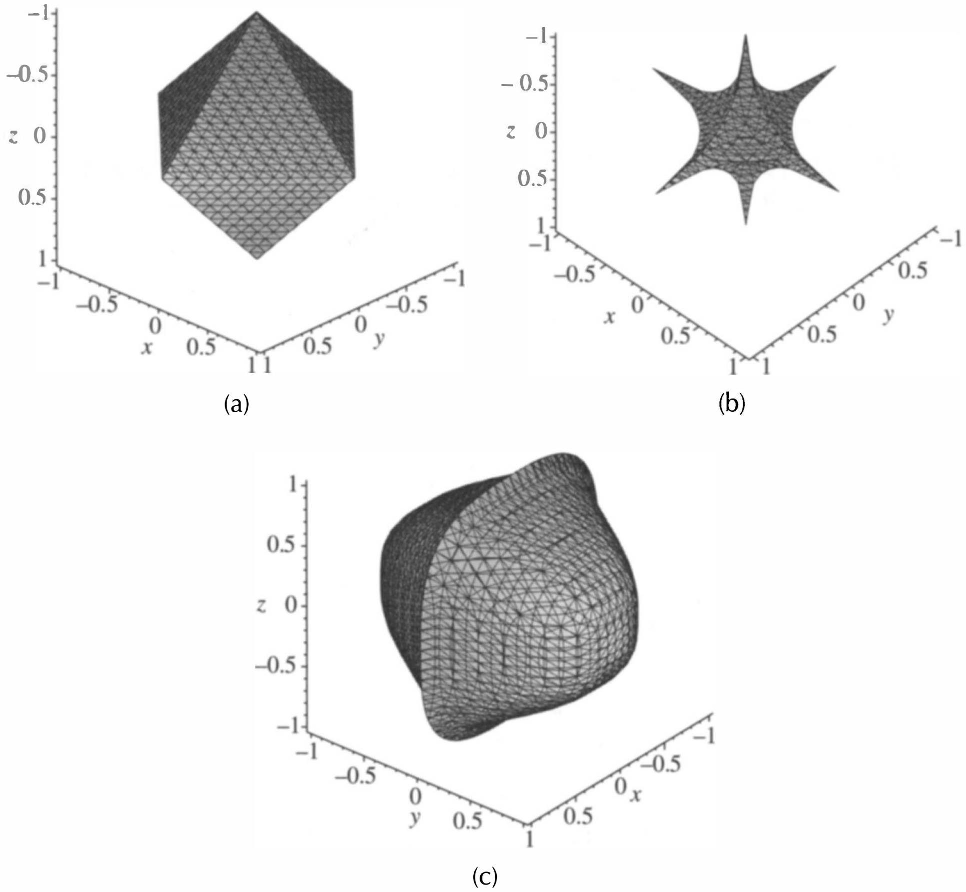


Figure 1 (a) $|x_1| + |x_2| + |x_3| \le 1$; (b) $|x_1|^{1/2} + |x_2|^{1/2} + |x_3|^{1/2} \le 1$; (c) $|x_1|^3 + |x_2|^{1/2} + |x_3|^3 \le 1$

$$2^n \frac{\Gamma(1 + 1/p_1) \cdots \Gamma(1 + 1/p_n)}{\Gamma(1/p_1 + 1/p_2 + \cdots + 1/p_n + 1)}. \tag{2}$$

The volume of the positive orthant part, where all x-values are positive, may be obtained by removing the factor of 2^n *from the formula.*

The formula involves the *gamma function*, which we review for readers who may be unfamiliar with it. For $0 < t < \infty$, we define

$$\Gamma(t) := \int_0^\infty s^{t-1} e^{-s}\, ds.$$

The integral converges for $t > 0$. The following facts will be needed: For $u > 0$ and $v > 0$, we have

$$\Gamma(u + 1) = u\Gamma(u), \tag{3}$$

and

$$\int_0^1 s^{u-1}(1-s)^{v-1}\,ds = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}. \tag{4}$$

Although the integral in $\Gamma(t)$ becomes infinite for $t \le 0$, (3) provides an analytic continuation formula to define $\Gamma(t)$ for $t < 0$. The function Γ has discontinuities only at $t = 0, -1, -2, \dots$. More details can be found in Folland [**6**, pp. 344–346].

Proof.

Step 1. We begin with the fact that

$$V(\mathbb{B}_{p_1\cdots p_n}) = \int_{\mathbb{B}_{p_1\cdots p_n}} 1\,d\mathbf{x}$$

and apply a change of variables that deforms the generalized ball into $\mathbb{B}$, the standard ball: Let $y_1 = x_1^{p_1/2}, \dots, y_n = x_n^{p_n/2}$. For the function

$$\phi(\mathbf{y}) := (y_1^{2/p_1}, \dots, y_n^{2/p_n}),$$

the Jacobian determinant is

$$J\phi(\mathbf{y}) = \frac{2}{p_1}\cdots\frac{2}{p_n} y_1^{\frac{2}{p_1}-1}\cdots y_n^{\frac{2}{p_n}-1}.$$

Readers may consult Folland [**6**, p. 432] for a detailed proof of the change of variables formula, which is our next ingredient. We use it to obtain

$$\int_{\mathbb{B}_{p_1\cdots p_n}} 1\,d\mathbf{x} = \int_{\mathbb{B}} \left|J\phi(y)\right|d\mathbf{y} = \frac{2^n}{p_1\cdots p_n}\int_{\mathbb{B}} |y_1|^{2/p_1-1}\dots|y_n|^{2/p_n-1}\,d\mathbf{y}.$$

Step 2. Assume $\alpha_1, \dots, \alpha_n > -1$. We claim:

$$\int_{\mathbb{B}} |x_1|^{\alpha_1}\dots|x_n|^{\alpha_n}\,d\mathbf{x} = \frac{\Gamma(\beta_1)\cdots\Gamma(\beta_n)}{\Gamma(\beta_1+\cdots+\beta_n+1)}, \tag{5}$$

where $\beta_i := (\alpha_i + 1)/2$ for $i = 1, \dots, n$.

To verify this claim, we develop a recursion formula. Let $I(\alpha_1, \dots, \alpha_n)$ denote the integral in (5). We then evaluate this as an iterated integral starting with x_1 as outermost variable.

$$I(\alpha_1, \dots, \alpha_n) = \int_{-1}^1 |x_1|^{\alpha_1} \int_{x_2^2+\cdots+x_n^2\le 1-x_1^2} |x_2|^{\alpha_2}\cdots|x_n|^{\alpha_n}dx_2\dots dx_n\,dx_1$$

The inner integration takes place over a ball of radius $r = \sqrt{1-x_1^2}$. Changing variables again, we set $(x_2, \dots, x_n) = r(y_2, \dots, y_n)$ to get

$$\int_{x_2^2+\cdots+x_n^2\le r^2} |x_2|^{\alpha_2}\cdots|x_n|^{\alpha_n}dx_2\cdots dx_n$$
$$= \int_{y_2^2+\cdots+y_n^2\le 1} r^{(n-1)+\alpha_2+\cdots+\alpha_n}|y_2|^{\alpha_2}\cdots|y_n|^{\alpha_n}dy_2\cdots dy_n.$$

This gives $I(\alpha_1, \dots, \alpha_n) =$

$$= \int_{-1}^{1} |x_1|^{\alpha_1}(1 - x_1^2)^{(n-1)/2+(\alpha_2+\dots+\alpha_n)/2} \int_{y_2^2+\dots+y_n^2 \le 1} |y_2|^{\alpha_2} \cdots |y_n|^{\alpha_n} dy_2 \dots dy_n\, dx_1$$

$$= 2\int_{0}^{1} x_1^{\alpha_1}(1 - x_1^2)^{(n-1)/2+(\alpha_2+\dots+\alpha_n)/2} dx_1 \cdot \int_{y_2^2+\dots+y_n^2 \le 1} |y_2|^{\alpha_2} \dots |y_n|^{\alpha_n} dy_2 \dots dy_n$$

$$= \int_{0}^{1} (x_1^2)^{(\alpha_1-1)/2}(1 - x_1^2)^{(\alpha_2+\dots+\alpha_n+n+1)/2-1} d(x_1^2) \int_{y_2^2+\dots+y_n^2 \le 1} |y_2|^{\alpha_2} \cdots |y_n|^{\alpha_n} dy_2 \dots dy_n.$$

Hence by (4),

$$I(\alpha_1, \dots, \alpha_n) = \frac{\Gamma((\alpha_1 + 1)/2)\Gamma((\alpha_2 + \dots + \alpha_n + n + 1)/2)}{\Gamma((\alpha_1 + \dots + \alpha_n + n + 2)/2)} \cdot I(\alpha_2, \dots, \alpha_n).$$

This provides a recursion formula connecting $I(\alpha_1, \alpha_2, \dots, \alpha_n)$ and $I(\alpha_2, \dots, \alpha_n)$. Applying the recursion formula $(n - 1)$ times, after cancellation, we obtain

$$I(\alpha_1, \dots, \alpha_n) = \frac{\Gamma(\frac{\alpha_1+1}{2}) \cdots \Gamma(\frac{\alpha_{n-1}+1}{2}) \frac{\alpha_n+1}{2} \Gamma(\frac{\alpha_n+1}{2})}{\Gamma(\frac{\alpha_1+\dots+\alpha_n+n}{2} + 1)} \cdot I(\alpha_n). \tag{6}$$

But

$$I(\alpha_n) = \int_{x^2 \le 1} |x|^{\alpha_n} dx = 2\int_0^1 x^{\alpha_n} dx = \frac{2}{\alpha_n + 1}.$$

Putting this into (6) yields (5).

Step 3. When $\alpha_i = 2/p_i - 1$ for $i = 1, \dots, n$, (5) gives

$$I(2/p_1 - 1, \dots, 2/p_n - 1) = \frac{\Gamma(1/p_1) \cdots \Gamma(1/p_n)}{\Gamma(1/p_1 + \dots + 1/p_n + 1)}.$$

Hence

$$\begin{aligned} V(\mathbb{B}_{p_1 \dots p_n}) &= 2^n \frac{1}{p_1} \cdots \frac{1}{p_n} I(2/p_1 - 1, \dots, 2/p_n - 1) \\ &= 2^n \frac{\Gamma(1 + 1/p_1) \cdots \Gamma(1 + 1/p_n)}{\Gamma(1/p_1 + \dots + 1/p_n + 1)}. \end{aligned}$$

The volume of positive orthant part follows from there being 2^n orthants in $\mathbb{R}^n$. ■

In (1), you might argue that p_i cannot be infinite, but, my dear readers, we can consider a limiting case. Let us write

$$x^\infty = \begin{cases} 0 & \text{if } 0 \le x < 1, \\ 1 & \text{if } x = 1. \end{cases}$$

We proceed to single out a few special cases (calculus students' delights):

1. Some $p_i = +\infty$: as Γ is continuous on $(0, +\infty)$, we have $V(\mathbb{B}_{p_1 \dots p_n}) =$

$$2 \cdot 2^{n-1} \frac{\Gamma(1 + 1/p_1) \cdots \Gamma(1 + 1/p_{i-1})\Gamma(1 + 1/p_{i+1}) \cdots \Gamma(1 + 1/p_n)}{\Gamma(1/p_1 + \dots + 1/p_{i-1} + 1/p_i + \dots + 1/p_n + 1)}.$$

In particular, when $p_1 = p_2 = \cdots = p_n = +\infty$, the volume of the ball is 2^n, and the shape is an n-dimensional hypercube (excluding the portions of its boundary where two or more x_is are simultaneously 1). When $p_1 = p_2 = 2$, $p_3 = \infty$, the generalized ball is a circular cylinder in $\mathbb{R}^3$.

2. When $p_1 = p_2 = \cdots = p_n = p > 0$, we have $V(\mathbb{B}_{p\ldots p}) =$

$$2^n \frac{(\Gamma(1+1/p))^n}{\Gamma(n/p+1)} = \frac{(2/p)^n (\Gamma(1/p))^n}{(n/p)\Gamma(n/p)}.$$

Recall that $\Gamma(1) = 1$, $\Gamma(1/2) = \pi^{1/2}$, and $\Gamma(n) = (n-1)!$. For $p = 2$, the generalized ball is the standard Euclidean ball with volume $2\pi^{n/2}/(n\Gamma(n/2))$. For $p = 1$, the generalized ball

$$\{(x_1, \ldots, x_n) : |x_1| + \cdots + |x_n| \le 1\},$$

is an n-dimensional diamond, and has volume $2^n/n!$. For $p = 1/2$, the generalized ball has volume $2^{2n}/(2n)!$, and its shape is an n-dimensional star. These are two of the balls shown in FIGURE 1.

Surprisingly, for $0 < p < \infty$ we find the n-dimensional ball has smaller volume when n becomes larger, and that

$$\lim_{n\to\infty} V(B_{\underbrace{p,p,\ldots,p}_{n \text{ terms}}}) = \lim_{n\to\infty} \frac{(2/p)^n (\Gamma(1/p))^n}{(n/p)\Gamma(n/p)} = 0.$$

Here we use Stirling's formula: $\Gamma(x) \sim \sqrt{2\pi} x^{x-1/2} e^{-x}$, where $\sim$ means that the ratio of the quantity on the left and right approaches 1 as $x \to \infty$ [**6**, p. 353].

3. For the ellipsoid $\{(x_1, \ldots, x_n) : |x_1|^{p_1}/a_1^{p_1} + \cdots + |x_n|^{p_n}/a_n^{p_n} \le 1\}$, with $a_i > 0$, a simple linear transformation reduces it to the form in (1) and the theorem yields its volume as

$$a_1 \cdots a_n \cdot 2^n \frac{\Gamma(1+1/p_1)\cdots\Gamma(1+1/p_n)}{\Gamma(1/p_1 + \cdots + 1/p_n + 1)}.$$

FIGURE 2 shows two ellipsoids in $\mathbb{R}^3$ to give readers an idea of their appearance.

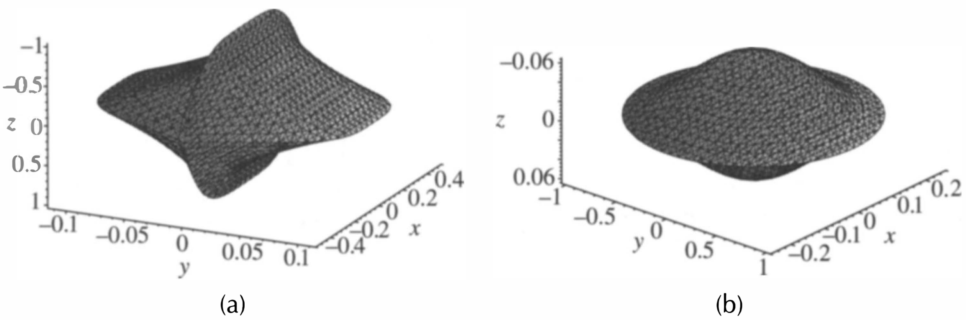


Figure 2 (a) $16|x_1|^4 + 3|x_2|^{1/2} + |x_3| \le 1$; (b) $16|x_1|^2 + |x_2|^2 + 4|x_3|^{1/2} \le 1$

Remark After I obtained this result, Dr. J. M. Borwein, at Simon Fraser University, informed me that the 19th-century French mathematician Dirichlet had obtained a similar result using a different method [**3**, pp. 153–159]. An *induction-free* proof to the volume formula of the l_p ball, via the *Laplace transform*, has been given by Borwein and Bailey in [**2**, pp. 195–197]. Similarly, one can derive an induction-free proof to the volume formula of generalized balls (2) using the Laplace transform. Finally, we remark that more properties on the gamma function and volume of Euclidean balls can be found in Stromberg [**7**, pp. 394–395].

Acknowledgment. This work was supported by NSERC and by the Grant In Aid of Okanagan University College.

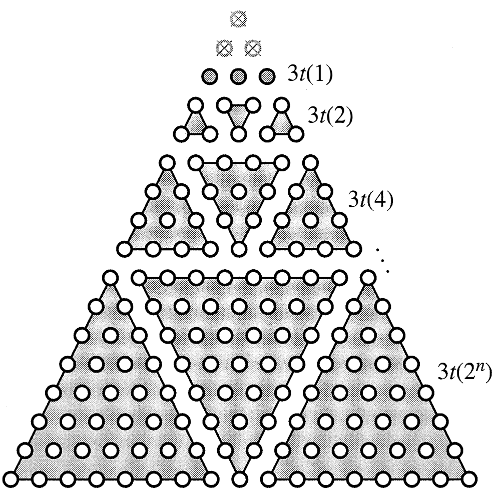
REFERENCES

1. J. A. Baker, Integration over spheres and the divergence theorem for balls, *Amer. Math. Monthly* **104** (1997), 36–47.
2. J. M. Borwein, D. H. Bailey, *Mathematics by Experiment: Plausible Reasoning in the 21st Century*, A.K. Peters, 2003.
3. J. Edwards, *A Treatise on the Integral Calculus*, Vol. II, Chelsea Publishing Company, New York, 1922.
4. W. Fleming, *Functions of Several Variables*, Springer-Verlag, New York, 1977.
5. G. B. Folland, How to integrate a polynomial over a sphere, *Amer. Math. Monthly* **108** (2001), 446–448.
6. ———, *Advanced Calculus*, Prentice Hall, Upper Saddle River, NJ, 2002.
7. K. R. Stromberg, *An Introduction to Classical Real Analysis*, Wadsworth International Mathematics Series, 1981.

Proof Without Words: A Triangular Sum

$$t(n) = 1 + 2 + \cdots + n \to \sum_{k=0}^{n} t(2^k) = \frac{1}{3} t(2^{n+1} + 1) - 1$$

$t(2^{n+1} + 1) - 3$:



$$3 \sum_{k=0}^{n} t(2^k) = t(2^{n+1} + 1) - 3$$

Exercises: (a) $\sum_{k=1}^{n} t(2^k - 1) = \frac{1}{3} t(2^{n+1} - 2)$

(b) $\sum_{k=0}^{n} t(3 \cdot 2^k - 1) = \frac{1}{3} \left[t(3 \cdot 2^{n+1} - 2) - 1 \right]$

—ROGER B. NELSEN
LEWIS & CLARK COLLEGE
PORTLAND OR 97219

Partitions into Consecutive Parts

M. D. HIRSCHHORN
University of New South Wales
Sydney 2052, Australia
m.hirschhorn@unsw.edu.au

P. M. HIRSCHHORN
5/312 Finchley Road, Hampstead
London NW3 7AG, U.K.
phirschhorn.mba2005@london.edu

It is known, though perhaps not as well as it should be, that the number of partitions of n into (one or more) consecutive parts is equal to the number of odd divisors of n. (This is the special case $k = 1$ of a theorem of J. J. Sylvester [**1**, §46], to the effect that the number of partitions of n into distinct parts with k sequences of consecutive parts is equal to the number of partitions of n into odd parts (repetitions allowed) precisely k of which are distinct.)

For instance,

$$15 = 7 + 8 = 4 + 5 + 6 = 1 + 2 + 3 + 4 + 5,$$

so 15 has four partitions into consecutive parts, and 15 has four odd divisors, 1, 3, 5, and 15.

We shall prove the following result.

THEOREM. *The number of partitions of n into an odd number of consecutive parts is equal to the number of odd divisors of n less than $\sqrt{2n}$, while the number of partitions into an even number of consecutive parts is equal to the number of odd divisors greater than $\sqrt{2n}$.*

Proof. Suppose n is the sum of an odd number of consecutive parts. Then the middle part is an integer and is the average of the parts. Suppose the middle part is a, and the number of parts is $2k + 1$. The partition of n is

$$n = (a - k) + \cdots + a + \cdots + (a + k)$$

and $n = (2k + 1)a$. So $d = 2k + 1$ is an odd divisor of n and its codivisor is $d' = a$. Note that $a - k \geq 1$, that is, $2a - (2k + 1) > 0$, $d < 2d'$, $d < 2n/d$, and $d^2 < 2n$. Conversely, suppose d is an odd divisor of n with $d^2 < 2n$, and codivisor d'. Then $d < 2d'$, and if we write $2k + 1 = d$, $a = d'$ then

$$n = (a - k) + \cdots + a + \cdots + (a + k)$$

is a partition of n into $2k + 1$ consecutive parts.

Next, suppose n is the sum of an even number, $2k$, of consecutive parts. Then the average part is $a + 1/2$ for some integer a, the partition of n is

$$n = (a + 1 - k) + \cdots + a + (a + 1) + \cdots + (a + k),$$

and $n = 2k(a + 1/2) = k(2a + 1)$. Then $d = 2a + 1$ is an odd divisor of n and its codivisor is $d' = k$. Note that $a - k \geq 0$, $(2a + 1) - 2k > 0$, $d > 2d'$, $d > 2n/d$, and $d^2 > 2n$.

Conversely, suppose d is an odd divisor of n with $d^2 > 2n$, with codivisor d'. Then $d > 2d'$, and if we write $2a + 1 = d$, $k = d'$, then

$$n = (a + 1 - k) + \cdots + a + (a + 1) + \cdots + (a + k)$$

is a partition of n into an even number of consecutive parts. ■

REFERENCE

1. J. J. Sylvester, A constructive theory of partitions, arranged in three acts, an interact and an exodion, *Amer. J. Math.* **5** (1882), 251–330.

Means Generated by an Integral

HONGWEI CHEN
Department of Mathematics
Christopher Newport University
Newport News, VA 23606
hchen@cnu.edu

For a pair of distinct positive numbers, a and b, a number of different expressions are known as *means*:

1. the arithmetic mean: $A(a, b) = (a + b)/2$
2. the geometric mean: $G(a, b) = \sqrt{ab}$
3. the harmonic mean: $H(a, b) = 2ab/(a + b)$
4. the logarithmic mean: $L(a, b) = (b - a)/(\ln b - \ln a)$
5. the Heronian mean: $N(a, b) = (a + \sqrt{ab} + b)/3$
6. the centroidal mean: $T(a, b) = 2(a^2 + ab + b^2)/3(a + b)$

Recently, Professor Howard Eves **[1]** showed how many of these means occur in geometrical figures. The integral in our title is

$$f(t) = \frac{\int_a^b x^{t+1}\,dx}{\int_a^b x^t\,dx}, \tag{1}$$

which encompasses all these means: particular values of t in (1) give each of the means on our list. Indeed, it is easy to verify that

$$f(-3) = H(a, b), \qquad f\left(-\frac{3}{2}\right) = G(a, b), \qquad f(-1) = L(a, b),$$
$$f\left(-\frac{1}{2}\right) = N(a, b), \qquad f(0) = A(a, b), \qquad f(1) = T(a, b).$$

Moreover, upon showing that $f(t)$ is strictly increasing, we can conclude that

$$H(a, b) \le G(a, b) \le L(a, b) \le N(a, b) \le A(a, b) \le T(a, b), \tag{2}$$

with equality if and only if $a = b$.

To prove that $f(t)$ is strictly increasing for $0 < a < b$, we show that $f'(t) > 0$. By the quotient rule,

$$f'(t) = \frac{\int_a^b x^{t+1} \ln x \, dx \int_a^b x^t \, dx - \int_a^b x^{t+1} \, dx \int_a^b x^t \ln x \, dx}{(\int_a^b x^t \, dx)^2}. \tag{3}$$

Since the bounds of the definite integrals are constant, the numerator of this quotient can be written

$$= \int_a^b x^{t+1} \ln x \, dx \int_a^b y^t \, dy - \int_a^b y^{t+1} \, dx \int_a^b x^t \ln x \, dx$$

$$= \int_a^b \int_a^b x^t y^t \ln x (x - y) \, dx dy.$$

Substituting in a different manner, we write the same numerator as

$$= \int_a^b y^{t+1} \ln y \, dy \int_a^b x^t \, dx - \int_a^b x^{t+1} \, dx \int_a^b y^t \ln y \, dy$$

$$= \int_a^b \int_a^b x^t y^t \ln y (y - x) \, dx \, dy.$$

Averaging the two equivalent expressions shows that this numerator is

$$\frac{1}{2} \int_a^b \int_a^b x^t y^t (x - y)(\ln x - \ln y) \, dx dy > 0,$$

as long as $0 < a < b$. In view of (3), this implies that $f'(t) > 0$. Thus, $f(t)$ is strictly increasing as desired.

We next turn to a refinement of (2). Since

$$f(-2) = \frac{ab(\ln b - \ln a)}{b - a} = \frac{G^2(a, b)}{L(a, b)},$$

the monotonicity of $f(t)$ allows us to deduce the following well-known interpolation inequality:

$$H(a, b) \leq \frac{G^2(a, b)}{L(a, b)} < G(a, b).$$

For more results, some of which have been obtained by other authors [**2**], we define the power mean by

$$M_p(a, b) = \left(\frac{a^p + b^p}{2} \right)^{1/p}.$$

Observing that $M_{1/2}(a, b) = (G(a, b) + A(a, b))/2$ and

$$N(a, b) = \frac{1}{3} \big(G(a, b) + 2A(a, b) \big),$$

we challenge the reader to choose values of t in (1) to show that

$$L(a, b) < M_{1/3}(a, b) < \frac{1}{3} \big(2G(a, b) + A(a, b) \big)$$

$$< M_{1/2}(a, b) < N(a, b) < M_{2/3}(a, b).$$

Following the excellent suggestion of an anonymous referee, for which the author is grateful, we put the discussion in a wider context by generalizing the means defined by (1). We state a set of axioms, which, if satisfied by a class of functions, will entitle those functions to be called means. The axioms will be chosen by abstracting the most important properties of $f(t)$ in (1).

We say that a function $F(a, b)$ *defines a mean* for $a, b > 0$ when

1. $F(a, b)$ is continuous in each variable,
2. $F(a, b)$ is strictly increasing in each variable,
3. $F(a, b) = F(b, a)$,
4. $F(ta, tb) = tF(a, b)$ for all $t > 0$,
5. $a < F(a, b) < b$ for $0 < a < b$.

The reader is invited to show that a necessary and sufficient condition for $F(a, b)$ to define a mean is that for $0 < a \leq b$,

$$F(a, b) = b\, f(a/b),$$

where $f(s)$ is positive, continuous and strictly increasing for $0 < s \leq 1$, and satisfies $s < f(s) \leq 1$, for $0 < s < 1$. In particular, if ϕ is a positive continuous function on $(0, 1]$ and if

$$f(s) = f_\phi(s) = \frac{\int_s^1 x\phi(x)\,dx}{\int_s^1 \phi(x)\,dx},$$

then f satisfies these conditions and

$$F(a, b) = bf(a/b) = \frac{\int_a^b x\phi(x/b)\,dx}{\int_a^b \phi(x/b)\,dx}$$

defines a mean. Moreover, if ψ is positive continuous on $(0, 1]$ and ψ/ϕ is strictly increasing, then $f_\phi < f_\psi$ on $(0, 1)$. This gives a general perspective on the topic of means.

REFERENCES

1. Howard Eves, Means appearing in geometric figures, this MAGAZINE **76** (2003), 292–294.
2. D. S. Mitrinovic, J. E. Pecaric, and A. E. Fink, *Classical and New Inequalities in Analysis*, Kluwer Academic Publishers, Boston, 1993, 21–48.

Nonattacking Queens on a Triangle

GABRIEL NIVASCH
Raanana 43108, Israel
gnivasch@yahoo.com

EYAL LEV
Weizmann Institute of Science
Rehovot 76100, Israel
elgr@actcom.co.il

Most readers are surely familiar with the problem of placing eight nonattacking queens on a chessboard, and its natural generalization to an $n \times n$ board (see the references at the end of this note). Here we consider an interesting variant of this problem, in which the board is triangular.

We are given a triangular board of side n. A queen on the board can move along a straight line parallel to any of the board's sides (see FIGURE 1). Our problem is to place on the board as many queens as possible, without any two queens attacking each other.

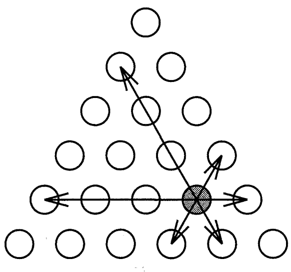


Figure 1 A queen on a triangular board

Obviously no more than n queens can be placed, since no row can contain more than one queen.

We now show how to get a tighter bound by counting in two different ways the total number of attacks by queens on cells. Let s be the number of times a cell is attacked by (that is, collinear with) a queen, summed over all the cells of the board. In this definition, we mean that a queen attacks its own cell three times—once for each direction of movement.

Each queen contributes exactly $2n + 1$ to s, no matter where it is placed. One way to see this is by projecting the cells attacked by a queen onto the board's bottom row, as shown in FIGURE 2. We have enough cells to cover the bottom row twice, plus one extra cell.

Therefore, if there are q nonattacking queens on the board, then

$$s = (2n + 1)q. \tag{1}$$

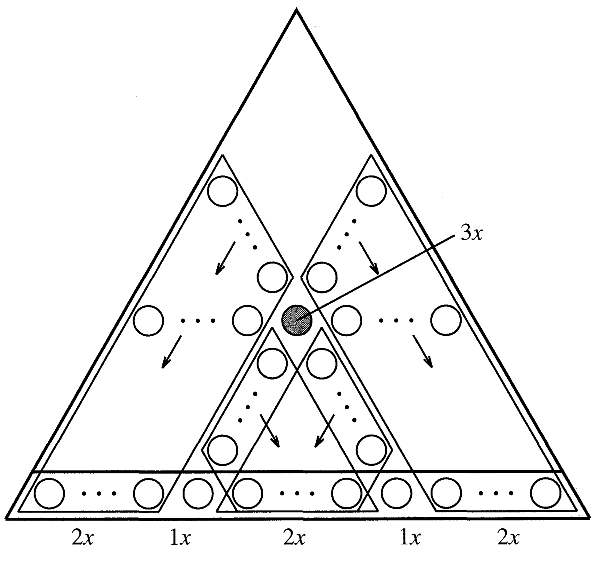


Figure 2 Each queen attacks $2n + 1$ cells

On the other hand, each cell can be attacked at most three times. And the number of cells is $1 + 2 + \cdots + n = n(n+1)/2$, the n-th triangular number. Therefore,

$$s \leq \frac{3n(n+1)}{2}. \tag{2}$$

Combining (1) and (2), we get

$$q \leq \frac{3n(n+1)}{2(2n+1)} < \frac{3n(n+1)}{4n} = \frac{3(n+1)}{4}.$$

Thus, we cannot place more than about $3n/4$ queens.

We can get a tighter bound by bounding the number of cells that can be actually attacked three times by a given configuration of queens.

Suppose q nonattacking queens are placed on an arbitrarily large board. Trace a line through each queen along each of the three directions of movement. Then, the cells attacked three times are exactly those where three lines intersect.

Let us ignore the queens for the moment, and concentrate on these three sets of lines, each set containing q parallel lines. Consider two of the sets of lines. The distance between adjacent lines may vary, but the lines will always produce a rhomboid of $q \times q$ intersections, as shown in FIGURE 3. Group the intersections into "layers" as indicated by the thick segments. The numbers of intersections in the layers will always be $1, 3, 5, \ldots, 2q-1$, the first q odd numbers (as can be easily shown by induction).

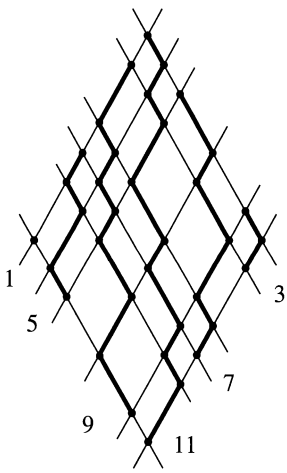


Figure 3 Intersections between two sets of lines, grouped into layers

Now, consider the third set of lines (which would be horizontal in the figure). Each such line can cross at most one intersection per layer, since no layer contains two horizontally-aligned intersections.

Therefore, to maximize the number of triple intersections, the most we can do is place the horizontal lines greedily one by one, each one passing through as many of the available intersections as possible. Thus, the first horizontal line can cross at most q intersections; the second line, at most $q-1$; the third line, again at most $q-1$; and so on, until we use up all the q horizontal lines. FIGURE 4 illustrates how to achieve the maximum number of triple intersections with equidistant lines.

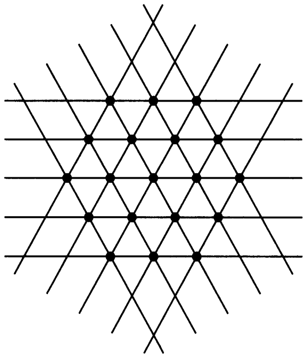


Figure 4 Maximizing the number of triple intersections

Therefore, the maximum number of triple intersections is, for q even,

$$q+2\left((q-1)+(q-2)+\cdots+\left(\frac{q}{2}+1\right)\right)+\frac{q}{2},$$

and, for q odd,

$$q+2\left((q-1)+(q-2)+\cdots+\frac{q+1}{2}\right).$$

The above sums are easily calculated in terms of triangular numbers; they equal $3q^2/4$ and $(3q^2+1)/4$, respectively. Thus, for all q, the number of triple intersections is no more than $(3q^2+1)/4$.

Now let us come back to our nonattacking queens on a triangle. It follows that s, the total number of times a cell is attacked by a queen, is bounded by

$$s \leq \frac{2n(n+1)}{2}+\frac{3q^2+1}{4}. \tag{3}$$

(We add 2 for each cell of the board, and then 1 for each cell that can be attacked a third time.)

Combining (1) and (3), we get a quadratic inequality for q. Surprisingly, the solution to this inequality does not involve radicals; the solution is

$$q \leq \frac{2n+1}{3} \quad \text{or} \quad q \geq 2n+1.$$

Obviously, q cannot be larger than n, and q must be an integer. Therefore,

$$q \leq \left\lfloor \frac{2n+1}{3} \right\rfloor. \tag{4}$$

An optimal solution We end by showing that the bound given by equation (4) is in fact tight. FIGURE 5 illustrates how to place $(2n+1)/3$ queens on a board of side n, when $n \equiv 1 \pmod 3$. If we number the board's columns 1 through $2n-1$ as shown, then the queens are placed on all cells in columns numbered $(2n+1)/3$ and $(4n+2)/3$.

For the cases $n \equiv 0$ and $n \equiv 2 \pmod 3$, we can use the same configuration and remove the board's bottom one or two rows, respectively, along with their queens. Therefore, we can always place $\lfloor (2n+1)/3 \rfloor$ queens on a board of side n.

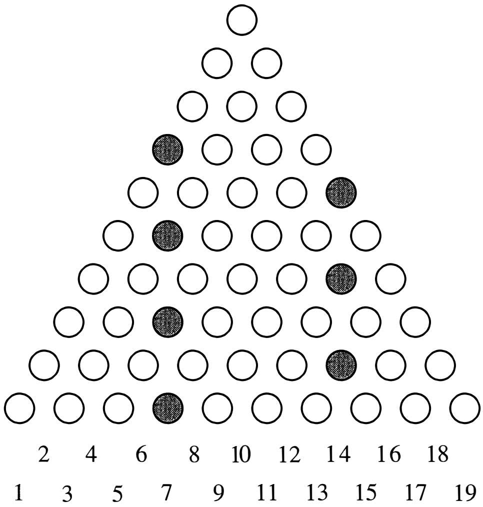


Figure 5 $(2n+1)/3 = 7$ nonattacking queens on a board of side $n = 10$

Acknowledgment. This work was done while the authors were at the Weizmann Institute of Science in Rehovot, Israel. The first author was partially supported by a grant from the German-Israeli Foundation for Scientific Research and Development. We would like to thank Professors Aviezri Fraenkel and Uriel Feige. Thanks also to the anonymous referees for helping us improve the presentation.

REFERENCES

1. Dean S. Clark and Oved Shisha, Proof without words: Inductive construction of an infinite chessboard with maximal placement of nonattacking queens, this MAGAZINE, **61** (1988), 98.
2. Cengiz Erbas and Murat M. Tanik, Generating solutions to the n-queens problem using 2-circulants, this MAGAZINE, **68** (1995), 343–356.
3. E. J. Hoffman, J. C. Loessi, and R. C. Moore, Constructions for the solution of the m queens problem, this MAGAZINE, **42** (1969), 66–72.
4. Matthias Reichling, A simplified solution of the n queens' problem, *Inform. Process. Lett.* **25** (1987), 253–255.

Editor's Note:

Late in our production period, it was pointed out that many of the results given in the Note "Nonattacking Queens on a Triangle" by Nivasch and Lev also appear in the popular book, *The Inquisitive Problem Solver*, by Vaderlind, Guy, and Larson, MAA, 2002. We hope that readers will nonetheless appreciate the treatment given here.

PROBLEMS

ELGIN H. JOHNSTON, *Editor*
Iowa State University

Assistant Editors: RĂZVAN GELCA, Texas Tech University; ROBERT GREGORAC, Iowa State University; GERALD HEUER, Concordia College; VANIA MASCIONI, Ball State University; BYRON WALDEN, Santa Clara University; PAUL ZEITZ, The University of San Francisco

Proposals

To be considered for publication, solutions should be received by May 1, 2006.

1731. *Proposed by Mowaffaq Hajja, Yarmouk University, Irbid, Jordan.*

Let A', B', and C' be points on sides $\overline{BC}$, $\overline{CA}$, and $\overline{AB}$, respectively, of triangle ABC, and let M be the point at which $\overline{AA'}$ intersects $\overline{B'C'}$. Prove that either $p(A'C'B) > p(A'C'M)$ or that $p(A'B'C) > p(A'B'M)$, where $p(XYZ)$ denotes the perimeter of triangle XYZ.

1732. *Proposed by Cafè Dalat Problem Solving Group, Washington D.C.*

Let n be a positive integer and define the function $f : [0,1]^n \longrightarrow [0,1]^2$ by

$$f(x_1, x_2, \dots, x_n) = \left(\frac{x_1 + x_2 + \cdots + x_n}{n},\ \sqrt[n]{x_1 x_2 \cdots x_n}\right).$$

Let $I(n)$ denote the range of f in $[0,1]^2$. Determine $\cup_{n=1}^{\infty} I(n)$.

1733. *Proposed by Ovidiu Furdui, student, Western Michigan University, Kalamazoo, MI.*

Let x be a fixed real number. For positive integer n define

$$a_n = \left(1 + \frac{1}{n}\right)^{nx} \quad \text{and} \quad b_n = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!}.$$

Determine

$$\lim_{n\to\infty} (a_{n+1} + a_{n+2} + \cdots + a_{2n} - b_{n+1} - b_{n+2} - \cdots - b_{2n}).$$

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.

Solutions and new proposals should be mailed to Elgin Johnston, Problems Editor, Department of Mathematics, Iowa State University, Ames IA 50011, or mailed electronically (ideally as a LaTeX file) to ehjohnst@iastate.edu. All communications should include the readers name, full address, and an e-mail address and/or FAX number.

1734. *Proposed by H. A. ShahAli, Tehran, Iran.*

Let α be a fixed irrational number and let P be a polynomial with integer coefficients and with $\deg(P) \geq 1$. Prove that there are infinitely many pairs (m, n) of integers such that $P(m) = \lfloor n\alpha \rfloor$.

1735. *Proposed by George Gilbert, Texas Christian University, Fort Worth, TX.*

Find the complex zeroes of the polynomial

$$p_n(z) = \det \begin{bmatrix} -1 & z & 0 & & & \cdots & 0 \\ 1-z & -1 & z & 0 & & \cdots & 0 \\ 0 & 1-z & -1 & z & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & 0 & \cdots & 0 & 1-z & -1 & z \\ 0 & 0 & \cdots & & 0 & 1-z & -1 \end{bmatrix},$$

where the matrix is an $n \times n$ tridiagonal matrix.

Quickies

Answers to the Quickies are on page 410.

Q955. *Proposed by Sadi Abu-Saymeh and Mowaffaq Hajja, Yarmouk University, Irbid, Jordan.*

Suppose the cevians $\overline{AA'}$, $\overline{BB'}$, and $\overline{CC'}$ of triangle ABC intersect at E. Prove that the quadrilaterals $EC'BA'$ and $EB'CA'$ have the same area if and only if $BA' = CA'$.

Q956. *Proposed by Steven Kahan, Queens College, Flushing, NY.*

Let x, y, z be positive numbers with

$$\frac{1}{x} + \frac{1}{y} + \frac{1}{z} \geq \frac{27}{8}.$$

Prove that

$$\sqrt{xyz} \leq \frac{x+y+z}{3}.$$

Solutions

Products and Logarithms **December 2004**

1706. *Proposed by Steve Edwards and James Whitenton, Southern Polytechnic State University, GA.*

Let $0 < a, b < 1$. Evaluate

$$\prod_{n=-\infty}^{\infty} \frac{1+b^{2^n}}{1+a^{2^n}}.$$

Solution by Cal Poly Pomona Problem Solving Group, Cal Poly Pomona, Pomona, CA.

$$\prod_{n=0}^{N}\frac{1+b^{2^n}}{1+a^{2^n}}=\frac{(1+b)(1+b^2)\cdots(1+b^{2^N})}{(1+a)(1+a^2)\cdots(1+a^{2^N})}$$

$$=\frac{1+b+b^2+\cdots+b^{2^{N+1}-1}}{1+a+a^2+\cdots+a^{2^{N+1}-1}}=\frac{1-b^{2^{N+1}}}{1-a^{2^{N+1}}}\cdot\frac{1-a}{1-b}.$$

Therefore,

$$\prod_{n=0}^{\infty}\frac{1+b^{2^n}}{1+a^{2^n}}=\lim_{N\to\infty}\frac{1-b^{2^{N+1}}}{1-a^{2^{N+1}}}\cdot\frac{1-a}{1-b}=\frac{1-a}{1-b}.$$

Next,

$$\prod_{n=-M}^{-1}\frac{1+b^{2^n}}{1+a^{2^n}}=\frac{(1+b^{1/2})(1+b^{1/2^2})\cdots(1+b^{1/2^M})}{(1+a^{1/2})(1+a^{1/2^2})\cdots(1+a^{1/2^M})}$$

$$=\frac{1+b^{1/2^M}+(b^{1/2^M})^2+\cdots+(b^{1/2^M})^{2^M-1}}{1+a^{1/2^M}+(a^{1/2^M})^2+\cdots+(a^{1/2^M})^{2^M-1}}=\frac{1-b}{1-a}\cdot\frac{1-a^{1/2^M}}{1-b^{1/2^M}}.$$

Hence, using L'Hôpital's rule,

$$\prod_{n=-\infty}^{-1}\frac{1+b^{2^n}}{1+a^{2^n}}=\lim_{M\to\infty}\left(\frac{1-b}{1-a}\cdot\frac{1-a^{1/2^M}}{1-b^{1/2^M}}\right)=\frac{1-b}{1-a}\cdot\frac{\ln a}{\ln b}.$$

Finally

$$\prod_{n=-\infty}^{\infty}\frac{1+b^{2^n}}{1+a^{2^n}}=\left(\prod_{n=-\infty}^{-1}\frac{1+b^{2^n}}{1+a^{2^n}}\right)\left(\prod_{n=0}^{\infty}\frac{1+b^{2^n}}{1+a^{2^n}}\right)=\frac{\ln a}{\ln b}.$$

Also solved by Hamza Ahmad, Tsehaye Andeberhan, Michael Andreoli, Micha Anholt (Israel), Michel Bataille (France), J. C. Binz (Switzerland), Jean Bogaert (Belgium), Paul Bracken, Minh Can, Juan Carlos Iglesias Castañón (Honduras), Con Amore Problem Group (Denmark), Chip Curtis, Jim Delany, Paul Deiermann, Daniele Donini (Italy), Joe Flowers, Michael Goldenberg and Mark Kaplan, G.R.A.20 Math Problems Group (Italy), Jim Hartman, W. P. Heidorn (Germany), Russell Jay Hendel, Parviz Khalili, Harris Kwong, Victor Y. Kutsenok, Sam McDonald and Kuzman Adzievski, Ioana Mihaila, José H. Nieto (Venezuela), Northwestern University Math Problem Solving Group, Rob Pratt, Manuel Reyes, Rolf Richberg (Germany), Thomas Richards, Volkhard Schindler (Germany), Heinz-Jürgen Seiffert (Germany), Nicholas C. Singer, Byron Siu, Albert Stadler (Switzerland), Ricardo M. Torrejón, Dave Trautman, Lucas Van der Merwe and Stan Byrd, Michael Vowe (Switzerland), Paul Weisenhorn (Germany), Chu Wenchang (Italy), Michael Woltermann, Li Zhou, and the proposer. There were two incorrect submissions.

An Urned Expectation **December 2004**

1707. *Proposed by Barthel Wayne Huff, Salt Lake City, Utah.*

Let k and n be positive integers. Evaluate

$$\sum_{j=1}^{n}\left(\frac{jk}{n-j+k}\prod_{r=1}^{j-1}\frac{n-r}{n-r+k}\right),$$

where the empty product is equal to 1.

I. *Solution by Rob Pratt, Raleigh, NC.*

First note that

$$\sum_{j=1}^{n} j\binom{k+n-j-1}{k-1}=\binom{k+n}{k+1},$$

which can be proved combinatorially by counting the number of $(k+1)$-subsets of $\{0, 1, \ldots, k+n-1\}$ in two different ways. It is clear that the right side is the number of such sets. The left side is obtained by conditioning on the second smallest element j of the subset: we then have j choices, $\{0, 1, \ldots, j-1\}$, for the smallest element, then must choose the remaining $k-1$ elements from $\{j+1, j+2, \ldots, k+n-1\}$.

The sum in the problem statement can now be rewritten as

$$\sum_{j=1}^{n}\left(jk\left(\prod_{r=1}^{j-1}(n-r)\right)\left(\prod_{r=1}^{j}\frac{1}{k+n-r}\right)\right)$$

$$=\sum_{j=1}^{n}\left(j\cdot\frac{k!}{(k-1)!}\cdot\frac{(n-1)!}{(n-j)!}\cdot\frac{(k+n-j-1)!}{(k+n-1)!}\right)$$

$$=\binom{k+n-1}{k}^{-1}\sum_{j=1}^{n}j\binom{k+n-j-1}{k-1}=\binom{k+n-1}{k}^{-1}\binom{k+n}{k+1}=\frac{k+n}{k+1}.$$

II. *Solution by the proposer.*

Consider an urn containing $n-1$ red balls and k blue balls. Balls are drawn at random from the urn and not replaced until a blue ball is obtained. Let T_n^k be the draw on which the first blue ball is obtained. Then

$$E[T_n^k]=1\cdot\frac{k}{n-1+k}+2\cdot\frac{n-1}{n-1+k}\cdot\frac{k}{n-2+k}+\cdots$$
$$+n\cdot\frac{n-1}{n-1+k}\cdot\cdots\cdot\frac{1}{k+1}\cdot\frac{k}{k}$$
$$=\sum_{j=1}^{n}\left(\frac{jk}{n-j+k}\prod_{r=1}^{j-1}\frac{n-r}{n-r+k}\right).$$

On the other hand, we intuitively expect the blue balls to be evenly spaced through the sequence of draws that empties the urn, that is the expected position of the first ball should be

$$E[T_n^k]=\frac{n-1}{k+1}+1=\frac{n+k}{k+1}.$$

This intuition can be shown to be correct by conditioning upon whether or not the first ball drawn is blue to obtain

$$E[T_n^k]=1\cdot P[T_n^k=1]+\left(1+E[T_{n-1}^k]\right)\cdot P[T_n^k\neq 1]$$
$$=1+P[T_n^k\neq 1]\cdot E[T_{n-1}^k]=1+\frac{n-1}{n-1+k}\cdot E[T_{n-1}^k].$$

Because it is clear that $E[T_1^k]=1$ (the no red ball case) a straight forward induction argument yields $E[T_n^k]=(n+k)/(k+1)$.

Also solved by Hamza Ahmad, Tsehaye Andeberhan, Michael Andreoli, Michel Bataille (France), J. C. Binz (Switzerland), Jean Bogaert (Belgium), Paul Bracken and N. Nadeau, Brian Bradie, Con Amore Problem Group (Denmark), Chip Curtis, Daniele Donini (Italy), Joe Flowers, Russell Jay Hendel, Michael Goldenberg and Mark Kaplan, Sam McDonald and Kuzman Adzievski, Daniel R. Patten, Rolf Richberg (Germany), Volkhard Schindler

(Germany), Heinz-Jürgen Seiffert (Germany), Byron Siu, Albert Stadler (Switzerland), Michael Vowe (Switzerland), Chu Wenchang (Italy), Paul Weisenhorn (Germany), and Li Zhou.

When a Parallelogram Is a Square **December 2004**

1708. *Proposed by Stephen J. Herschkorn, Highland Park, NJ.*

It is well known that the area of a square is half the square of the length of its diagonal. Show that if the area of a parallelogram is half the square of one of its diagonals, and if the area and each side have rational measure, then the parallelogram is a square.

Solution by Andrew J. Miller, Belmont University, Nashville, TN.

Without loss of generality, we may assume that the length of one side of the parallelogram is 1 and the other side length is a. Let θ be the measure of one of the smaller vertex angles. Then the area of the parallelogram is $a \sin\theta$ and the squares of the lengths of its diagonals are $a^2 \pm 2a\cos\theta + 1$. Because the area is equal to half the square of the length of one of the diagonals, we have

$$2a\sin\theta = a^2 \pm 2a\cos\theta + 1, \tag{1}$$

from which

$$a^2 - 2a\mu + 1 = 0, \tag{2}$$

where $\mu = \sin\theta \pm \cos\theta$. Equation (2) has real solutions for a if and only if $\mu^2 - 1 \geq 0$. Because $\mu^2 - 1 = \pm\sin(2\theta)$ and $0 \leq \theta \leq \frac{\pi}{2}$, it follows that $\mu = \sin\theta + \cos\theta$. In this case the solutions to (2) are

$$a = \sin\theta + \cos\theta \pm \sqrt{\sin 2\theta}. \tag{3}$$

Because a and the area of the parallelogram are rational, it follows that $\sin\theta$ is rational, and from (1) and (3) that $\cos\theta$ and $\sqrt{\sin 2\theta}$ are also rational. Thus there is a rational number x such that

$$(\cos\theta,\ \sin\theta) = \left(\frac{2x}{1+x^2},\ \frac{1-x^2}{1+x^2}\right),$$

and in addition

$$\sqrt{\sin 2\theta} = \frac{2}{1+x^2}\sqrt{x(1-x^2)}.$$

Since both x and $\sqrt{\sin 2\theta}$ are rational, this equation implies that there is a rational number y such that

$$x - x^3 = y^2.$$

But it is well known that the only rational solutions to this equation are $(x, y) = (0, 0)$ and $(x, y) = (\pm 1, 0)$. Thus $\sin 2\theta = 0$, so $\theta = \pi/2$. It follows from (1) that $a = 1$, so the parallelogram is a square.

Also solved by Tsehaye Andeberhan, Roy Barbara (Lebanon), Henry J. Barten, Michel Bataille (France), Claude Begin (Canada), Minh Can, John Christopher, Con Amore Problem Group (Denmark), Chip Curtis, Jim Delany, Daniele Donini (Italy), Timothy Eckert, Michael Goldenberg and Mark Kaplan, G.R.A.20 Math Problems Group (Italy), Victor Y. Kutsenok, Allen J. Mauney, Sam McDonald and Kuzman Adzievski, Volkhard Schindler (Germany), H. T. Tang, Michael Vowe (Switzerland), Michael Woltermann, Li Zhou, and the proposer. There was one incorrect submission.

A Product Inequality **December 2004**

1709. *Proposed by Mihály Bencze, Săcele-Négyfalu, Romania.*

Let $x_1, x_2, \ldots, x_{3n} \geq 0$. Prove that

$$2^n \prod_{k=1}^{3n} \frac{1+x_k^2}{1+x_k} \geq \left(1 + \prod_{k=1}^{3n} x_k^{1/n}\right)^n .$$

Solution by Michael Reid, University of Central Florida, Orlando, FL.

The statement of the problem makes sense and remains true if $3n$ is a nonnegative integer, though n need not be. We prove this slightly more general statement, and show that equality holds if and only if each $x_k = 1$. We first prove a simple lemma.

LEMMA. *Let $a_1, a_2, \ldots, a_m$ be nonnegative real numbers and let g be their geometric mean. Then $\prod_{k=1}^{m}(1+a_k) \geq (1+g)^m$.*

Proof. Expand the product to get

$$\prod_{k=1}^{m}(1+a_k) = 1 + s_1 + s_2 + \cdots + s_m,$$

where the s_ks are the elementary symmetric functions of the a_ks. By the arithmetic-geometric mean inequality, $s_k \geq \binom{m}{k} g^k$ for $1 \leq k \leq m$. Thus

$$\prod_{k=1}^{m}(1+a_k) = 1 + \sum_{k=1}^{m} s_k \geq 1 + \sum_{k=1}^{m} \binom{m}{k} g^k = (1+g)^m.$$

We now prove the desired inequality. Let $m = 3n$ be a nonnegative integer. Note that for any real number t, we have

$$2(1+t^2)^3 = (1+t^3)(1+t)^3 + (1-t^3)(1-t)^3 \geq (1+t^3)(1+t)^3,$$

with equality if and only if $t = 1$. In particular, for nonnegative t,

$$2^{1/3} \frac{1+t^2}{1+t} \geq (1+t^3)^{1/3}, \qquad (*)$$

with equality if and only if $t = 1$. This is the required inequality for the case $m = 1$. Apply $(*)$ with $t = x_1, x_2, \ldots, x_m$, take the product, then apply the lemma to get

$$2^{m/3} \prod_{k=1}^{m} \frac{1+x_k^2}{1+x_k} \geq \left(\prod_{k=1}^{m}(1+x_k^3)\right)^{1/3} \geq \left(1 + \prod_{k=1}^{m} x_k^{3/m}\right)^{m/3},$$

which is the desired inequality. If equality holds, then it must hold in the first inequality above, which implies that $x_1 = x_2 = \cdots = x_m = 1$. ∎

Also solved by Tsehaye Andeberhan, Michel Bataille (France), Daniele Donini (Italy), Ovidiu Furdui, G.R.A.20 Problems Group (Italy), W. P. Heidorn (Germany), Michael G. Neubauer, Rolf Richberg (Germany), Li Zhou, and the proposer.

A GCD Relationship **December 2004**

1710. *Proposed by William D. Weakley, Indiana-Purdue University at Fort Wayne, Fort Wayne, IN.*

Let $n > 1$ be an integer and let $(a_1\ a_2 \cdots a_n)$ and $(b_1\ b_2 \cdots b_n)$ be $1 \times n$ matrices with integer entries. Show that $\text{GCD}(a_1, a_2, \ldots, a_n) = \text{GCD}(b_1, b_2, \ldots, b_n)$ if and only if there is an $n \times n$ matrix M with integer entries and $\det(M) = 1$ such that

$$(a_1\ a_2 \cdots a_n)M = (b_1\ b_2 \cdots b_n).$$

Solution by Jim Delany, California Polytechnic State University, San Luis Obispo, CA.

The $n \times n$ matrices M with integer entries and $\det(M) = 1$ constitute the group $SL_n(\mathbb{Z})$. Define an equivalence relation on $\mathbb{Z}^n$ by defining $\mathbf{a}, \mathbf{b} \in \mathbb{Z}^n$ to be equivalent if there is an $M \in SL_n(\mathbb{Z})$ such that $\mathbf{a}M = \mathbf{b}$. We first show that $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{Z}^n$ is equivalent to $(d, 0, \ldots, 0)$, where $d = \text{GCD}(a_1, a_2, \ldots, a_n)$.

Let $d_1 = a_1$. For $2 \le k \le n$ let $d_k = \text{GCD}(a_1, a_2, \ldots, a_k) = \text{GCD}(d_{k-1}, a_k)$, (where we define $\text{GCD}(0, 0) = 0$) and let x_k, y_k be integers such that $d_{k-1}x_k + a_k y_k = d_k$. If $d_n = 0$, then $\mathbf{a} = (0, 0, \ldots, 0)$ and the assertion is trivial. If $d_n > 0$, let $m = \min\{k \ge 2 : d_k > 0\}$. (Thus, if $a_1 \ne 0$, then $m = 2$, and if $a_1 = 0$, then $m = \min\{k : a_k > 0\}$.)

Define matrices $M_1, M_2, \ldots, M_n \in SL_n(\mathbb{Z})$ as follows. For $k < m$ let $M_k = I_n$, the $n \times n$ identity matrix. For $m \le k \le n$, let $M_k = M_{k-1}P_k$ where

$$P_k = \left(p_{ij}^{(k)}\right)_{1 \le 1,\, j \le n} \in SL_n(\mathbb{Z})$$

is the matrix with entries

$$p_{11}^{(k)} = x_k, \quad p_{k1}^{(k)} = y_k, \quad p_{1k}^{(k)} = -\frac{a_k}{d_k}, \quad p_{kk}^{(k)} = \frac{d_{k-1}}{d_k},$$

and

$$p_{ij}^{(k)} = \delta_{ij} \quad \text{otherwise.}$$

We then have $\mathbf{a}M_k = \mathbf{a}$ for $1 \le k < m$. By induction,

$$\mathbf{a}M_k = (d_k, 0, \ldots, 0, a_{k+1}, \ldots, a_n)$$

for $m \le k < n$ and $\mathbf{a}M_n = (d_n, 0, \ldots, 0)$. This confirms our assertion. Thus any two n-tuples with the same GCD are equivalent.

For the converse we need to show that if $c, d \ge 0$, and if $(c, 0, \ldots, 0)$ and $(d, 0, \ldots, 0)$ are equivalent, then $c = d$. Suppose that $(c, 0, \ldots, 0)M = (d, 0, \ldots, 0)$ for some $M \in SL_n(\mathbb{Z})$. Then $(d, 0, \ldots, 0)M^{-1} = (c, 0, \ldots, 0)$. It follows that $m_{11}c = d$ and $m_{11}^{(-1)}d = c$, so $c \mid d$, $d \mid c$, and $c = d$.

Also solved by Hamza Ahmad, Michel Bataille (France), Con Amore Problem Group (Denmark), Randall J. Covill, Daniele Donini (Italy), Victor Y. Kutsenok, Sam McDonald and Kuzman Adzievski, Manuel Reyes, Rolf Richberg (Germany), Achilleas Sinefakopoulos, and the proposer.

Answers

Solutions to the Quickies from page 405.

A955. If $BA' = CA'$, then it follows from Ceva's Theorem that $AB'/B'C = AC'/C'B$ and hence that $\overline{B'C'} \parallel \overline{BC}$. Therefore $[C'BC] = [B'CB]$, where $[\cdots]$ denotes area. It also follows from $BA' = CA'$ that $[BEA'] = [CEA']$. Therefore, $[EC'BA'] = [EB'CA']$, as desired.

For the converse, assume that $BA' < A'C$ and let M be the midpoint of $\overline{BC}$. Let $\overline{AM}$ intersect $\overline{BB'}$ at F, and extend $\overline{CF}$ through F to intersect $\overline{AB}$ at C''. Then by the first part of the proof,

$$[EC'BA'] < [FC''BM] = [FB'CM] < [EB'CA'].$$

A956. From the given inequality it follows that $yz + xz + xy \geq \frac{27}{8}xyz$. Next note that

$$\begin{aligned} x^2 + y^2 + z^2 + 2xy - 2xz - 2yz &= (x + y - z)^2 \geq 0 \\ x^2 + y^2 + z^2 - 2xy + 2xz - 2yz &= (x - y + z)^2 \geq 0 \\ x^2 + y^2 + z^2 - 2xy - 2xz + 2yz &= (-x + y + z)^2 \geq 0. \end{aligned}$$

Adding these inequalities leads to

$$x^2 + y^2 + z^2 \geq \frac{2}{3}(xy + xz + yz).$$

Then

$$\begin{aligned} (x + y + z)^2 &= x^2 + y^2 + z^2 + 2xy + 2xz + 2yz \\ &\geq \frac{8}{3}(xy + xz + yz) \geq \frac{8}{3}\left(\frac{27}{8}xyz\right) = 9xyz. \end{aligned}$$

Dividing by 9 and taking square roots gives the desired inequality.

The MAGAZINE in Numbers—A Snapshot

Starting in 2000, the last year of the old millennium, we began entering MAGAZINE records in a database. It holds all the information about manuscripts, their authors and referees, and their status at various times. As of October 27, 2005, here are a few notable numbers from the database.

- 1,372 manuscripts are recorded in the system, including Articles, Notes, Proofs Without Words, and other miscellaneous material.
- 1,787 authors' names were attached to these manuscripts.
- 290 manuscripts were published, 931 not accepted, and 100 withdrawn.
- 1,570 referee reports were requested from volunteer referees, including 468 requests sent to Associate Editors.
- 270 letters were sent to ask authors to revise manuscripts. We received 192 revisions, with 11 still pending.

Notes: We stopped receiving new manuscripts at the start of 2005. Some manuscripts are being passed on to Allen Schwenk for publication in our 2006 issues; some decisions are still pending. The count of published manuscripts includes some received from Paul Zorn, the past editor, which approximately balance those sent on to Allen Schwenk.

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.

Kammen, Daniel M., and David M. Hassenzahl, *Should We Risk It? Exploring Environmental, Health, and Technological Problem Solving*, Princeton University Press, 1999; xx + 404 pp, \$32.95 (P). ISBN 0–691–07457–7. Harte, John, *Consider a Cylindrical Cow: More Adventures in Environmental Problem Solving*, University Science Books, 2001; xii + 212 pp, \$38 (P). ISBN 1–891389–17–3.

You may wonder why I am reviewing books that are several years old. Well, I didn't notice them earlier; they aren't exhibited at the meetings of mathematical societies, or advertised to mathematicians, so you probably didn't notice them, either. What courses in applied mathematics does your department teach? Most college students, apart from taking a required "distribution" course in a mathematics department, go on to learn their needed applied mathematics elsewhere; more applied mathematics is taught outside of mathematics departments than inside them. Some pure mathematicians are not interested in applied mathematics (and that's OK); others (including many teaching assistants) have no substantial background in any science or engineering and hence are incapable of teaching truly applied mathematics. Yet students must be equipped to address important questions—for example, involving the environment and its effects on humans (and vice versa)—that require quantitative analysis. Must students learn that only in a Department of Environmental Studies, of Energy and Society, or of Energy and Resources? The two books listed above, plus the predecessors *Consider a Spherical Cow: A Course in Environmental Problem Solving* (1988) by Harte and its accompanying computer models in Leonard J. Soltzberg's *The Dynamic Environment* (1996), point out a direction in which mathematics departments could expand their relevance. These books require (in places) background in calculus, matrix algebra, differential equations, probability, and statistics; but they also demand of instructor and students knowledge of (or willingness to learn) some physics, chemistry, and biology. Kammen and Hassenzahl set out the techniques of risk assessment (including Bayesian analysis, toxicology, epidemiology, and exposure assessment), while Harte in this second "cow" book focuses on probability, optimization, scaling, differential equations, stability, and feedback in environmental contexts. If only mathematicians knew a little more about the world, they could bring so much more mathematics to it.

Cipra, Barry, et al. (eds.), *Tribute to a Mathemagician*, A K Peters, 2005; x + 252 pp + 4 color plates, \$38. ISBN 1–56881–204–3.

This is the third book from "Gathering for Gardner" (G4G) meetings of mathematicians inspired by Martin Gardner. Details of some previous meetings, together with material on recreational math, are at `http://www.g4g4.com/`. This volume contains papers from G4G6; topics include interlocking puzzles, sliding-coin and sliding-block puzzles, origami, configuration games, domino portraits, and much more. Nick Baxter's explanation of the Pólya-Burnside lemma "in language that puzzlers can understand" may be useful in teaching students how to count symmetries, and I particularly enjoyed Underwood Dudley on how and why recreational mathematics beats phony applications problems ("[mathematics] is by far the best way to teach people to reason"). G4G7 will take place in March 2006, with participation by invitation.

Danesi, Marcel, *The Liar Paradox and the Towers of Hanoi: The Ten Greatest Math Puzzles of All Time*, Wiley, 2004; vii + 248 pp, $15.95 (P). ISBN 0–471–64816–7.

The recent feverish popularity of Sudoku proves that mathematical puzzles can excite vast numbers of people. Here are some of the "Sudokus" of yesteryear (and today): the riddle of the Sphinx, river-crossing puzzles, Fibonacci's rabbits, the four-color problem, Loyd's "Get Off the Earth" puzzle, the Cretan labyrinth, the Tower of Hanoi puzzle, the Königsberg bridges, magic squares, and the liar paradox. Each chapter includes further "Explorations" (don't call them exercises—they are puzzles!), with answers and explanations in the back of the book.

Hayes, Brian, Group theory in the bedroom, *American Scientist* (September-October 2005), `http://www.americanscientist.org/template/AssetDetail/assetid/47438` . Letters to the editor: Building a better mattress flip, (November-December 2005), `http://www.americanscientist.org/template/AssetDetail/assetid/47439` .

How should you "turn" a mattress to have it systematically pass through all of its (four) states, thus evening the wear? There is no "golden rule," a "set of geometric maneuvers that you could perform in the same way every time in order to cycle through all the configurations." The mathematical content of this article (which features "mattress multiplication" and an introduction to group theory) is that the Klein 4-group cannot be generated by a single element. A reader suggests "an alternating sequence of two basic flipping operations," demonstrating that the group can be generated by (any) two elements—Hayes dubs it a "silver rule." Now, if only king-size mattresses were square (instead of 76" by 80"), there would be some further "exploration" for the reader. (Note: The group that Hayes gives for tire rotation is interesting for didactic purposes (lots of golden rules) but applies only to nonradial tires; radial tires are supposed to remain on the same side of the car, hence their rotation subgroup is considerably smaller.)

Hadlock, Charles R. (ed.), *Mathematics in Service to the Community: Concepts and Models for Service-Learning in the Mathematical Sciences*, Mathematical Association of America, 2005; xi + 264 pp, $45.50 (P).

"What is service-learning and why should I be interested in it?" No, it's not necessarily tutoring. It's activities that "enhance delivery or impact of curricular material... in a service framework [of] civic engagement or social contribution." How can mathematics fit in? Natural arenas are mathematical modeling, statistics, and education-related activities. This book includes case studies in each category, from optimal staffing of a fire department to efficient snowplow routes, from community surveys to data organization, and from implementing use of graphing calculators to bringing teacher candidates together with children and their parents. There is plenty of advice about how to organize a service-learning course, but I wished for more accounts of projects, with testimonials from clients outside of academia. As editor Hadlock observes, there are "untapped possibilities." (Blatant advertisement from another editor in my chair wearing a different hat of mine: *The UMAP Journal* has for some years featured such case studies in its occasional MathServe section.)

Mankiewicz, Richard, *The Story of Mathematics*, Princeton University Press, 2004; 192 pp, $19.95 (P). ISBN 0–691–12046–3.

This book would be a great coffee-table book about mathematics at twice its size, enlarging the too-small print to readability and the many color illustrations to irresistible attractiveness (however, the quotations, in an ugly neo-Celtic font, would look even worse). The goal, "to illustrate how the mathematical sciences were intimately linked to the interests and aspirations of the civilizations in which they flourished," is achieved in a fast panorama virtually without equations or mathematical symbolism. Much of the content is thus what you would expect (astronomy, ancient and medieval mathematics from various cultures, Newton, cartography, the quintic, noneuclidean geometry, infinity, chance). Four of the 24 chapters relate somewhat to mathematics since 1900: game theory, noneuclidean dimensions of modern art, computing, and chaos/fractals. The illustrations are spectacular, the prose inspiring; this book deserves (besides a coffee-table edition) a permanent place in the (tiny) mathematics section of bookstores.

/eject

NEWS AND LETTERS

Acknowledgments

In addition to our Associate Editors, the following referees have assisted the MAGAZINE *during the past year. We thank them for their time and care.*

Adams, Colin, *Williams College, Williamstown, MA*

Allman, Elizabeth S., *University of Southern Maine, Falmouth, ME*

Axtell, Michael C., *Wabash College, Crawfordsville, IN*

Baeth, Nicholas R., *Central Missouri State University, Warrensburg, MO*

Barksdale, Jr., James B., *Western Kentucky University, Bowling Green, KY*

Bauer, Craig P., *York College, York, PA*

Beals, Richard W., *Yale University, New Haven, CT*

Beauregard, Raymond A., *University of Rhode Island, Kingston, RI*

Beezer, Robert A., *University of Puget Sound, Tacoma, WA*

Bennett, Curtis D., *Loyola Marymount University, Los Angeles, CA*

Borwein, Jonathan, *Dalhousie University, Halifax, Canada*

Bressoud, David M., *Macalester College, St. Paul, MN*

Bruckman, Paul S., *Sointula, Canada*

Bullington, Grady D., *University of Wisconsin Oshkosh, Oshkosh, WI*

Callan, C. David, *University of Wisconsin, Madison, WI*

Case, Jeremy S., *Taylor University, Upland, IN*

Chakerian, G. Don, *University of California, Davis, Davis, CA*

Chen, Hang, *Central Missouri State University, Warrensburg, MO*

Chinn, Phyllis Z., *Humboldt State University, Arcata, CA*

Davis, Phillip J., *Brown University, Providence, RI*

De Angelis, Valerio, *Xavier University, New Orleans, LA*

DeTemple, Duane W., *Washington State University, Pullman, WA*

Dodge, Clayton W., *University of Maine, Orono, ME*

Dunne, Edward G., *American Mathematical Society, Providence, RI*

Eenigenburg, Paul, *Western Michigan University, Kalamazoo, MI*

Eisenberg, Bennett, *Lehigh University, Bethlehem, PA*

Elderkin, Richard H., *Pomona College, Claremont, CA*

Ensley, Douglas E., *Shippensburg University, Shippensburg, PA*

Eroh, Linda, *University of Wisconsin Oshkosh, Oshkosh, WI*

Evans, Anthony, *Wright State University, Dayton, OH*

Feil, Todd, *Denison University, Granville, OH*

Feroe, John A., *Vassar College, Poughkeepsie, NY*

Fisher, Evan, *Lafayette College, Easton, PA*

Fisher, J. Chris, *University of Regina, Regina, Canada*

Flint, Donna L., *South Dakota State University, Brookings, SD*

Frantz, Marc, *Indiana University, Bloomington, IN*

Fraser, Craig, *University of Toronto, Toronto, Canada*

Fredricks, Gregory A., *Lewis & Clark College, Portland, OR*

Fung, Maria G., *Western Oregon University, Monmouth, OR*

Goddart, Wayne, *Clemson University, Clemson, SC*

Gordon, Russell A., *Whitman College, Walla Walla, WA*

Grabiner, Judith V., *Pitzer College, Claremont, CA*

Green, Euline I., *Abilene Christian University, Abilene, TX*

Grosshans, Frank D., *West Chester University of Pennsylvania, Baltimore, MD*

Grossman, Jerrold W., *Oakland University, Rochester, MI*

Grunbaum, Branko, *University of Washington, Seattle, WA*

Guy, Richard K., *University of Calgary, Calgary, Canada*

Hamburger, Peter, *Indiana University-Purdue, Fort Wayne, IN*

Hayashi, Elmer K., *Wake Forest University, Winston-Salem, NC*

Henle, Jim, *Smith College, Northhampton, MA*

Hodge, Jonathan, *Grand Valley State University, Allendale, MI*

Hoehn, Larry P., *Austin Peay State University, Clarksville, TN*

Horn, Roger, *University of Utah, Salt Lake City, UT*

Hull, Thomas, *Merrimack College, North Andover, MA*

Isaac, Richard, *Herbert H. Lehman College CUNY, Bronx, NY*

Janke, Steven J., *Colorado College, Colorado Springs, CO*

Jepsen, Charles H., *Grinnell College, Grinnell, IA*

Johnson, Brenda, *Union College, Schenectady, NY*

Johnson, Diane L., *Humboldt State University, Arcata, CA*

Johnson, Warren P., *Connecticut College, New London, CT*

Kallaher, Michael J., *Washington State University, Pullman, WA*

Kerckhove, Michael G., *University of Richmond, Richmond, VA*

Kimberling, Clark, *University of Evansville, Evansville, IN*

Kleiner, Israel, *York University, North York, Canada*

Kuzmanovich, James J., *Wake Forest University, Winston-Salem, NC*

Kwong, Harris, *State University of New York at Fredonia, Fredonia, NY*

Lang, Robert J., *Alamo, CA*

Larson, Dean S., *Gonzaga University, Spokane, WA*

Lautzenheiser, Roger G., *Rose-Hulman Institute of Technology, Terre Haute, IN*

Ledyaev, Yuri, *Western Michigan University, Kalamazoo, MI*

Loud, Warren S., *Minneapolis, MN*

Mackin, Gail, *Northern Kentucky University, Highland Heights, KY*

Mazur, Mark S., *Duquesne University, Pittsburgh, PA*

McCleary, John H., *Vassar College, Poughkeepsie, NY*

Merrill, Kathy D., *Mancos, CO*

Monson, Barry, *University of New Brunswick, Fredericton, Canada*

Morics, Steven W., *University of Redlands, Redlands, CA*

Moses, Peter, *Moparmatic Co., Astwood Bank, UK*

Neidinger, Richard D., *Davidson College, Davidson, NC*

Nelsen, Roger B., *Lewis & Clark College, Portland, OR*

Nelson, Donald J., *Western Michigan University, Kalamazoo, MI*

Ostrov, Daniel N., *Santa Clara University, Santa Clara, CA*

Otero, Daniel E., *Xavier University, Cincinnati, OH*

Pfiefer, Richard E., *San José State University, San Jose, CA*

Ratliff, Thomas C., *Wheaton College, Norton, MA*

Ross, Kenneth A., *University of Oregon, Eugene, OR*

Ryan, Richard F., *Marymount College, Rancho Palos Verdes, CA*

Sander, Evelyn, *George Mason University, Fairfax, VA*

Savage, Carla, *North Carolina State, Raleigh, NC*

Schaefer, Edward F., *Santa Clara University, Santa Clara, CA*

Scheinerman, Edward S., *Johns Hopkins University, Baltimore, MD*

Schumer, Peter D., *Middlebury College, Middlebury, VT*

Schwenk, Allen J., *Western Michigan University, Kalamazoo, MI*

Scott, Richard A., *Santa Clara University, Santa Clara, CA*

Sepanski, Peter D., *Western Kentucky University, Bowling Green, KY*

Shell-Gellasch, Amy E., *Grafenwoehr, Germany*

Sherman, Gary, *Rose-Hulman Institute of Technology, Terre Haute, IN*

Shifrin, Theodore, *University of Georgia, Athens, GA*

Sloughter, Daniel C., *Furman University, Greenville, SC*

Solheid, Ernie S., *California State University, Fullerton, Fullerton, CA*

Stockmeyer, Paul, *College of William and Mary, Williamsburg, VA*

Stone, Alexander P., *University of New Mexico, Albuquerque, NM*

Straffin, Philip D., *Beloit College, Beloit, WI*

Szabó, Tamás, *Weber State University, Ogden, UT*

Takács, Lajos F., *Case Western Reserve University, Cleveland Heights, OH*

Teets, Donald, *South Dakota School of Mines and Technology, Rapid City, SD*

Thomas, Hugh, *University of New Brunswick, Fredericton, Canada*

Thrall, Anthony, *Stanford Alumni Organization, Menlo Park, CA*

Villalobos, Ph. D., Cristina, *The University of Texas-Pan American, Edinburg, TX*

Walden, Byron, *Santa Clara University, Santa Clara, CA*

White, Arthur T., *Western Michigan University, Kalamazoo, MI*

Worner, Tamara S., *Wayne State College, Wayne, NE*

Zeleke, Melkamu, *William Patterson University, Wayne, NJ*

Zhang, Ping, *Western Michigan University, Kalamazoo, MI*

Zhu, Qiji J., *Western Michigan University, Kalamazoo, MI*

Index to Volume 78

AUTHORS

TITLES

PROBLEMS

The letters P, Q, and S refer to Proposals, Quickies, and Solutions, respectively; page numbers appear in parentheses. For example, P1622(155) refers to Proposal 1622, which appears on page 155.

Thanks, too, to the following colleagues who have helped with refereeing of problems: Steve Dunbar, A. M. Fink, Gary Lieberman, Paul Sacks, and Ananda Weerasinghe.

A Letter from the Editor

Dear Readers:

With this issue, I complete my term as Editor of the MAGAZINE and turn over our operations to the capable hands of Editor-Elect Allen Schwenk, whose February 2006 issue is already underway.

I am grateful for the opportunity to serve as Editor, a job I have found challenging and rewarding. The editors who served before me, and the authors who wrote for them, set a remarkable standard of quality, which I have done my best to maintain.

Copious thanks are due to many people, especially our authors—416 of them over the past five years. My excellent Editorial Assistant, Martha Giannini, kept MAGAZINE operations at Santa Clara University in fine running order. Managing Editor Harry Waldman did the same for our business at MAA headquarters. The wizards at Integre Technical Publishing, especially composatrix Dianne Parish, made our issues look dapper. Over the past summer, Keith Thompson, a recent Santa Clara graduate, made a significant contribution as Student Editorial Assistant. Our expert board of Associate Editors, most notably Problems Editor Elgin Johnston and Reviews Editor Paul Campbell, deserve a great deal of credit.

My wish is that our world-wide community will continue to read the MAGAZINE with passion and enthusiasm for many volumes to come. Sometimes we academics get the message that our highest priority should be to write. May the MAGAZINE remind us all how important it is to read.

Frank A. Farris
Santa Clara University

United States Postal Service

Statement of Ownership, Management, and Circulation

1 Publication Title	2 Publication Number	3 Filing Date
Mathematics Magazine	0025-570X	9/8/05
4 Issue Frequency	**5 Number of Issues Published Annually**	**6 Annual Subscription Price**
Bimonthly, except July and August	5	$144

7 Complete Mailing Address of Known Office of Publication *(Not printer) (Street, city, county, state, and ZIP+4)*
Mathematical Association of America, 1529 18th St, NW Washington, DC 20036
Contact Person: Harry Waldman
Telephone: 202-387-5200

8 Complete Mailing Address of Headquarters or General Business Office of Publisher *(Not printer)*
Mathematical Association of America, 1529 18th St, NW Washington, DC 20036

9 Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do not leave blank)*

Publisher *(Name and complete mailing address)*
Mathematical Association of America, 1529 18th St, NW Washington, DC 20036

Editor *(Name and complete mailing address)*
Frank Farris, Santa Clara U, Santa Clara, CA 95053

Managing Editor *(Name and complete mailing address)*
Harry Waldman
Mathematical Association of America, 1529 18th St, NW Washington, DC 20036

10 Owner *(Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)*

Full Name	Complete Mailing Address
Mathematical Association of America	1529 18th St, NW Washington, DC 20036

11 Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → ☒ None

Full Name	Complete Mailing Address

12 Tax Status *(For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)*
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes
☒ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months *(Publisher must submit explanation of change with this statement)*

PS Form **3526**, October 1999 *(See Instructions on Reverse)*

13 Publication Title	14 Issue Date for Circulation Data Below
Mathematics Magazine	June 2005

15 Extent and Nature of Circulation		Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a Total Number of Copies *(Net press run)*		14,000	15,000
b Paid and/or Requested Circulation	(1) Paid/Requested Outside-County Mail Subscriptions Stated on Form 3541 *(Include advertiser's proof and exchange copies)*	12,500	13,500
	(2) Paid In-County Subscriptions Stated on Form 3541 *(Include advertiser's proof and exchange copies)*	0	0
	(3) Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Non-USPS Paid Distribution	0	0
	(4) Other Classes Mailed Through the USPS	0	0
c Total Paid and/or Requested Circulation *[Sum of 15b (1), (2),(3),and (4)]* ▶		12,500	13,500
d Free Distribution by Mail *(Samples, complimentary, and other free)*	(1) Outside-County as Stated on Form 3541	0	0
	(2) In-County as Stated on Form 3541	0	0
	(3) Other Classes Mailed Through the USPS	1,000	1,000
e Free Distribution Outside the Mail *(Carriers or other means)*		0	0
f Total Free Distribution *(Sum of 15d and 15e)* ▶		1,000	1,000
g Total Distribution *(Sum of 15c and 15f)* ▶		13,500	14,500
h Copies not Distributed		500	500
i Total *(Sum of 15g and h)* ▶		14,000	15,000
j Percent Paid and/or Requested Circulation *(15c divided by 15g times 100)*		90%	90%

16 Publication of Statement of Ownership
☒ Publication required. Will be printed in the Dec. 2005 issue of this publication ☐ Publication not required

17 Signature and Title of Editor, Publisher, Business Manager, or Owner: Donald J. Albers Date: 9/8/05

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

Instructions to Publishers

1 Complete and file one copy of this form with your postmaster annually on or before October 1. Keep a copy of the completed form for your records.

2 In cases where the stockholder or security holder is a trustee, include in items 10 and 11 the name of the person or corporation for whom the trustee is acting. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing corporation. In item 11, if none, check the box. Use blank sheets if more space is required.

3 Be sure to furnish all circulation information called for in item 15. Free circulation must be shown in items 15d, e, and f.

4 Item 15h, Copies not Distributed, must include (1) newsstand copies originally stated on Form 3541, and returned to the publisher, (2) estimated returns from news agents, and (3), copies for office use, leftovers, spoiled, and all other copies not distributed.

5 If the publication had Periodicals authorization as a general or requester publication, this Statement of Ownership, Management, and Circulation must be published, it must be printed in any issue in October or, if the publication is not published during October, the first issue printed after October.

6 In item 16, indicate the date of the issue in which this Statement of Ownership will be published.

7 Item 17 must be signed.

Failure to file or publish a statement of ownership may lead to suspension of Periodicals authorization.

PS Form **3526**, October 1999 *(Reverse)*

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW
Washington, DC 20036