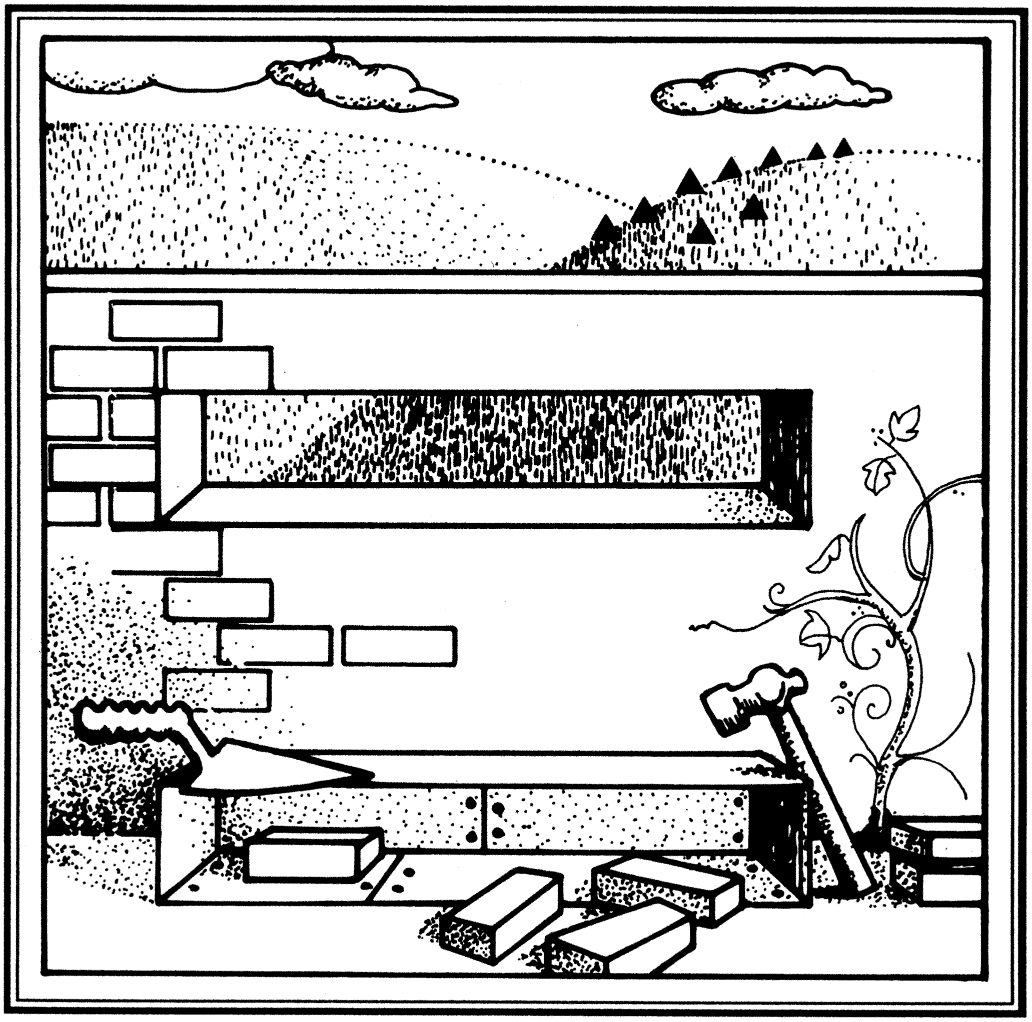
Vol. 64, No. 2 April 1991

MATHEMATICS MAGAZINE



- Plane Curves, Polar Coordinates, and Winding Numbers
- Packing a Box with Bricks
- Complements of Linear Subspaces

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

The aim of *Mathematics Magazine* is to provide lively and appealing mathematical exposition. This is not a research journal and, in general, the terse style appropriate for such a journal (lemma-theorem-proof-corollary) is not appropriate for an article for the *Magazine*. Articles should include examples, applications, historical background, and illustrations, where appropriate. They should be attractive and accessible to undergraduates and would, ideally, be helpful in supplementing undergraduate courses or in stimulating student investigations. Articles on pedagogy alone, unaccompanied by interesting mathematics, are not suitable. Neither are articles consisting mainly of computer programs unless these are essential to the presentation of some good mathematics. Manuscripts on history are especially welcome, as are those showing relationships between various branches of mathematics and between mathematics and other disciplines.

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

Send new manuscripts to: Martha Siegel, Editor, Mathematics Magazine, Towson State University, Towson, MD 21204. Manuscripts should be typewritten and double spaced and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit the original and two copies and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Illustrations should be carefully prepared on separate sheets in black ink, the original without lettering and two copies with lettering added. Do not use staples.

Cover illustration: Carolyn Westbrook

AUTHORS

John Baker received his B.A. and M.A. from the University of Saskatchewan and his Ph.D. from the University of Waterloo. His interest in mathematics was kindled by his father who had not been to a university but who understood the nature of mathematics and could communicate his enthusiasm. Other major influences on the author were his calculus teacher, K. G. Toews, and his colleague, Mike McKiernan. The present article has its roots in Dr. Baker's attempt, many years ago, to explain to his students how the giants of analysis might have thought about winding numbers.

Vol. 64, No. 2 April 1991

MATHEMATICS MAGAZINE

The *MATHEMATICS MAGAZINE* (ISSN 0025–570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.
The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $16 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $68 per year.
Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

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

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

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

PRINTED IN THE UNITED STATES OF AMERICA

ARTICLES

Plane Curves, Polar Coordinates and Winding Numbers

JOHN A. BAKER
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

This article begins with an elementary proof of (and motivation for) the fact that a plane curve can be suitably represented in polar coordinates. Our aim is to demonstrate that this fact can be used

(i) to give natural definitions of "winding number" and "degree" (in two dimensions); and thereby,

(ii) to deduce, with relative ease, several important, well-known results from diverse areas of mathematics including real and complex analysis, algebra, plane topology and differential equations.

Among these results are the fundamental theorem of algebra, Brouwer's Fixed Point Theorem in the plane, an open mapping theorem in two dimensions, and assertions concerning oscillation and stability of differential equations.

1. Introduction

Suppose γ is a (plane) *curve*, by which we mean a continuous function, $\gamma: [a, b] \to \mathbb{R}^2$, where $a, b \in \mathbb{R}$ (the set of all real numbers) and $a < b$. Intuitively, we may think of γ as describing the motion of a "particle" moving in $\mathbb{R}^2$, $\gamma(t)$ being its position at "time" t for t in the "time interval" $[a, b]$. Thus γ is not a subset of $\mathbb{R}^2$ although useful information concerning γ can be discerned from its range, $\{\gamma(t): a \leqslant t \leqslant b\}$, which we will denote by $[\gamma]$. Let $\gamma(t) = (x(t), y(t))$ for $a \leqslant t \leqslant b$ so that x and y are continuous real-valued functions defined on $[a, b]$. We say that γ is a C^1 curve provided both x and y are continuously differentiable on $[a, b]$ (with one-sided derivatives at a and b). Such functions are said to be C^1. In this case we let $\gamma'(t) = (x'(t), y'(t))$ for $a \leqslant t \leqslant b$ and we call γ' the *derivative* of γ. We say γ is *piecewise* C^1 provided there exists a partition of $[a, b]$ such that the restriction of γ to each of the associated closed subintervals is C^1 ($a = a_0 < a_1 < a_2 < \cdots < a_n = b$ with $\gamma|_{[a_{j-1}, a_j]}$ a C^1 curve for each $j = 1, 2, \ldots, n$). For elementary complex analysis it suffices to consider curves that are piecewise C^1. We will identify $\mathbb{R}^2$ with $\mathbb{C}$, the complex plane, mainly because of the advantages of Euler's formula: $e^{it} = (\cos t, \sin t)$ for $t \in \mathbb{R}$.

Suppose $\gamma: [a, b] \to \mathbb{C}$ is a *closed* curve ($\gamma(a) = \gamma(b)$), $p \in \mathbb{C}$ and p is *not on* γ, i.e., $p \notin [\gamma]$. In most contemporary, rigorous accounts of complex analysis the important concept of the *winding number* (or index) of γ with respect to p, $W(\gamma; p)$, is defined (in case γ is piecewise C^1) by a complex line integral:

$$W(\gamma;p)=\frac{1}{2\pi i}\int_{\gamma}\frac{dz}{z-p}. \tag{1}$$

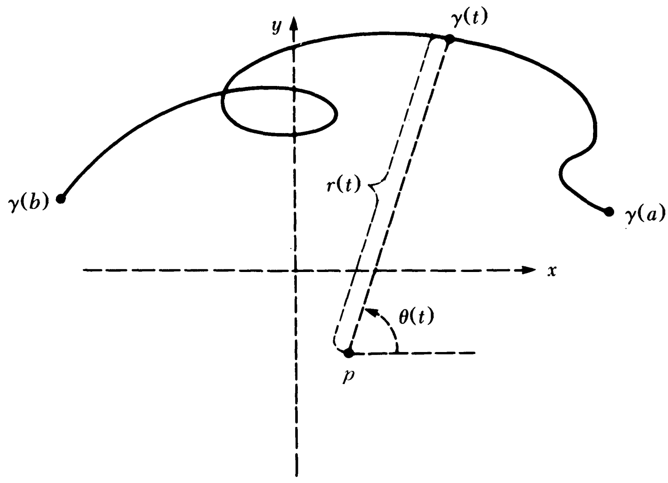
While concise, elegant and useful, this definition disguises the intuitive geometric idea that the terminology is intended to suggest, namely that $W(\gamma;p)$ is an integer (!), which indicates the net number of times our particle "winds around" or "revolves about" p. Moreover the sign of $W(\gamma;p)$ tells us whether the net effect has been counterclockwise (positive) or clockwise (negative). This definition of winding number is unnecessarily complicated and may convey the false impression that it depends in some essential way on complex analysis.

An alternative definition of winding number can be based on the following intuitively evident observation: If $\gamma\colon[a,b]\to\mathbb{C}$ is a curve, $p=(x_0,y_0)\in\mathbb{C}$ and p is not on γ then there exist continuous functions $r,\theta\colon[a,b]\to\mathbb{R}$ such that $r(t)>0$ and

$$\gamma(t)=(x_0+r(t)\cos\theta(t),y_0+r(t)\sin\theta(t))\quad\text{for } a\leqslant t\leqslant b \tag{2}$$

or, by Euler's formula,

$$\gamma(t)=p+r(t)\exp(i\theta(t))\quad\text{for } a\leqslant t\leqslant b. \tag{2$'$}$$



Thus $r(t)$ and $\theta(t)$ are polar coordinates of $\gamma(t)-p$ for $a\leqslant t\leqslant b$. We will refer to (2) or (2)′ as a *polar representation* of γ. The requirement that r be positive forces us to choose $r(t)=|\gamma(t)-p|$ for $t\in[a,b]$. The only nontrivial thing to check in order to prove the existence of a polar representation is that θ can be chosen to be continuous. Although θ is not uniquely determined by γ and p, it will be easy to show that the continuity assumption determines it "up to a multiple of 2π." If γ is closed then $r(a)\exp(i\theta(a))=r(b)\exp(i\theta(b))$; hence $r(a)=r(b)$ and $\theta(b)-\theta(a)$ is an integer multiple of 2π; moreover this integer is independent of the choice of θ. When γ is closed we may therefore unambiguously define

$$W(\gamma;p)=\frac{\theta(b)-\theta(a)}{2\pi} \tag{3}$$

and the geometric meaning is clear. This approach is followed by Beardon [2] and Burckel [4]. Beardon deduces the polar representation by "piecing together" branches of the argument; Burckel uses branches of the complex logarithm. In both cases a little compactness argument is involved. With (3) as the definition, (1) follows quickly from (2)′ and the definition of complex line integral when γ is piecewise C^1.

In this article we will show how the polar representation of a C^1 curve can be easily deduced from the fundamental theorem of calculus and elementary properties of the sine and cosine. In fact, we don't need the interval of definition of γ to be compact. This fact, as we shall illustrate, is useful in the study of differential equations.

The existence of a polar representation of a piecewise C^1 curve follows easily from the C^1 case. To get from the piecewise C^1 case to the general case we will use an approximation argument. This allows us to construct a "degree theory" in two dimensions. This theory yields a generalization of Rouché's Theorem from which it is possible (as in [5]) to easily deduce the fundamental theorem of algebra and Brouwer's fixed point theorem in the plane. We will also relate these ideas to an open mapping theorem and an inverse function theorem in two dimensions and raise a question concerning the inverse function theorem.

I doubt that there is anything new in this article. I hope, however, it will show that these ideas, while intuitive and not difficult to prove, have important implications for many parts of mathematics and deserve to be more widely known. I also hope that this article will stimulate interest in degree theory and shed light on real and complex analysis.

For $x=(x_1,x_2,\dots,x_n)$ in $\mathbb{R}^n$, let $|x|=(x_1^2+\cdots+x_n^2)^{1/2}$.

2. The Polar Representation

The development rests on the following result.

PROPOSITION 1. *Suppose I is a nondegenerate interval in $\mathbb{R}$ and $x, y\colon I\to\mathbb{R}$ are C^1 functions such that $(x(t),y(t))\neq(0,0)$ for all $t\in I$. Suppose that $t_0\in I$, $r_0>0$ and $\theta_0\in\mathbb{R}$ are chosen so that $x(t_0)=r_0\cos\theta_0$ and $y(t_0)=r_0\sin\theta_0$. Then there exist C^1 functions $r,\theta\colon I\to\mathbb{R}$ such that $r(t_0)=r_0$, $\theta(t_0)=\theta_0$, $r(t)>0$ and*

$$x(t)=r(t)\cos\theta(t) \quad \textit{and} \quad y(t)=r(t)\sin\theta(t) \quad \textit{for all } t\in I. \tag{4}$$

Proof. If such r and θ exist then (4) and the positivity of r imply that

$$r(t)=\left(x(t)^2+y(t)^2\right)^{1/2} \quad \text{for all } t\in I.$$

So define r accordingly; note that r is C^1 and $r(t_0)=r_0$.

Now let $u(t)=x(t)/r(t)$ and $v(t)=y(t)/r(t)$ for $t\in I$ so that u and v are C^1 functions on I, $u(t_0)=\cos\theta_0$, $v(t_0)=\sin\theta_0$ and

$$u(t)^2+v(t)^2=1 \quad \text{for all } t\in I. \tag{5}$$

We want to find a C^1 function $\theta\colon I\to\mathbb{R}$ such that

$$u(t)=\cos\theta(t) \quad \text{and} \quad v(t)=\sin\theta(t) \quad \text{for all } t\in I \tag{6}$$

and $\theta(t_0)=\theta_0$. If such a θ exists then (6) implies

$$u'(t)=-v(t)\theta'(t) \quad \text{and} \quad v'(t)=u(t)\theta'(t) \quad \text{for } t\in I. \tag{7}$$

If we multiply the first equation in (7) by $v(t)$, the second by $u(t)$ and subtract the resulting equations we find, in light of (5), that

$$\theta'(t) = -v(t)u'(t) + u(t)v'(t) \quad \text{for } t \in I. \tag{8}$$

But we want $\theta(t_0) = \theta_0$. Thus, according to the Fundamental Theorem of Calculus, we must have

$$\theta(t) = \theta_0 + \int_{t_0}^{t} \{-v(s)u'(s) + u(s)v'(s)\}\, ds \quad \text{for } t \in I. \tag{9}$$

We thus define θ by (9), noting that θ is C^1, $\theta(t_0) = \theta_0$ and that (8) holds. From (5) it follows that

$$u(t)u'(t) + v(t)v'(t) = 0 \quad \text{for all } t \in I. \tag{10}$$

Viewing (8) and (10) as a pair of linear equations in $u'(t)$ and $v'(t)$ and recalling (5) we find that (7) holds.

Let $g(t) = (u(t) - \cos\theta(t))^2 + (v(t) - \sin\theta(t))^2$ for $t \in I$. It follows from (7) and (10) that $g'(t) = 0$ for all $t \in I$. But $g(t_0) = 0$ so that $g(t) = 0$ for all $t \in I$. Thus (6) holds.

In most treatments of complex analysis in which the winding number is defined by a line integral, the definition is followed immediately by a proof of the fact that it is indeed an integer; this proof typically uses similar ideas to those used in the above proof. See, for example, [1] or [17].

Proposition 1 suffices for the applications to differential equations included at the end of this article. It allows us to derive the polar representation for piecewise C^1 curves with ease. From now on we will use the Euler notation when convenient.

Proposition 2. *Suppose $\gamma: [a,b] \to \mathbb{C}$ is a piecewise C^1 curve, $p \in \mathbb{C}$ and p is not on γ. Then there exist piecewise C^1 functions $r, \theta: [a,b] \to \mathbb{R}$ such that $r(t) > 0$ and*

$$\gamma(t) = p + r(t)\exp(i\theta(t)) \quad \textit{for all } t \in [a,b].$$

Proof. Let $(x(t), y(t)) = \gamma(t) - p$ and let $r(t) = (x(t)^2 + y(t)^2)^{1/2}$ for $a \leqslant t \leqslant b$. Let $a = a_0 < a_1 < a_2 < \dots < a_n = b$ such that the restrictions of x and y to $[a_{j-1}, a_j]$ are C^1 for each $j = 1, 2, \dots, n$. According to Proposition 1 there exists a C^1 function θ_1: $[a_0, a_1] \to \mathbb{R}$ such that $\gamma(t) - p = r(t)\exp(i\theta_1(t))$ for $a_0 \leqslant t \leqslant a_1$. Using Proposition 1 again we deduce that there exists a C^1 function θ_2: $[a_1, a_2] \to \mathbb{R}$ such that $\theta_2(a_1) = \theta_1(a_1)$ and

$$\gamma(t) - p = r(t)\exp(i\theta_2(t)) \quad \text{for } a_1 \leqslant t \leqslant a_2.$$

By induction, for each $j = 1, 2, \dots, n$ there exists a C^1 function θ_j: $[a_{j-1}, a_j] \to \mathbb{R}$ such that $\theta_j(a_{j-1}) = \theta_{j-1}(a_{j-1})$ for $1 < j \leqslant n$ and

$$\gamma(t) - p = r(t)\exp\big(i\theta_j(t)\big) \quad \text{for } a_{j-1} \leqslant t \leqslant a_j.$$

The existence assertion now follows by defining

$$\theta(t) = \theta_j(t) \text{ in case } a_{j-1} \leqslant t \leqslant a_j \text{ for some } j = 1, 2, \dots, n.$$

To obtain the polar representation of an arbitrary curve we will use the following lemmas.

LEMMA 1. *Suppose* $s, t \in \mathbb{R}$, $|e^{is} - e^{it}| < \varepsilon < 1$ *and* $|s - t| < \pi$. *Then* $|s - t| < \varepsilon\pi/2$.

Proof. We will use the elementary fact that $2|x|/\pi \leqslant |\sin x| \leqslant |x|$ for $-\pi/2 \leqslant x \leqslant \pi/2$. If $\pi/2 < |s - t| < \pi$ then $\cos(s - t) < 0$ so that $|e^{is} - e^{it}| = |e^{i(s-t)} - 1| \geqslant 1 - \cos(s - t) > 1$. Hence it is not possible to have $\pi/2 < |s - t| \leqslant \pi$. Since $|s - t| \leqslant \pi/2$ we must have $|\sin(s - t)| \leqslant |(\cos(s - t) - 1) + i\sin(s - t)| = |e^{i(s-t)} - 1| = |e^{is} - e^{it}| < \varepsilon$ and hence $2|s - t|/\pi \leqslant |\sin(s - t)| < \varepsilon$.

LEMMA 2. *Suppose* γ: $[a, b] \to \mathbb{C}$ *is a curve and* $\varepsilon > 0$. *Then there exists a piecewise* C^1 *curve* η: $[a, b] \to \mathbb{C}$ *such that* $\gamma(a) = \eta(a)$, $\gamma(b) = \eta(b)$ *and* $|\gamma(t) - \eta(t)| < \varepsilon$ *for all* $t \in [a, b]$.

Proof. Since γ is continuous and $[a, b]$ is compact, γ is uniformly continuous. Choose $\delta > 0$ so that $|\gamma(s) - \gamma(t)| < \varepsilon/2$ whenever $s, t \in [a, b]$ and $|s - t| \leqslant \delta$. Let $a = a_0 < a_1 < a_2 < \cdots < a_n = b$ such that $a_j - a_{j-1} < \delta$ for $1 \leqslant j \leqslant n$. Define η: $[a, b] \to \mathbb{C}$ by

$$\eta(t) = \gamma(a_{j-1}) + \left(\frac{t - a_{j-1}}{a_j - a_{j-1}}\right)(\gamma(a_j) - \gamma(a_{j-1}))$$

in case $a_{j-1} \leqslant t \leqslant a_j$ for some $j = 1, 2, \ldots, n$. The curve η has the desired properties; it was constructed from γ by a standard construction known as piecewise linear (some would say affine) interpolation.

PROPOSITION 3. *Suppose* γ: $[a, b] \to \mathbb{C}$ *is a curve,* $p \in \mathbb{C}$ *and* p *is not on* γ. *Then there exist continuous functions* r, θ: $[a, b] \to \mathbb{R}$ *such that* $r(t) > 0$ *and*

$$\gamma(t) = p + r(t)\exp(i\theta(t)) \quad \text{for } a \leqslant t \leqslant b.$$

Moreover, r *is unique and* θ *is determined up to a multiple of* 2π.

Proof. Choose $r_0 > 0$ and $\theta_0 \in \mathbb{R}$ so that $\gamma(a) - p = r_0 \exp(i\theta_0)$ and let $r(t) = |\gamma(t) - p| > 0$ for $a \leqslant t \leqslant b$.

For each natural number n, choose a piecewise C^1 curve γ_n: $[a, b] \to \mathbb{C}$ such that $\gamma_n(a) = \gamma(a)$, $\gamma_n(b) = \gamma(b)$ and $|\gamma_n(t) - \gamma(t)| < 1/n$ for all $t \in [a, b]$. Thus $\{\gamma_n\}_{n=1}^{\infty}$ converges uniformly to γ on $[a, b]$. Hence $|\gamma_n(t) - p| \to |\gamma(t) - p| = r(t)$ as $n \to \infty$ uniformly for $t \in [a, b]$ and so there exists a natural number N and a positive real number ε_0 such that

$$|\gamma_n(t) - p| > \varepsilon_0 > 0 \quad \text{for all } t \in [a, b] \quad \text{provided } n > N.$$

By Proposition 2, for each $n > N$, we may choose a piecewise C^1 function θ_n: $[a, b] \to \mathbb{R}$ such that $\theta_n(a) = \theta_0$ and

$$\gamma_n(t) - p = |\gamma_n(t) - p|\exp(i\theta_n(t)) \quad \text{for } a \leqslant t \leqslant b.$$

We claim that $\{\theta_n\}_{n>N}$ converges uniformly on $[a, b]$. To see this, we will prove that it is a uniform Cauchy sequence. Let $0 < \varepsilon < 1$. Since

$$\exp(i\theta_n(t)) = \frac{\gamma_n(t) - p}{|\gamma_n(t) - p|} \to \frac{\gamma(t) - p}{|\gamma(t) - p|} \quad \text{as } n \to \infty$$

uniformly on $[a, b]$, there exists a natural number $M \geqslant N$ such that

$$|\exp(i\theta_n(t)) - \exp(i\theta_m(t))| < \varepsilon < 1 \quad \text{for all } t \in [a, b] \qquad (*)$$

provided $m, n > M$. But $\theta_n(a) = \theta_m(a) = \theta_0$ for $m, n > N$ and hence we must have

$$|\theta_n(t) - \theta_m(t)| < \pi \quad \text{for } t \in [a, b] \quad \text{and} \quad m, n > M:$$

otherwise, by the intermediate value theorem, there would exist a τ in $[a, b]$ such that

$$|\theta_n(\tau) - \theta_m(\tau)| = \pi$$

thereby contradicting $(*)$.

By Lemma 1, $|\theta_n(t) - \theta_m(t)| < \pi\varepsilon/2$ for $t \in [a, b]$ and $m, n > M$. Thus $\{\theta_n\}_{n > M}$ converges uniformly on $[a, b]$ to a continuous function θ: $[a, b] \to \mathbb{R}$ and, for all $t \in [a, b]$,

$$\gamma(t) - p = \lim_{n \to \infty} \gamma_n(t) - p = \lim_{n \to \infty} |\gamma_n(t) - p| \exp(i\theta_n(t)) = r(t)\exp(i\theta(t)).$$

To prove the uniqueness assertion first note that, since $r(t) > 0$, we must have $r(t) = |\gamma(t) - p|$ for all $t \in [a, b]$.

Suppose ω: $[a, b] \to \mathbb{R}$ is continuous and

$$\gamma(t) = p + r(t)\exp(i\omega(t)) \quad \text{for all } t \in [a, b].$$

Then, for each $t \in [a, b]$, $\exp(i\theta(t)) = \exp(i\omega(t))$ and hence $\theta(t) - \omega(t) = 2\pi k(t)$ for some integer $k(t)$. Since θ and ω are continuous, so is k. But k is also integer valued. Hence, by the intermediate value theorem, k is constant.

3. The Winding Number

Suppose γ: $[a, b] \to \mathbb{R}^2$ is a closed curve, $p \in \mathbb{R}^2$ and p is not on γ. Choose continuous functions r, θ: $[a, b] \to \mathbb{R}$ such that $r(t) > 0$ and

$$\gamma(t) = p + r(t)\exp(i\theta(t)) \quad \text{for } a \leqslant t \leqslant b.$$

Since $\gamma(a) = \gamma(b)$, $\theta(b) - \theta(a)$ is an integer multiple of 2π. Moreover, if ω: $[a, b] \to \mathbb{R}$ is continuous and such that

$$\gamma(t) = p + r(t)\exp(i\omega(t)) \quad \text{for } a \leqslant t \leqslant b$$

then, according to Proposition 3, there exists an integer, k, such that $\omega(t) = \theta(t) + 2\pi k$ for all $t \in [a, b]$. Hence $\theta(b) - \theta(a) = \omega(b) - \omega(a)$ and we may therefore unambiguously define $W(\gamma; p)$, *the winding number of* γ *with respect to* p, by $W(\gamma; p) = [\theta(b) - \theta(a)]/2\pi$. As noted above, the winding number is an integer.

The following properties of winding numbers follow directly from the definition of winding number (and, in case of $W.3$, the definitions of real and complex line integrals). In these assertions γ: $[a, b] \to \mathbb{R}^2$ is a closed curve, $p \in \mathbb{R}^2$ and p is not on γ.

$W.1$. If γ is constant (i.e., there is a q in $\mathbb{R}^2$ such that $\gamma(t) = q$ for all $t \in [a, b]$) then $W(\gamma; p) = 0$.

$W.2$. If $a = 0, b = 2\pi, n$ is an integer, $\rho > 0$ and $\gamma(t) = p + \rho \exp(int)$ for $0 \leqslant t \leqslant 2\pi$ then $W(\gamma; p) = n$.

$W.3$. If γ is piecewise C^1 and $p=(\alpha,\beta)$ then

$$W(\gamma;p)=\frac{1}{2\pi i}\int_\gamma\frac{dz}{z-p}=\frac{1}{2\pi}\int_\gamma\frac{-(y-\beta)\,dx+(x-\alpha)\,dy}{(x-\alpha)^2+(y-\beta)^2}.$$

$W.4$. If $\phi\colon[c,d]\to[a,b]$ is continuous, $\phi(c)=a$, $\phi(d)=b$ and $\eta(s)=\gamma(\phi(s))$ for $c\leqslant s\leqslant d$ then p is not on η and

$$W(\eta;p)=W(\gamma;p).$$

$W.5$. If we define γ^- (the *opposite* of γ) by $\gamma^-(t)=\gamma(a+b-t)$ for $a\leqslant t\leqslant b$ then $[\gamma]=[\gamma^-]$, p is not on γ^- and $W(\gamma^-;p)=-W(\gamma;p)$.

$W.6$. If we define $(\gamma-p)(t)=\gamma(t)-p$ for $t\in[a,b]$ then 0 is not on $\gamma-p$ and $W(\gamma;p)=W(\gamma-p;0)$.

If in $W.4$ ϕ is also one-to-one (or, equivalently, ϕ is strictly increasing) then we say that η is equivalent to γ and conclude that equivalent curves have equal winding numbers. In this case η is often referred to as a reparametrization of γ.

THEOREM 1. *Suppose $\gamma,\eta\colon[a,b]\to\mathbb{R}^2$ are closed curves, $p\in\mathbb{R}^2$ and*

$$|\gamma(t)-\eta(t)|<|\gamma(t)-p|+|\eta(t)-p|\quad \textit{for all } t\in[a,b].$$

Then p is not on either γ or η and $W(\gamma;p)=W(\eta;p)$.

To understand the intuitive meaning of this theorem let us observe what the inequality means geometrically. If we write the inequality as

$$|(\gamma(t)-p)-(\eta(t)-p)|<|\gamma(t)-p|+|\eta(t)-p|$$

(strict inequality in the triangle inequality) we see that it means that the vector $\gamma(t)-p$ is not a negative multiple of the vector $\eta(t)-p$ or, equivalently, for no t does p lie on the line segment joining $\gamma(t)$ to $\eta(t)$. Thus if γ describes the motion of a particle A and η describes the motion of a particle B, it is intuitively evident that A winds around p the same number of times as does B.

Proof. If $\gamma(t)=p$ for some $t\in[a,b]$ then the inequality implies $|p-\eta(t)|<|\eta(t)-p|$. Hence p cannot be on γ and, similarly, p is not on η.

Choose continuous functions $r,\rho,\theta,\omega\colon[a,b]\to\mathbb{R}$ such that $r(t)>0$, $\rho(t)>0$, $\gamma(t)=p+r(t)\exp(i\theta(t))$ and $\eta(t)=p+\rho(t)\exp(i\omega(t))$ for all $t\in[a,b]$. Let $\phi(t)=\theta(t)-\omega(t)$ for $a\leqslant t\leqslant b$.

We claim that $\phi(t)\neq k\pi$ for all $t\in[a,b]$ and all odd integers k. For suppose $\phi(\tau)=k\pi$ for some $\tau\in[a,b]$ and some odd integer k. Then $\theta(t)=\omega(\tau)+k\pi$ so that $\exp(i\theta(\tau))=-\exp(i\omega(\tau))$ and hence

$$\begin{aligned}|\gamma(\tau)-\eta(\tau)|&=|r(\tau)\exp(i\theta(\tau))+\rho(\tau)\exp(i\theta(\tau))|\\&=r(\tau)+\rho(\tau)=|\gamma(\tau)-p|+|\eta(\tau)-p|,\end{aligned}$$

contradicting our assumption.

Since ϕ is continuous and its range contains no odd multiples of π, it follows from the intermediate value theorem that the range of ϕ must be contained in an interval of length less than 2π. Thus $|\phi(b)-\phi(a)|<2\pi$ so that

$$|W(\gamma;p)-W(\gamma;p)|=\left|\frac{\theta(b)-\theta(a)}{2\pi}-\frac{\omega(b)-\omega(a)}{2\pi}\right|=\left|\frac{\phi(b)-\phi(a)}{2\pi}\right|<1.$$

But $W(\gamma; p)$ and $W(\eta; p)$ are integers and hence are equal.

COROLLARY 1. *Suppose* $\gamma\colon [a,b] \to \mathbb{R}^2$ *is a closed curve,* $p, q \in \mathbb{R}^2$ *and* $|p-q| < |\gamma(t) - p|$ *for all* $t \in [a,b]$. *Then* $p \notin [\gamma]$, $q \notin [\gamma]$, *and* $W(\gamma; p) = W(\gamma; q)$.

Proof. Since $|(\gamma(t)-q)-(\gamma(t)-p)| < |\gamma(t)-q| + |\gamma(t)-p|$ for all $t \in [a,b]$, it follows from Theorem 1 that $W(\gamma - q; 0) = W(\gamma - p; 0)$ and the rest follows from property $W.6$.

COROLLARY 2. *Suppose that for each natural number* n, $\gamma_n\colon [a,b] \to \mathbb{R}^2$ *is a closed curve and that* $\{\gamma_n\}_{n=1}^{\infty}$ *converges uniformly on* $[a,b]$ *to* $\gamma\colon [a,b] \to \mathbb{R}^2$. *Then* γ *is a closed curve. Moreover, if* $p \in \mathbb{R}^2 \setminus [\gamma]$, *then, for all sufficiently large* n, $p \notin [\gamma_n]$ *and* $W(\gamma; p) = W(\gamma_n; p)$.

Proof. Since each γ_n is continuous and the convergence is uniform, γ is continuous, i.e., γ is a curve. Since $\gamma(a) = \lim_{n\to\infty} \gamma_n(a) = \lim_{n\to\infty} \gamma_n(b) = \gamma(b)$, γ is closed. Let $0 < \Delta \leqslant |\gamma(t) - p|$ for $t \in [a,b]$ and choose a natural number N such that

$$|\gamma_n(t) - \gamma(t)| < \Delta \quad \text{for all } t \in [a,b] \quad \text{when } n > N.$$

Then $|\gamma_n(t) - \gamma(t)| < \Delta \leqslant |\gamma(t) - p| \leqslant |\gamma_n(t) - p| + |\gamma(t) - p|$ for all $t \in [a,b]$ when $n > N$. Hence, by Theorem 1, $W(\gamma_n; p) = W(\gamma; p)$ for $n > N$.

Theorem 1 and its corollaries suggest that $W(\gamma; p)$ depends, in some sense, continuously on both γ and p. In fact, Corollary 1 implies that, for fixed γ, the function $p \to W(\gamma; p)$ is constant on each component of $\mathbb{R}^2 \setminus [\gamma]$. The next corollary implies that $W(\gamma; p) = 0$ for p belonging to the unbounded component of $\mathbb{R}^2 \setminus [\gamma]$.

COROLLARY 3. *Suppose* $\gamma\colon [a,b] \to \mathbb{R}^2$ *is a closed curve,* $p \in \mathbb{R}^2$ *and* $|\gamma(t)| < |p|$ *for all* $t \in [a,b]$. *Then* $W(\gamma; p) = 0$.

Proof. Let $\eta(t) = -p$ for $t \in [a,b]$. Then $|(\gamma(t) - p) - \eta(t)| < |\eta(t)| \leqslant |\gamma(t) - p| + |\eta(t)|$ for all $t \in [a,b]$. Hence, by Theorem 1, $W(\gamma - p; 0) = W(\eta; 0)$. But $W(\eta; 0) = 0$ by property $W.1$. and $W(\gamma - p; 0) = W(\gamma; p)$ by property $W.6$.

We now aim to prove that $W(\gamma; p)$ depends continuously on γ (in a sense now to be defined). Suppose U is a non-empty open subset of $\mathbb{R}^2$ and $\gamma, \eta\colon [a,b] \to \mathbb{R}^2$ are closed curves in U (i.e., $[\gamma], [\eta] \subseteq U$). We say γ is *homotopic to* η *in* U provided there exists a function $H\colon [a,b] \times [0,1] \to \mathbb{R}^2$ such that

i) H is continuous on $[a,b] \times [0,1]$,
ii) $H(t,\lambda) \in U$ for every $t \in [a,b]$ and every $\lambda \in [0,1]$,
iii) $H(a,\lambda) = H(b,\lambda)$ for every $\lambda \in [0,1]$, and
iv) $H(t,0) = \gamma(t)$ and $H(t,1) = \eta(t)$ for every $t \in [a,b]$.

Such an H is called a *homotopy* in U. With H as in the definition of homotopy, for $0 \leqslant \lambda \leqslant 1$ let $H_\lambda(t) = H(t,\lambda)$ for $a \leqslant t \leqslant b$. Then it follows that H_λ is a closed curve in U for each $\lambda \in [0,1]$, $H_0 = \gamma$ and $H_1 = \eta$. Intuitively, as λ varies from 0 to 1, H_λ varies continuously from γ to η with $[H_\lambda] \subseteq U$ for every $\lambda \in [0,1]$. Thus H can be thought of as a continuous deformation of γ to η *inside* U. Property i) implies that if $\lambda_k \to \lambda_0$ in $[0,1]$ then $H_{\lambda_k} \to H_{\lambda_0}$ uniformly on $[a,b]$. This can be proved by noting that i) implies that H is uniformly continuous. Note that if $\gamma, \eta\colon [a,b] \to \mathbb{R}^2$ are closed curves then γ is homotopic to η in $\mathbb{R}^2$; simply let $H(t,\lambda) = (1-\lambda)\gamma(t) + \lambda\eta(t)$ for $a \leqslant t \leqslant b$ and $0 \leqslant \lambda \leqslant 1$.

THEOREM 2. *Suppose* $p \in \mathbb{R}^2$ *and* γ, η: $[a,b] \to \mathbb{R}^2$ *are closed curves that are homotopic in* $\mathbb{R}^2 \setminus \{p\}$. *Then* $W(\gamma; p) = W(\eta; p)$.

Proof. Choose a suitable homotopy H and let $\psi(\lambda) = W(H_\lambda; p)$ for $0 \leqslant \lambda \leqslant 1$. If $\lambda_k \to \lambda_0$ in $[0,1]$ then $H_{\lambda_k} \to H_{\lambda_0}$ uniformly on $[a,b]$ so that, by Corollary 2, $\psi(\lambda_k) = \psi(\lambda_0)$ for all sufficiently large k. It follows that ψ is continuous on $[0,1]$. But ψ is integer valued and therefore constant, according to the intermediate value theorem. Thus $W(\gamma; p) = W(H_0; p) = W(H_1; p) = W(\eta; p)$.

4. Degree Theory in Two Dimensions

Suppose $D = \{(x,y) \in \mathbb{R}^2 : (x - x_0)^2 + (y - y_0)^2 < \rho^2\}$ denotes the disk with center $z_0 = (x_0, y_0)$ and radius $\rho > 0$. Let ∂D: $[0, 2\pi] \to \mathbb{R}^2$ be defined by

$$\partial D(t) = (x_0 + \rho \cos t, y_0 + \rho \sin t) \quad \text{for } 0 \leqslant t \leqslant 2\pi$$

so that ∂D is a closed curve and $[\partial D] = \{(x,y) \in \mathbb{R}^2 : (x - x_0)^2 + (y - y_0)^2 = \rho^2\} = bdD$, the boundary of D. Let $\overline{D} = D \cup bdD$, the closure of D. Suppose too that F: $\overline{D} \to \mathbb{R}^2$ is continuous, $p \in \mathbb{R}^2$ and $F(\partial D(t)) \neq p$ for all $t \in [0, 2\pi]$; that is, $F(z) \neq p$ for all $z \in bdD$. Thus $F \circ \partial D$, the composition of F with ∂D, is a closed curve and $p \notin [F \circ \partial D]$. We define $d(F; D; p)$, the *degree of* F *relative to* D *at* p by

$$d(F; D; p) = W(F \circ \partial D; p).$$

If F is analytic in an open set containing $\overline{D}$ then

$$d(F; D; p) = \frac{1}{2\pi i} \int_{F \circ \partial D} \frac{1}{\zeta - p} d\zeta = \frac{1}{2\pi i} \int_0^{2\pi} \frac{F'(z_0 + \rho e^{it}) i \rho e^{it}}{F(z_0 + \rho e^{it}) - p} dt$$
$$= \frac{1}{2\pi i} \int_{\partial D} \frac{F'(z)}{F(z) - p} dz$$

and hence, according to the "argument principle" of complex analysis, $d(F; D; p)$ is a nonnegative integer and is, in fact, the number of solutions, z, in D of the equation $F(z) = p$, counted according to multiplicities (see [4, p. 179]). In particular, if $d(F; D; p) \neq 0$, there is at least one z in D such that $F(z) = p$.

In the general case (with F not necessarily analytic) $d(F; D; p)$, may be negative. Intuitively, if $d(F; D; p) \neq 0$ then p is "enclosed" by the image under F of bdD; we will see that, in this case too, the equation $F(z) = p$ necessarily has at least one solution, z, in D. This result can thus be thought of as a generalization of the intermediate value theorem to two dimensions as well as a partial generalization of the argument principle.

We now record some properties of degree. In these assertions F, D and p are as above.

D.1. If $F(z) = z$ for all $z \in bdD$ and $|p - z_0| < \rho$ then $d(F; D; p) = 1$.

D.2. If we let $(F - p)(z) = F(z) - p$ for $z \in \overline{D}$ then $d(F; D; p) = d(F - p; D; 0)$.

Property *D*.1 follows easily from property *W*.2 and Corollary 1. Property *D*.2 follows from *W*.6. It is also easy to deduce the following partial generalization of Rouché's Theorem from Theorem 1 (see [4, pp. 179 and 190]).

THEOREM 3. *Suppose D is a disk in $\mathbb{R}^2$, $F,G\colon \overline{D} \to \mathbb{R}^2$ are continuous, $p \in \mathbb{R}^2$ and*

$$|F(z) - G(z)| < |F(z) - p| + |G(z) - p| \quad \textit{for } z \in bdD.$$

Then $F(z) \neq p$ and $G(z) \neq p$ for all $z \in bdD$ and $d(F;\, D;\, p) = d(G;\, D;\, p)$.

From Theorem 2 we can easily prove the following result.

THEOREM 4. *Suppose D is a disk in $\mathbb{R}^2, p \in \mathbb{R}^2$, $F\colon \overline{D} \times [0,1] \to \mathbb{R}^2$ is continuous and such that*

$$p \neq F(z,\lambda) \ \textit{for } z \in bdD \textit{ and } 0 \leqslant \lambda \leqslant 1.$$

Then $d(F_\lambda, D, p) = d(F_\mu, D, p)$ for $\lambda, \mu \in [0,1]$ where, for $0 \leqslant \lambda \leqslant 1$, $F_\lambda(z) = F(z,\lambda)$ for $z \in \overline{D}$.

The following result is one of the most important properties of degree. As mentioned earlier it can be thought of as a generalization of both the intermediate value theorem and the argument principle.

THEOREM 5. *Suppose D is a disk in $\mathbb{R}^2, p \in \mathbb{R}^2$, $F\colon \overline{D} \to \mathbb{R}^2$ is continuous, $p \neq F(z)$ for all $z \in bdD$ and $d(F;\, D;\, p) \neq 0$. Then there exists a z in D such that $F(z) = p$. In fact there exists a $\Delta > 0$ such that, if $q \in \mathbb{R}^2$ such that $|p - q| < \Delta$, there exists $w \in D$ such that $F(w) = q$.*

Proof. Suppose, on the contrary, that $F(z) \neq p$ for all $z \in \overline{D}$. Suppose D has center z_0 and radius $\rho > 0$. For $0 \leqslant t \leqslant 2\pi$ and $0 \leqslant \lambda \leqslant 1$ let $H(t,\lambda) = F(z_0 + \lambda\rho e^{it})$.

If we let $\gamma(t) = H(t,0) = F(z_0)$ and $\eta(t) = H(t,1) = F(z_0 + \rho e^{it})$ for $0 \leqslant t \leqslant \pi$, then it follows that γ is homotopic to η in $\mathbb{R}^2 \setminus \{p\}$. Hence, according to Theorem 2, $W(\gamma;\, p) = W(\eta;\, p)$. But $W(\gamma;\, p) = 0$ by property $W.1$ and $W(\eta;\, p) = d(F;\, D;\, p)$ by the definition of degree. Thus $d(F;\, D;\, p) = 0$, contradicting our assumption according to Theorem 4. The last assertion follows from Corollary 1 of Theorem 1.

One might wonder if, under the hypotheses of Theorem 5, the equation $F(z) = p$ has exactly $|d(F;\, D;\, p)|$ solutions. The following examples show that this is not necessarily so.

Examples. Let $D = \{(x,y) \in \mathbb{R}^2 \colon x^2 + y^2 < 1\}$ and define $F, G, H\colon \mathbb{R}^2 \to \mathbb{R}^2$ as follows. For $(x,y) \in \mathbb{R}^2$, $F(x,y) = (x, |y|)$, $G(x,y) = (x^2 - y^2, -2xy)$ and $H(x,y) = \phi(x^2 + y^2)(x,y)$ where

$$\phi(t) = \begin{cases} 0 & \text{for } 0 \leqslant t \leqslant 1/2 \\ 2t - 1 & \text{for } 1/2 < t \leqslant 1. \end{cases}$$

Notice that $G(z) = \bar{z}^2$ in complex notation. Then $bdD = \{(x,y) \in \mathbb{R}^2 \colon x^2 + y^2 = 1\}$, $F(bdD) = \{(x,y) \in \mathbb{R}^2 \colon x^2 + y^2 = 1,\ y \geqslant 0\}$, $G(bdD) = bdD$ and $H(bdD) = bdD$.

It is not difficult to check that

i) $d(F\colon D;\, p) = 0$ for all $p \in \mathbb{R}^2 \setminus F(bdD)$, but the equation $F(x,y) = (0,0)$ has exactly one solution in D, the equation $F(x,y) = (0,1/2)$ has exactly two solutions in D, and the equation $F(x,y) = (0,2)$ has no solutions in D.

ii) $d(G;\, D;\, p) = -2$ for all $p \in D$, the equation $G(x,y) = (0,0)$ has exactly one solution in D but the equation $G(x,y) = (1/4, 0)$ has exactly two solutions in D.

iii) $d(H;\, D;\, p) = 1$ for all $p \in D$, the equation $H(x,y) = (0,0)$ has infinitely many solutions in D whereas the equation $H(x,y) = (3/4, 0)$ has exactly one solution in D.

The next two sections concern some well-known applications of degree theory (see **[5]** and **[11]**).

5. The Fundamental Theorem of Algebra

Suppose n is a natural number, $c_0, c_1, \dots, c_{n-1} \in \mathbb{C}$ and

$$F(z) = z^n + c_{n-1}z^{n-1} + \cdots + c_1 z + c_0 \quad \textit{for } z \in \mathbb{C}.$$

Then there exists $z_0 \in \mathbb{C}$ such that $F(z_0) = 0$.

Proof. Choose $r > 0$ so large that

$$\left|\frac{c_{n-1}}{z} + \cdots + \frac{c_1}{z^{n-1}} + \frac{c_0}{z^n}\right| < 1 \quad \text{when } z \in \mathbb{C} \text{ and } |z| \geqslant r.$$

Let $G(z) = z^n$ for $z \in \mathbb{C}$ and let $D = \{z \in \mathbb{C}: |z| < r\}$. From property $W.2$ and the definition of degree, it follows that $d(G, D, 0) = n$.

If $z \in bdD$ then $|z| = r$ and

$$|F(z) - G(z)| = |z^n|\left|\frac{c_{n-1}}{z} + \cdots + \frac{c_0}{z^n}\right| < |z^n| = |G(z)| \leqslant |F(z)| + |G(z)|.$$

Thus, by Theorem 3, $d(F; D; 0) = d(G; D; 0)$. Hence $d(F; D; 0) = n \neq 0$ and therefore, by Theorem 5, there exists $z_0 \in D$ such that $F(z_0) = 0$.

6. Brouwer's Fixed Point Theorem in the Plane

In this section $D = \{z \in \mathbb{C}: |z| < 1\}$. The results of this section and the main ideas of their proofs are well known (e.g., [**5**, pp. 106–110]).

PROPOSITION. *Suppose G: $\overline{D} \to \mathbb{R}^2$, G is continuous and $G(z) = z$ for all $z \in bdD$. Then $G(D) \supseteq D$. That is, for every $p \in D$ there exits $z \in D$ such that $G(z) = p$.*

Proof. If $p \in D$ then $d(G, D, p) = 1$ by the definition of degree and property $W.2$. The assertion now follows from Theorem 5 and Corollary 1.

COROLLARY. *There does not exist a continuous function G: $\overline{D} \to bdD$ such that $G(z) = z$ for all $z \in bdD$.*

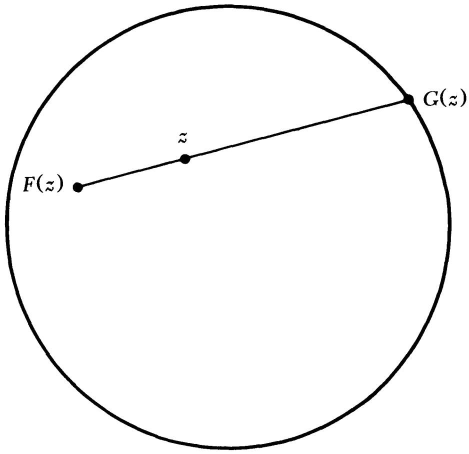
This corollary follows immediately from the proposition and will be used to prove the following two-dimensional version of the famous

BROUWER'S FIXED POINT THEOREM. *Suppose F: $\overline{D} \to \overline{D}$ is continuous. Then F has a fixed point, that is, there exists a z in $\overline{D}$ such that $F(z) = z$.*

Proof. Suppose, on the contrary, that $F(z) \neq z$ for all $z \in \overline{D}$.

We will define a continuous function G: $\overline{D} \to bdD$ such that $G(z) = z$ for all $z \in bdD$. This will contradict the last corollary and thereby prove the result.

The G we have in mind can be described geometrically as follows. For $z \in \overline{D}$, $F(z) \neq z$ and hence there is a unique ray from $F(z)$ through z. It cuts bdD in a unique point; call it $G(z)$. Clearly, $G(z) = z$ for $z \in bdD$ and, intuitively, G is continuous.



To prove analytically the existence of such a G it suffices to show that there is a continuous function $\lambda\colon \overline{D} \to [1,\infty)$ such that

$$|F(z)+\lambda(z)(z-F(z))| = 1 \quad \text{for all } z \in \overline{D} \tag{$\dagger$}$$

and to define $G(z) = F(z) + \lambda(z)(z - F(z))$ for $z \in \overline{D}$.

Fix z temporarily in $\overline{D}$ and let $q(\lambda) = |F(z) + \lambda(z - F(z))|^2 - 1 = |F(z)|^2 + 2\lambda F(z)\cdot(z - F(z)) + \lambda^2|z - F(z)|^2 - 1$ for $\lambda \in \mathbb{R}$. Then $q(\lambda)$ is a polynomial of degree two in λ whose leading coefficient is positive and hence $\lim_{\lambda\to+\infty} q(\lambda) = +\infty$. But $q(1) = |z|^2 - 1 \leqslant 0$. Hence, according to the intermediate value theorem, q has a unique root in $[1,\infty)$; call it $\lambda(z)$. The quadratic formula implies that λ is continuous on $\overline{D}$.

It is not difficult to generalize Brouwer's Theorem to certain sets other than $\overline{D}$. Suppose that a subset K of $\mathbb{R}^n$ (or any topological space) is *homeomorphic* to $\overline{D}$, that is, suppose there exists a one-to-one, continuous function ϕ from $\overline{D}$ onto K such that ϕ^{-1} is also continuous. Suppose $G\colon K \to K$ is continuous. Let $F(z) = \phi^{-1}(G(\phi(z)))$ for $z \in \overline{D}$. Then, by Brouwer's Theorem, there exists a z in D such that $F(z) = z$. If $w = \phi(z)$ it follows that $G(w) = w$. We have thus proved that any subset, K, of $\mathbb{R}^n$ that is homeomorphic to $\overline{D}$ has the *fixed point property*; if $G\colon K \to K$ is continuous then there exists a $w \in K$ such that $G(w) = w$.

7. Remarks on Degree Theory

Before mentioning extensions of degree theory to higher dimensions, we will illustrate how the theory can be extended so as to apply, in two dimensions, to certain sets other than disks.

Suppose Δ is a nondegenerate triangle in $\mathbb{R}^2$ with vertices a, b and c. That is, $\Delta = \{sa + tb + uc\colon s,t,u \geqslant 0 \text{ and } s+t+u = 1\}$ where a, b and c are not collinear. Define $\partial\Delta\colon [0,3] \to \mathbb{R}^2$ by

$$(\partial\Delta)(t) = \begin{cases} a + t(b-a) & \text{if } 0 \leqslant t \leqslant 1 \\ b + (t-1)(c-b) & \text{if } 1 < t \leqslant 2 \\ c + (t-2)(a-c) & \text{if } 2 < t \leqslant 3 \end{cases}$$

so that $\partial\Delta$ is a piecewise C^1, closed curve and $[\partial\Delta] = bd\Delta$. If $F\colon \Delta \to \mathbb{R}^2$ is continuous, $p \in \mathbb{R}^2$ and $F(z) \neq p$ for all $z \in bd\Delta$, define

$$d(F;\Delta;p) = W(F \circ \partial\Delta;p).$$

With these definitions one can prove the analogues of Theorems 3, 4, 5 and Brouwer's Theorem in the same way in which we have proved them for closed disks.

As an illustration we will present a proof, in the case $n = 3$, of the following theorem of Perron ([7, p. 64]): if A is a real $n \times n$ matrix with positive entries then A has a positive eigenvalue and an associated eigenvector with nonnegative components.

Suppose $A = (a_{ij})$ is a real, 3×3 matrix with $a_{ij} > 0$ for $1 \leqslant i, j \leqslant 3$. Let $K = \{(x, y, z) \in \mathbb{R}^3\colon x, y, z \geqslant 0 \text{ and } x + y + z = 1\}$. For $v = (x, y, z) \in \mathbb{R}^3$ let $l(v) = x + y + z$. For $X \in K$ define $F(X) = l((AX^T)^T)^{-1}(AX^T)^T$ where X^T is the transpose of X. It follows that F maps K into K and is continuous. The mapping $(s, t) \to (s, t, 1 - s - t)$ is a homeomorphism of the triangle $\{(s, t)\colon s, t \geqslant 0, s + t \leqslant 1\}$ onto K. Hence, by Brouwer's Theorem, F has a fixed point, say X_0. Thus

$$\lambda X_0^T = AX_0^T, \text{ where } \lambda = l\Big((AX^T)^T\Big) > 0,$$

from which the desired conclusion easily follows.

Other applications of winding numbers can be found in the very readable book of Chinn and Steenrod [5]. We also recommend the elegant and elementary article of M. K. Fort [6], which contains similar ideas.

The theory of degree can be extended to higher dimensions in the following sense. Given a bounded, open subset, D, of $\mathbb{R}^n$, given a continuous function $F\colon \overline{D} \to \mathbb{R}^n$, and given $p \in \mathbb{R}^n$ such that $F(x) \neq p$ for all $x \in bdD$, it is possible to define an integer, $d(F; D; p)$, with properties analogous to those enunciated in Theorems 3, 4 and 5. This allows one to extend Brouwer's Theorem to $\mathbb{R}^n$ and, for example, to prove the theorem of Perron in full generality. The main difficulty in this development is to *define* $d(F; D; p)$. This was originally accomplished by Brouwer [3] near the turn of the century using combinatorial ideas. In the early 1950s, Nagumo [13] found an analytic approach based on the differential calculus of functions of several real variables. A theory of degree based on integral calculus was developed by E. Heinz [8] in the late '50s. The book of Lloyd [11] contains a readable account of degree theory in $\mathbb{R}^n$ based on the works of Nagumo and Heinz together with generalizations to infinite-dimensional spaces.

Brouwer's Theorem can be proved without using degree theory. See, for example, the elegant papers of Milnor [12] and Rogers [15]. Brouwer's Theorem was extended to infinite dimensions by Schauder [18]. His generalization, Schauder's Fixed Point Theorem, can be used to deduce existence theorems for integral and differential equations.

The concept of degree was extended to infinite dimensions by Leray and Schauder [10]. The reader may find details in Lloyd's book, which also contains references to many other works on degree theory and its applications.

8. The Inverse Function Theorem

Suppose U is an open subset of $\mathbb{R}^2$, $F\colon U \to \mathbb{R}^2$ is differentiable and $\det J_F(x, y) \neq 0$ for all $(x, y) \in U$ where $J_F(x, y)$ is the Jacobian matrix (of the Fréchet derivative or total derivative) of F at (x, y).

The inverse function theorem (in two dimensions) states that if F is *continuously* differentiable on U, then every $(a,b) \in U$ has a neighbourhood, $V \subseteq U$, such that $F(V) = W$ is open and the restriction of F to V is one-to-one and has a continuously differentiable inverse. This may be proved by establishing the following three facts:

i) F is an open mapping,
ii) F is locally one-to-one, and
iii) the local inverses are continuously differentiable.

Steps i) and ii) are usually proved together by using the contraction mapping principle (Banach's fixed point theorem) as in Rudin [16]. Spivak [19] employs a simpler argument. Step iii) is relatively easy.

The inverse function theorem in one dimension follows from Darboux's Theorem (the intermediate value property of derivatives) without assuming the continuity of the derivative. Thus one might ask with M. Laczkovich (personal communication) if the inverse function theorem holds in higher dimensions *without the assumption of continuity of the derivative* (Jacobian matrix). As far as I know, the answer to this question is not known. However, we will use degree theory to prove i), without assuming the continuity of the derivative.

AN OPEN MAPPING THEOREM. *Let U and F be as in the first paragraph of this section. If Ω is an open subset of U then $F(\Omega)$ is open.*

Proof. Let $(a,b) \in U$ and let L denote the derivative of F at (a,b). Thus L is a one-to-one linear mapping of $\mathbb{R}^2$ onto $\mathbb{R}^2$ since its matrix, relative to the usual basis for $\mathbb{R}^2$, is $J_F(a,b)$ and $\det J_F(a,b) \neq 0$. Moreover, for each $\varepsilon > 0$ there exists a $\delta > 0$ such that $(a+h, b+k) \in U$ and

$$|F(a+h,b+k) - F(a,b) - L(h,k)| < \varepsilon|(h,k)| \tag{\#}$$

whenever $(h,k) \in \mathbb{R}^2$ and $|(h,k)|^2 = h^2 + k^2 \leqslant \delta^2$. Thus

$$|L\{L^{-1}[F(a+h,b+k) - F(a,b)] - (h,k)\}| < \varepsilon|(h,k)|$$

if $(h,k) \in \mathbb{R}^2$ and $|(h,k)| \leqslant \delta$.

Since L is linear, one-to-one and continuous, the mapping $(u,v) \to |L(u,v)|$ has a positive minimum on the compact set $\{(u,v) \in \mathbb{R}^2 | u^2 + v^2 = 1\}$. If we denote this minimum value by m then $m > 0$ and, by homogeneity, we find that

$$m|(s,t)| \leqslant |L(s,t)| \quad \text{for all } (s,t) \in \mathbb{R}^2.$$

Now choose $\varepsilon > 0$ so that $\varepsilon < m$ and select $\delta > 0$ so that (#) holds. Then

$$m|L^{-1}[F(a+h,b+k) - F(a,b)] - (h,k)| < \varepsilon|(h,k)|$$

whenever $(h,k) \in \mathbb{R}^2$ and $h^2 + k^2 \leqslant \delta^2$. Let

$$G(h,k) = L^{-1}[F(a+h,b+k) - F(a,b)]$$

for $(h,k) \in \overline{D} = \{(h,k) \in \mathbb{R}^2\colon h^2 + k^2 \leqslant \delta^2\}$. Then G is continuous on $\overline{D}$ and, for all $(h,k) \in \overline{D}$,

$$|G(h,k) - I(h,k)| < \frac{\varepsilon}{m}|(h,k)| < |I(h,k)|$$

where I denotes the identity mapping on $\overline{D}$. According to Theorem 3, $d(G; D; (0,0)) = d(I; D; (0,0))$ and $d(I; D; (0,0)) = 1$ by property $D.1$. Moreover, according to Theorem 5, there exists a $\Delta > 0$ such that, for each $(u,v) \in \mathbb{R}^2$ satisfying $u^2 + v^2 < \Delta^2$, there exists $(h,k) \in D$ such that $G(h,k) = (u,v)$.

We have shown that $G(D) \supseteq \{(u,v) \in \mathbb{R}^2 | u^2 + v^2 < \Delta^2\}$. But $F(a+h, b+k) = F(a,b) + L(G(h,k))$ for $(h,k) \in D$ and L maps open sets to open sets. It follows that the image under F of $\{(x_0 + h, y_0 + k) | h^2 + k^2 < \Delta^2\}$ contains a neighbourhood of $F(a,b)$. Hence, since (a,b) was chosen arbitrarily from U, $F(U)$ contains a neighbourhood of each of its members and therefore $F(U)$ is open. The same argument applies to the restriction of F to any nonempty open subset Ω of U.

Remark. We have actually proved the following: If U is an open subset of $\mathbb{R}^2$, $F\colon U \to \mathbb{R}^2$ is continuous, F is differentiable at some $(a,b) \in U$ and $\det J_F(a,b) \neq 0$ then $F(a,b)$ is an interior point of $F(U)$.

9. Differential Equations

The aim of this final section is to illustrate briefly the utility of the polar representation in the study of differential equations. A more thorough discussion can be found in [9] and [14]. According to [9] the ideas discussed here are due to H. Prüfer.

Suppose I is a nondegenerate interval in $\mathbb{R}$ and $a, b, c, d\colon I \to \mathbb{R}$ are continuous functions. We are interested in the linear system of differential equations

$$\begin{aligned} x'(t) &= a(t)x(t) + b(t)y(t) \\ y'(t) &= c(t)x(t) + d(t)y(t) \quad \text{for } t \in I. \end{aligned} \tag{S}$$

Suppose $x, y\colon I \to \mathbb{R}$ are such that (S) holds. Then x and y are C^1 on I. If we write members of $\mathbb{R}^2$ as column vectors and let

$$\gamma(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} \quad \text{and} \quad A(t) = \begin{pmatrix} a(t) & b(t) \\ c(t) & d(t) \end{pmatrix} \quad \text{for } t \in I$$

then (S) becomes

$$\gamma'(t) = A(t)\gamma(t) \quad \text{for } t \in I. \tag{S$'$}$$

To begin with we will show that if $(S)'$ holds for some $\gamma\colon I \to \mathbb{R}^2$ and $\gamma(t_0) = \binom{0}{0}$ for some $t_0 \in I$ then $\gamma(t) = \binom{0}{0}$ for all $t \in I$. It follows that a solution, γ, of $(S)'$ either vanishes everywhere or nowhere. Thus a nontrivial solution of (S) has a polar representation.

LEMMA. *Suppose $\gamma\colon I \to \mathbb{R}^2$ satisfies $(S)'$ and $\gamma(t_0) = \binom{0}{0}$ for some $t_0 \in I$. Then $\gamma(t) = \binom{0}{0}$ for all $t \in I$.*

Proof. Let $\phi(t) = |\gamma(t)|^2 = \gamma(t) \cdot \gamma(t)$ for $t \in I$. Then ϕ is C^1 on I, and

$$\phi'(t) = 2\gamma(t) \cdot \gamma'(t) \quad \text{for every } t \in I.$$

Suppose there exists $t_1 \in I$ such that $t_1 > t_0$ and let $\lambda = \max\{|U \cdot A(t)U| \colon t_0 \leqslant t \leqslant t_1,\ U \in \mathbb{R}^2,\ |U| = 1\}$. Then $|X \cdot A(t)X| \leqslant \lambda |X|^2$ for all $X \in \mathbb{R}^2$ so that

$$|\phi'(t)| = |2\gamma(t) \cdot A(t)\gamma(t)| \leqslant 2\lambda |\gamma(t)|^2 = 2\lambda\phi(t)$$

for all $t \in [t_0, t_1]$. Thus, for $t_0 \leqslant t \leqslant t_1$,

$$0 \leqslant \phi(t) = |\phi(t) - \phi(t_0)| = \left| \int_{t_0}^{t} \phi'(s)\, ds \right| \leqslant \int_{t_0}^{t} 2\lambda\phi(s)\, ds.$$

It follows from Gronwall's lemma [14, p. 17] that $\phi(t) = 0$ for all $t \in [t_0, t_1]$. Thus $\phi(t) = 0$ for all $t \in I$ such that $t \geqslant t_0$. Similarly $\phi(t) = 0$ for all $t \in I$ such that $t \leqslant t_0$. Hence $\phi(t) = 0$ for all $t \in I$.

Now suppose x, y: $I \to \mathbb{R}$ satisfy (S) and, for some $t_0 \in I$, either $x(t_0) \neq 0$ or $y(t_0) \neq 0$. Then $(x(t), y(t)) \neq (0,0)$ for all $t \in I$ and hence, according to Proposition 1, there exist C^1 functions r, θ: $I \to \mathbb{R}$ such that $r(t) > 0$ and

$$x(t) = r(t)\cos\theta(t) \quad \text{and} \quad y(t) = r(t)\sin\theta(t) \quad \text{for all } t \in I. \tag{P}$$

Substituting (P) into (S) we obtain a pair of equations that can be regarded as a linear system in $r'(t)$ and $\theta'(t)$. Moreover the coefficient matrix is invertible and we deduce that, for all $t \in I$,

$$\begin{aligned} \frac{r'(t)}{r(t)} &= a(t)\cos^2\theta(t) + [b(t) + c(t)]\cos\theta(t)\sin\theta(t) + d(t)\sin^2\theta(t) \\ \theta'(t) &= c(t)\cos^2\theta(t) + [d(t) - a(t)]\cos\theta(t)\sin\theta(t) - b(t)\sin^2\theta(t). \end{aligned} \tag{SP}$$

Conversely, if r, θ: $I \to \mathbb{R}$ satisfy (SP) with $r(t) > 0$ for all $t \in I$ and if we define x and y by (P) then (S) holds. Thus (S) is equivalent to (SP).

Observe that (SP) is "uncoupled" so that existence and uniqueness questions for the system (S) are reduced to existence and uniqueness questions for the single equation involving $\theta'(t)$ in (SP). The dependence of the right-hand side of this equation in $\theta(t)$ is such that Picard's theorem guarantees a unique global solution satisfying a given initial condition (see [14, p. 240]).

Let us now suppose that $I = [0, \infty)$ and let

$$B(t) = \frac{1}{2}\left[A(t) + A(t)^T\right] \text{ and } C(t) = \begin{pmatrix} c(t) & \dfrac{d(t) - a(t)}{2} \\ \dfrac{d(t) - a(t)}{2} & -b(t) \end{pmatrix} \quad \text{for } t \geqslant 0.$$

Then (SP) can be written

$$\begin{aligned} \frac{r'(t)}{r(t)} &= (\cos\theta(t), \sin\theta(t))B(t)(\cos\theta(t), \sin\theta(t))^T \\ \theta'(t) &= (\cos\theta(t), \sin\theta(t))C(t)(\cos\theta(t), \sin\theta(t))^T. \end{aligned} \tag{SP}'$$

Suppose C is "uniformly positive definite," that is, there exists $\lambda > 0$ such that

$$XC(t)X^T \geqslant \lambda|X|^2 \quad \text{for all } X \in \mathbb{R}^2 \quad \text{and all } t \geqslant 0.$$

Then $\theta'(t) \geqslant \lambda$ for all $t \in I$ so that θ is strictly increasing and $\lim_{t\to\infty}\theta(t) = \infty$. Thus x and y "oscillate" and the zeros of x and y are "intertwined." A similar assertion holds if C is uniformly negative definite.

If B is uniformly negative definite, it follows that, for some $\mu > 0$,

$$\frac{d}{dt}\{\ln r(t)\} = \frac{r'(t)}{r(t)} \leqslant -\mu \quad \text{for all } t \geqslant 0$$

and hence r is decreasing and $\lim_{t \to \infty} r(t) = 0$. In this case we find that $\lim_{t \to \infty} x(t) = \lim_{t \to \infty} y(t) = 0$.

REFERENCES

1. Lars V. Ahlfors, *Complex Analysis*, 3rd edition, McGraw-Hill, New York, 1979.
2. A. F. Beardon, *Complex Analysis*, John Wiley and Sons, Inc., New York, 1979.
3. L. E. J. Brouwer, Über Abbildung von Mannigfaltigkeiten, *Math. Ann.* 71 (1912), 97–115.
4. Robert B. Burckel, *An Introduction to Classical Complex Analysis*, Vol. 1, Academic Press, New York, 1979.
5. W. G. Chinn and N. E. Steenrod, *First Concepts of Topology*, New Mathematical Library Series of The MAA, 1966.
6. M. K. Fort, Jr., Continuous square roots of mappings, from *Lectures on Calculus*, edited by K. O. May, Holden-Day, Oakland, CA, 1967.
7. F. R. Gantmacher, *Applications of the Theory of Matrices*, Interscience Publishers, Inc., New York, 1959.
8. E. Heinz, An elementary analytic theory of the degree of a mapping in n-dimensional space, *J. Math. and Mech.* 8 (1959), 213–47.
9. Einar Hille, *Lectures on Ordinary Differential Equations*, Addison-Wesley, Reading, MA, 1969.
10. J. Leray and J. Schauder, Topologie et équations fonctionnelles, *Ann. Sci. École. Norm. Sup. Sér. 3* 51(1934), 45–78.
11. N. G. Lloyd, *Degree Theory*, Cambridge Tracts in Mathematics, Cambridge University Press, New York, 1978.
12. John Milnor, Analytic proofs of the "hairy ball theorem" and the Brouwer Fixed Point Theorem, *Amer. Math. Monthly* 85(1978), 521–524.
13. M. Nagumo, A theory of degree of mappings based on infinitesimal analysis, *Amer. J. Math.* 73 (1951), 485–496.
14. L. C. Piccinini, G. Stampacchia, and G. Vidossich, *Ordinary Differential Equations in* $\mathbb{R}^n$, #39, Applied Mathematical Sciences Series, Springer-Verlag, New York, 1984.
15. C. A. Rogers, A less strange version of Milnor's proof of Brouwer's Fixed Point Theorem, *Amer. Math. Monthly*, 87(1980), 525–527.
16. Walter Rudin, *Principles of Mathematical Analysis*, 3rd edition, McGraw-Hill, New York, 1976.
17. ______, *Real and Complex Analysis*, 3rd edition, McGraw-Hill, New York, 1987.
18. J. Schauder, Der Fixpunktsatz in Funktionalräumen, *Studia Math.* 2(1930), 171–180.
19. Michael Spivak, *Calculus on Manifolds*, Benjamin, Princeton, NJ, 1965.

Limerick

A teacher once, having some fun,
In presenting that two equals one,
Remained quite aloof
From his rigorous proof;
But his class was convinced and undone.

—ARTHUR WHITE
WESTERN MICHIGAN UNIVERSITY
KALAMAZOO, MI 49008

and hence r is decreasing and $\lim_{t \to \infty} r(t) = 0$. In this case we find that $\lim_{t \to \infty} x(t) = \lim_{t \to \infty} y(t) = 0$.

REFERENCES

1. Lars V. Ahlfors, *Complex Analysis*, 3rd edition, McGraw-Hill, New York, 1979.
2. A. F. Beardon, *Complex Analysis*, John Wiley and Sons, Inc., New York, 1979.
3. L. E. J. Brouwer, Über Abbildung von Mannigfaltigkeiten, *Math. Ann.* 71 (1912), 97–115.
4. Robert B. Burckel, *An Introduction to Classical Complex Analysis*, Vol. 1, Academic Press, New York, 1979.
5. W. G. Chinn and N. E. Steenrod, *First Concepts of Topology*, New Mathematical Library Series of The MAA, 1966.
6. M. K. Fort, Jr., Continuous square roots of mappings, from *Lectures on Calculus*, edited by K. O. May, Holden-Day, Oakland, CA, 1967.
7. F. R. Gantmacher, *Applications of the Theory of Matrices*, Interscience Publishers, Inc., New York, 1959.
8. E. Heinz, An elementary analytic theory of the degree of a mapping in n-dimensional space, *J. Math. and Mech.* 8 (1959), 213–47.
9. Einar Hille, *Lectures on Ordinary Differential Equations*, Addison-Wesley, Reading, MA, 1969.
10. J. Leray and J. Schauder, Topologie et équations fonctionnelles, *Ann. Sci. École. Norm. Sup. Sér. 3* 51(1934), 45–78.
11. N. G. Lloyd, *Degree Theory*, Cambridge Tracts in Mathematics, Cambridge University Press, New York, 1978.
12. John Milnor, Analytic proofs of the "hairy ball theorem" and the Brouwer Fixed Point Theorem, *Amer. Math. Monthly* 85(1978), 521–524.
13. M. Nagumo, A theory of degree of mappings based on infinitesimal analysis, *Amer. J. Math.* 73 (1951), 485–496.
14. L. C. Piccinini, G. Stampacchia, and G. Vidossich, *Ordinary Differential Equations in* $\mathbb{R}^n$, #39, Applied Mathematical Sciences Series, Springer-Verlag, New York, 1984.
15. C. A. Rogers, A less strange version of Milnor's proof of Brouwer's Fixed Point Theorem, *Amer. Math. Monthly*, 87(1980), 525–527.
16. Walter Rudin, *Principles of Mathematical Analysis*, 3rd edition, McGraw-Hill, New York, 1976.
17. ______, *Real and Complex Analysis*, 3rd edition, McGraw-Hill, New York, 1987.
18. J. Schauder, Der Fixpunktsatz in Funktionalräumen, *Studia Math.* 2(1930), 171–180.
19. Michael Spivak, *Calculus on Manifolds*, Benjamin, Princeton, NJ, 1965.

Limerick

A teacher once, having some fun,
In presenting that two equals one,
Remained quite aloof
From his rigorous proof;
But his class was convinced and undone.

—Arthur White
Western Michigan University
Kalamazoo, MI 49008

NOTES

Packing a Box with Bricks

CHARLES H. JEPSEN
Grinnell College
Grinnell, IA 50112

The following problem was proposed for a recent International Mathematical Olympiad:

> A $2 \times 2 \times 12$ hole in a wall is to be filled with twenty-four $1 \times 1 \times 2$ bricks. In how many different ways can this be done if the bricks are indistinguishable?

(See [2]. The problem is C/1 on p. 14; a combinatorial solution is given on pp. 99–100.)

Suppose the hole in the wall is replaced by a box. If the box is assumed to have a fixed orientation (as in Problem 1 below), packing the box is identical to filling the hole. However, in box-packing it seems natural to take rigid motions into account. That is, we consider two packings to be the same if a rigid motion of the box transforms one packing into the other. (See Problem 2 below.) In this note, we present a solution of Problem 2 based on Burnside's Theorem and our matrix algebra solution of Problem 1. Our approach illustrates an interaction among several areas of undergraduate mathematics: geometry, combinatorics, linear algebra, and group theory.

Problem 1. A $2 \times 2 \times n$ box is to be packed with identical $1 \times 1 \times 2$ bricks. Suppose the box has a fixed top and front. Find a formula for a_n, the number of different packings of such a box.

Solution. Think of the $2 \times 2 \times n$ box as consisting of n layers, numbered 1 through n from bottom to top. Given a packing of the box, for each k, a slice at the top of the k^{th} layer intersects either 0 bricks, 2 adjacent bricks, or 4 bricks.
For any k, consider three types of packings at level k:

Type (a): The bottom k layers of the box are packed (so that a slice at the top of the k^{th} layer intersects 0 bricks). Denote by a_k the number of packings of type (a).

Type (b): The bottom k layers plus half of the $(k+1)^{\text{st}}$ layer are packed so that a slice at the top of the k^{th} layer intersects 2 bricks. Let b_k denote the number of packings of type (b).

Type (c): The bottom $k+1$ layers are packed so that a slice at the top of the k^{th} layer intersects 4 bricks. Let c_k be the number of packings of type (c).

Let also $a_0 = 1$, $b_0 = 0$, $c_0 = 0$.

If we consider the various possibilities in moving from the k^{th} to the $(k+1)^{\text{st}}$ layer, we get the recurrence relationships

$$a_{k+1} = 2a_k + b_k + c_k$$

$$b_{k+1} = 4a_k + b_k$$
$$c_{k+1} = a_k,$$

each coefficient being obtained by direct count. Let

$$Y_k = \begin{bmatrix} a_k \\ b_k \\ c_k \end{bmatrix}, \qquad A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

and get the matrix equations $Y_{k+1} = AY_k$, $k = 0, 1, \dots, n-1$. From linear algebra, Y_k can be expressed in terms of the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ and eigenvectors X_1, X_2, X_3 of A:

$$Y_k = \alpha_1 \lambda_1^k X_1 + \alpha_2 \lambda_2^k X_2 + \alpha_3 \lambda_3^k X_3,$$

where the constants $\alpha_1, \alpha_2, \alpha_3$ are derived from the initial condition $Y_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ by solving $PC = Y_0$ for C where $P = [X_1\ X_2\ X_3]$, $C = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$.

Straightforward algebraic computations on the above matrix A yield:

$$\lambda_1 = -1, \qquad \lambda_2 = 2 + \sqrt{3}, \qquad \lambda_3 = 2 - \sqrt{3},$$

$$X_1 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \qquad X_2 = \begin{bmatrix} 2+\sqrt{3} \\ 2+2\sqrt{3} \\ 1 \end{bmatrix}, \quad X_3 = \begin{bmatrix} 2-\sqrt{3} \\ 2-2\sqrt{3} \\ 1 \end{bmatrix},$$

$$\alpha_1 = -\frac{1}{3}, \qquad \alpha_2 = \frac{1}{6}, \qquad \alpha_3 = \frac{1}{6}.$$

Hence,

$$Y_k = \left(-\frac{1}{3}\right)(-1)^k \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} + \frac{1}{6}(2+\sqrt{3})^k \begin{bmatrix} 2+\sqrt{3} \\ 2+2\sqrt{3} \\ 1 \end{bmatrix} + \frac{1}{6}(2-\sqrt{3})^k \begin{bmatrix} 2-\sqrt{3} \\ 2-2\sqrt{3} \\ 1 \end{bmatrix}.$$

We will use a_k, b_k, c_k in Problem 2. The formula that solves Problem 1 is

$$a_n = \frac{1}{3}(-1)^n + \frac{1}{6}(2+\sqrt{3})^{n+1} + \frac{1}{6}(2-\sqrt{3})^{n+1}.$$

Problem 2. A $2 \times 2 \times n$ box is to be packed with identical $1 \times 1 \times 2$ bricks. The box has no fixed top or front so two packings are the same if a rigid motion of the box transforms one into the other. Find a formula for p_n, the number of different packings of such a box.

Solution. By direct count, we have $p_0 = 1$, $p_1 = 1$, $p_2 = 2$ so assume $n \geqslant 3$. Let G be the group of rigid motions of the box. If the corners of the box are numbered 1 through 8 (as in FIGURE 1), an element of G can be considered as a permutation of the set $\{1, \dots, 8\}$.

Hence, $G = \{e, \rho, \rho^2, \rho^3, \sigma, \rho\sigma, \rho^2\sigma, \rho^3\sigma\}$, where e = identity,

$$\rho = (1\ 2\ 3\ 4)(5\ 6\ 7\ 8),$$
$$\rho^2 = (1\ 3)(2\ 4)(5\ 7)(6\ 8),$$

$$\begin{aligned}\rho^3 &= (1\ 4\ 3\ 2)(5\ 8\ 7\ 6),\\ \sigma &= (1\ 6)(2\ 5)(3\ 8)(4\ 7),\\ \rho\sigma &= (1\ 5)(2\ 8)(3\ 7)(4\ 6),\\ \rho^2\sigma &= (1\ 8)(2\ 7)(3\ 6)(4\ 5),\\ \rho^3\sigma &= (1\ 7)(2\ 6)(3\ 5)(4\ 8).\end{aligned}$$

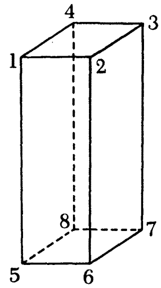


FIGURE 1

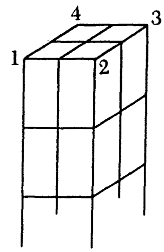


FIGURE 2

Let X be the set of all packings of the $2 \times 2 \times n$ box (with fixed orientation as in Problem 1). We can consider G as a group of maps on X: For $\phi \in G$, $x \in X$, let $\phi(x)$ be the packing that results from applying the rigid motion ϕ to the packing x. For $x, y \in X$, define $x \equiv y$ if there exists $\phi \in G$ such that $y = \phi(x)$. An easy check shows that $\equiv$ is an equivalence relation on X and that p_n is the number of equivalence classes of $\equiv$. To count p_n, we use the following result.

THEOREM (Burnside).

$$p_n = \frac{1}{|G|} \sum_{\phi \in G} \Psi(\phi),$$

where $|G|$ is the order of G and $\Psi(\phi)$ is the number of elements in X that are fixed under ϕ.

(See [1] for a discussion and proof.)

Given $\phi \in G$, we need to count the number of packings of the box that are invariant under the rigid motion ϕ. Observe first that by symmetry $\Psi(\rho) = \Psi(\rho^3)$, $\Psi(\sigma) = \Psi(\rho^2\sigma)$, $\Psi(\rho\sigma) = \Psi(\rho^3\sigma)$.

$\phi = e$: Since all packings are invariant under the identity motion, $\Psi(e) = a_n$.

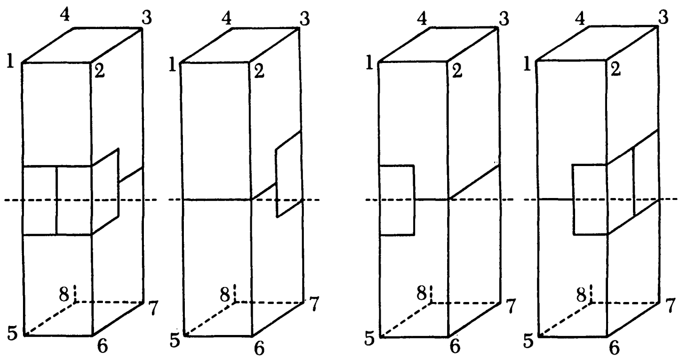
$\phi = \rho$: For a packing to be fixed under a 90° rotation, the bricks must be placed as in FIGURE 2. Hence,

$$\Psi(\rho) = \begin{cases} 1, & \text{if } n \text{ is even} \\ 0, & \text{if } n \text{ is odd.} \end{cases}$$

$\phi = \rho^2$: If a packing is to be invariant under such a 180° rotation, a slice at the top of the k^{th} layer must intersect 0 bricks or 4 bricks for any k. The number d_n of such packings can be derived by an argument similar to that in Problem 1. (Here the matrix equations are $Y_{k+1} = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} Y_k$, $Y_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.) Omitting the details, we get

$$\Psi(\rho^2) = d_n = \frac{\sqrt{2}}{4}(1+\sqrt{2})^{n+1} - \frac{\sqrt{2}}{4}(1-\sqrt{2})^{n+1}.$$

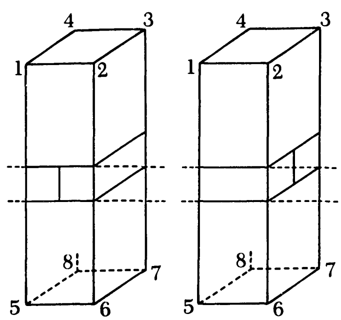
$\phi = \sigma$: Suppose first that n is even, say $n = 2m$, and consider the three types of packings at level m. A packing of type (a) determines a packing of the $2 \times 2 \times 2m$ box that is invariant under σ (use σ to rotate the packing of the bottom m layers to pack the top m layers). The same is true for each packing of type (c). Also, half of the packings of type (b) yield packings of the entire box that are invariant under σ and half do not (FIGURE 3).



invariant under σ not invariant under σ

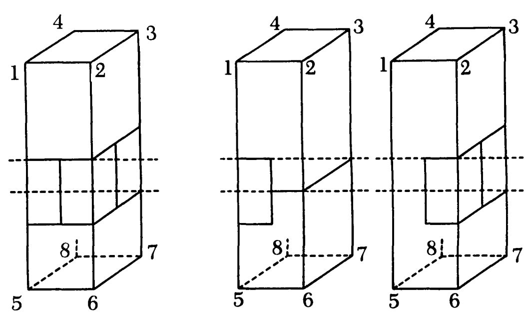
FIGURE 3

Now let $n = 2m + 1$ and again consider the three types of packings at level m. Here a packing of type (a) determines two packings of the $2 \times 2 \times (2m + 1)$ box that are invariant under σ (FIGURE 4). No packing of type (c) leads to a packing of the whole box that is invariant under σ, and half of the packings of type (b) lead to such packings (FIGURE 5).



invariant under σ

FIGURE 4



not invariant under σ invariant under σ

FIGURE 5

Hence, we have

$$\Psi(\sigma)=\begin{cases} a_m+\frac{1}{2}b_m+c_m, & \text{if } n=2m\\ 2a_m+\frac{1}{2}b_m, & \text{if } n=2m+1.\end{cases}$$

$\phi=\rho\sigma$: Let $n=2m$. As in the preceding case, the a_m+c_m packings of types (a) and (c) at level m lead to packings of the $2\times 2\times 2m$ box that are invariant under $\rho\sigma$. However, the 90° rotation ρ prevents a packing of type (b) at level m from yielding a packing of the entire box that is invariant under $\rho\sigma$. If $n=2m+1$, the rotation ρ prohibits any packing from being invariant under $\rho\sigma$. Thus

$$\Psi(\rho\sigma)=\begin{cases} a_m+c_m, & \text{if } n=2m\\ 0, & \text{if } n=2m+1.\end{cases}$$

From Burnside's Theorem, we have

$$\begin{aligned} p_n&=\tfrac{1}{8}\big(\Psi(e)+2\Psi(\rho)+\Psi(\rho^2)+2\Psi(\sigma)+2\Psi(\rho\sigma)\big)\\ &=\begin{cases}\frac{1}{8}\big(a_{2m}+2+d_{2m}+2\big(a_m+\frac{1}{2}b_m+c_m\big)+2(a_m+c_m)\big), & \text{if } n=2m\\ \frac{1}{8}\big(a_{2m+1}+0+d_{2m+1}+2\big(2a_m+\frac{1}{2}b_m\big)+2(0)\big), & \text{if } n=2m+1\end{cases}\end{aligned}$$

or

$$\begin{cases} p_{2m}=\frac{1}{8}(a_{2m}+b_{m+1}+4c_m+d_{2m}+2),\\ p_{2m+1}=\frac{1}{8}(a_{2m+1}+b_{m+1}+d_{2m+1}).\end{cases}$$

Substituting values and simplifying, we get

$$\begin{aligned} p_{2m}=&\frac{7}{24}-\frac{1}{12}(-1)^m+\frac{1}{48}(2+\sqrt{3})^{2m+1}+\frac{1}{48}(2-\sqrt{3})^{2m+1}\\ &+\frac{1}{24}(7+3\sqrt{3})(2+\sqrt{3})^m+\frac{1}{24}(7-3\sqrt{3})(2-\sqrt{3})^m\\ &+\frac{\sqrt{2}}{32}(1+\sqrt{2})^{2m+1}-\frac{\sqrt{2}}{32}(1-\sqrt{2})^{2m+1},\end{aligned}$$

TABLE 1

n	a_n	p_n
0	1	1
1	2	1
2	9	2
3	32	7
4	121	26
5	450	71
6	1,681	258
7	6,272	857
8	23,409	3,148
9	87,362	11,300
10	326,041	41,841
11	1,216,800	154,140
12	4,541,161	573,201

$$p_{2m+1} = -\frac{1}{24} - \frac{1}{12}(-1)^m + \frac{1}{48}(2+\sqrt{3})^{2m} + \frac{1}{48}(2-\sqrt{3})^{2m} + \frac{1}{24}(1+\sqrt{3})(2+\sqrt{3})^m + \frac{1}{24}(1-\sqrt{3})(2-\sqrt{3})^m + \frac{\sqrt{2}}{32}(1+\sqrt{2})^{2m} - \frac{\sqrt{2}}{32}(1-\sqrt{2})^{2m}.$$

The table gives values of a_n and p_n for $n = 0, 1, \ldots, 12$.

Note that $\lim_{n\to\infty} a_n/p_n = 8$. This agrees with the intuitive notion that, for large n, a given packing of a $2 \times 2 \times n$ box will in general be transformed into a different packing by a rigid motion of the box. Consequently, most of the p_n packings give rise to 8 packings among the a_n.

REFERENCES

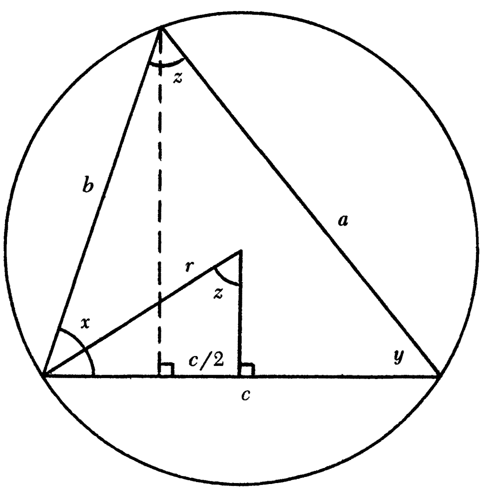
1. C. Berge, *Principles of Combinatorics*, Academic Press, New York, 1971.
2. Murray S. Klamkin, *International Mathematical Olympiads 1979–1985*, New Mathematical Library, Vol. 31, MAA, Washington, DC, 1986.

Proof without Words:
$\sin(x+y) = \sin x \cos y + \cos x \sin y \quad \text{for } x+y<\pi$

$$c = a\cos y + b\cos x$$

$$r = 1/2 \Rightarrow \sin z = (c/2)/(1/2) = c, \quad \sin x = a, \quad \sin y = b;$$

$$\sin(x+y) = \sin(\pi - (x+y)) = \sin z = c = \sin x \cos y + \sin y \cos x$$



—Sidney H. Kung
Jacksonville University
Jacksonville, FL 32211

$$p_{2m+1} = -\frac{1}{24} - \frac{1}{12}(-1)^m + \frac{1}{48}(2+\sqrt{3})^{2m} + \frac{1}{48}(2-\sqrt{3})^{2m} + \frac{1}{24}(1+\sqrt{3})(2+\sqrt{3})^m + \frac{1}{24}(1-\sqrt{3})(2-\sqrt{3})^m + \frac{\sqrt{2}}{32}(1+\sqrt{2})^{2m} - \frac{\sqrt{2}}{32}(1-\sqrt{2})^{2m}.$$

The table gives values of a_n and p_n for $n = 0, 1, \ldots, 12$.

Note that $\lim_{n\to\infty} a_n/p_n = 8$. This agrees with the intuitive notion that, for large n, a given packing of a $2 \times 2 \times n$ box will in general be transformed into a different packing by a rigid motion of the box. Consequently, most of the p_n packings give rise to 8 packings among the a_n.

REFERENCES

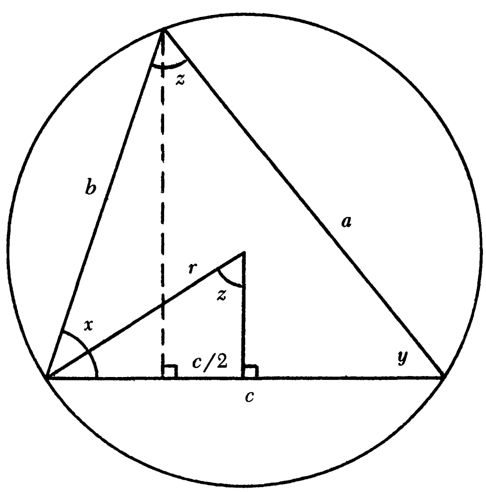
1. C. Berge, *Principles of Combinatorics*, Academic Press, New York, 1971.
2. Murray S. Klamkin, *International Mathematical Olympiads 1979–1985*, New Mathematical Library, Vol. 31, MAA, Washington, DC, 1986.

Proof without Words: $\sin(x+y) = \sin x \cos y + \cos x \sin y$ for $x+y<\pi$

$$c = a\cos y + b\cos x$$

$$r = 1/2 \Rightarrow \sin z = (c/2)/(1/2) = c, \quad \sin x = a, \quad \sin y = b;$$

$$\sin(x+y) = \sin(\pi - (x+y)) = \sin z = c = \sin x \cos y + \sin y \cos x$$



—Sidney H. Kung
Jacksonville University
Jacksonville, FL 32211

Complements of Linear Subspaces

DARYL TINGLEY
University of New Brunswick
Fredericton, N.B., Canada E3B 5A3

1. Introduction In "Covers by Linear Subspaces" [2], A. Todd shows that a finite dimensional linear space over an uncountable field has no countable cover by proper linear subspaces. This result is used to prove that if E is a finite dimensional linear space over an uncountable field, and if $\underline{F}$ is a countable collection of proper linear subspaces of E, each of the same dimension k, then there is a common complement G for $\underline{F}$ in E (i.e., a subspace G such that $E = F \oplus G$ for each F in $\underline{F}$). This result was earlier proved by N. J. Lord in [1] when the field is $\mathbb{R}$ or $\mathbb{C}$. Lord used the Baire Category Theorem, while Todd uses only elementary techniques. Both authors show that there are, in fact, an uncountable number of choices for G.

Using a modification of Todd's methods, this paper obtains similar results whenever the cardinality of the underlying field is greater than the cardinality of the collection of subspaces in question. (When the field is finite these cardinalities may be equal.) For infinite fields the results and proofs remain valid, virtually unchanged. When the field is finite, however, the results are slightly, but only slightly, more difficult to handle. The extra difficulties are most evident when the number of elements of a particular collection of subspaces is counted. The peculiarities of the arithmetic of infinite cardinals allow simpler constructions to be used, occasionally giving stronger results, when the field is infinite.

The results presented here are, perhaps, more interesting when the field in question is finite. The proofs are designed with this in mind but, for the most part, remain valid when the field is infinite.

After generalizing the work in [2], we consider some interesting combinatorial questions. In [1], Lord asked: If E^n is an n-dimensional linear space over the finite field K, and if $1 \leq k \leq n-1$, what is the maximum number of distinct k-dimensional subspaces of E^n that have a common complement? In this paper we answer this question as well as the following closely related question: With E^n, K, and k as above, what is the maximum value for m such that every m k-dimensional subspaces of E^n have a common complement?

One might also ask how many common complements a given collection of subspaces has. We derive a lower bound in Theorem 5, but it appears difficult to determine the exact number. However, for the special case of a single subspace the number of complements is determined in Theorem 6. This result is used to answer the question of Lord referred to above (Theorem 7), as well as to calculate the number of complements for certain pairs of subspaces in Section 3.

2. Theorems and proofs Where the proofs differ little from those in [2] they are omitted, and the reader is referred to that paper.

In what follows, E^n will be used to represent an n-dimensional linear space over the field K, while E is used for an arbitrary linear space, over K. Subspaces of E will be designated by F, G, H, and V. The symbols $\underline{F}$ and $\underline{G}$ will be used to designate collections of subspaces.

For a set S, $|S|$ means the cardinality of S. The condition $|S| < |K| + 1$ is frequently used. If $|K|$ were finite, our primary concern, this could be stated as $|S| \leqq |K|$, while if $|K|$ were infinite it is equivalent to $|S| < |K|$. So that the results remain valid for $|K|$ infinite, we use the more awkward condition.

The first few results (which are similar to results in [2]) concern covers by linear subspaces. The proofs of [2], Theorem 2 and our Theorem 4 show how such results relate to common complements for a collection of subspaces.

THEOREM 1. *Let E be a linear space over the field K and let $\underline{F}_1$ and $\underline{F}_2$ be families of linear subspaces of E. Suppose that $|\underline{F}_1 \cup \underline{F}_2| < |K| + 1$. Then if $\underline{F}_1 \cup \underline{F}_2$ covers E, either $\underline{F}_1$ or $\underline{F}_2$ covers E.*

Proof. Assume neither $\underline{F}_1$ nor $\underline{F}_2$ covers E and let x and y be elements of $E \setminus \cup \underline{F}_1$ and $E \setminus \cup \underline{F}_2$ respectively. Since $\underline{F}_1 \cup \underline{F}_2$ covers E, x and y are distinct and not linearly dependent.

Let $L = \{rx + (1-r)y\colon r \in K\}$. Then $|L| = |K|$ and $y - x \notin L$ (or else x would be in span $\{y\}$). If an element F of $\underline{F}_1 \cup \underline{F}_2$ contains two different points of $L \cup \{y - x\}$, then, since F is a linear subspace, F would contain both x and y, a contradiction. Thus $\underline{F}_1 \cup \underline{F}_2$ covers at most $|\underline{F}_1 \cup \underline{F}_2|$ elements of $L \cup \{y - x\}$. Since $|\underline{F}_1 \cup \underline{F}_2| < |K| + 1 = |L \cup \{y - x\}|$, $\underline{F}_1 \cup \underline{F}_2$ does not cover $L \cup \{y - x\}$, contradicting the assumption that $\underline{F}_1 \cup \underline{F}_2$ covers E.

This is a slight generalization of Proposition 2 in [2]. The proof there is valid, with minor changes, whenever $|F_1 \cup F_2| < |K|$. The modification made here is to consider $L \cup \{y - x\}$ rather than the set L.

In [1], Lord gives an ad hoc argument to show that a countable collection of lines through the origin in Euclidean 2-space has a common complement: Simply notice that they cannot cover the unit circle, and a line through the origin and any point not covered is a common complement. In [2], Todd suggests that his proof of Theorem 1 formalizes this argument. Applied to Lord's setting, Todd's proof relies on the fact that a countable collection of lines through the origin cannot cover a line *not* through the origin. Thus he has used a line rather than a circle; after all, circles don't exist in general linear spaces, while lines do.

The points of the circle in Lord's result and the line in Todd's version of Theorem 1 are only used to count the lines through the origin of the space. In the case of the circle there are, in a sense, too many points: Both x and $-x$ correspond to the same line. In the case of the line, there are not enough points: The line through the origin, parallel to the given line is not accounted for. For infinite cardinalities, none of this matters, but for finite cardinalities, to achieve the best results, a set that corresponds in a $1-1$ fashion with all the lines through the origin is desirable. The set $L \cup \{y - x\}$ is such a set.

COROLLARY. *If E is a linear space over a field K and $\underline{F}$ is any cover of E by proper linear subspaces of E such that $|\underline{F}| < |K| + 1$, then each cofinite subcollection of $\underline{F}$ covers E.*

Notice that a collection $\underline{F}$ satisfying the hypothesis of this corollary cannot be finite. If it were, the empty collection would be a cofinite subcollection of $\underline{F}$. This establishes part of the following theorem.

THEOREM 2. *If either E is a finite dimensional linear space or K is a finite field then E has no cover by a family of proper linear subspaces $\underline{F}$ such that $|\underline{F}| < |K| + 1$.*

Proof. If $|K|$ is finite the result is established by the above remarks, while if E is finite dimensional a proof similar to [2, Theorem 1] suffices.

Theorem 3 is a generalization of the main result in [1] and [2], the proof of which is similar to that in [2].

THEOREM 3. *Let E^n be a finite dimensional linear space over the field K and let $\underline{F}$ be a collection of k-dimensional subspaces of E^n with $|\underline{F}| < |K| + 1$. Then there is a linear subspace G of E^n such that G is a common complement for $\underline{F}$.*

The finite dimension hypothesis cannot, in general, be eliminated from the above two theorems. To see this, recall (cf. [2], Proposition 1) the well-known fact that any infinite dimensional linear space has a countable cover by proper linear subspaces. Such a collection has no common complement. It may be worth noting that if E is infinite dimensional, K is a finite field, and $\underline{F}$ is a collection of subspaces of E with the same finite codimension, then the conclusion of Theorem 3 remains valid, the proof being much the same.

THEOREM 4. *For every linear space E over the field K and for every cardinal number β, $0 < \beta < \dim E$, there is a collection of $|K| + 1$ distinct β-dimensional subspaces of E that cover a $(\beta + 1)$-dimensional subspace of E and have no common complement.*

Proof. Let G be any subspace of E of dimension $\beta - 1$. Let $v \in E \setminus G$ and let $V = G \oplus \operatorname{span}\{v\}$. Then $\dim V = \beta < \dim E$, so let $y \in E \setminus V$. The line $L = \{rv + y\colon r \in K\}$ does not intersect V (or y would be in V). Thus for each $r \in K$, letting $x_r = rv + y$ and $F_r = G \oplus \operatorname{span}\{x_r\}$ we have $\dim F_r = \beta$, $F_r \neq V$ and G has codimension one in F_r, as well as in V. It follows that $F_r \cap V = G$ for all $r \in K$. This is illustrated in FIGURE 1 for $n = 3$ and $\beta = 2$.

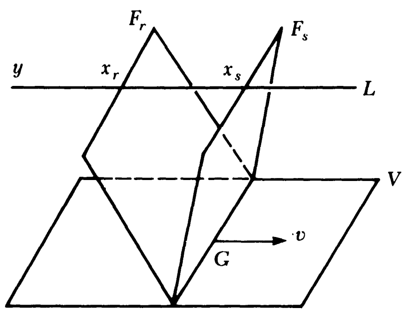


FIGURE 1

Let r and s be in K, $r \neq s$. If $F_r = F_s$ it follows that v is in F_r, and hence $V \subset F_r$. This contradicts $V \cap F_r = G$. Thus if $\underline{F} = \{F_r\colon r \in K\} \cup \{V\}$, $\underline{F}$ is a collection of $|K| + 1$ distinct subspaces of E, each with dimension β. Furthermore, it is easy to see that $\cup \underline{F} = G \oplus \operatorname{span}\{v\} \oplus \operatorname{span}\{y\}$, a $(\beta + 1)$-dimensional subspace of E. We claim that $\underline{F}$ has no common complement in E. Suppose that it does, and call this complement H. Then for $F \in \underline{F}$, $H \cap F = \{0\}$ and $H \oplus F = E$. Thus $H \cap (\cup \underline{F}) = \{0\}$ and $E \supset H \oplus \cup \underline{F} \supset H \oplus V = E$. Thus H is a complement for $\cup \underline{F}$. Consider $rv + y$ for some $r \in K$. Since $H \oplus V = E$ and $rv + y \notin V$, $rv + y = z + h$ where $z \in V$, $h \in H$ and $h \neq 0$. Thus $h = rv + y - z$. Since $rv + y$ and z are in $\cup \underline{F}$, and $\cup \underline{F}$ is a subspace of E, $h \in \cup \underline{F}$, contradicting $H \cap (\cup \underline{F}) = \{0\}$. Thus $\cup \underline{F}$ is a $(\beta + 1)$-dimensional subspace of E and $\underline{F}$ has no common complement in E.

Theorems 3 and 4 together show that every m distinct k-dimensional subspaces of E^n have a common complement if and only if $m < |K| + 1$. Hence for $|K|$ finite, the maximum possible value for m is $|K|$. This answers the second of the combinatorial questions mentioned in the introduction. (Note that the maximum value for m is independent of k.)

We have just considered a question about the number of subspaces having a common complement. We can also ask about the number of complements for a given subspace or, more generally, about the number of common complements for a given family of subspaces.

Lord [1] considered a countable family of k-dimensional subspaces in a finite dimensional real or complex linear space. He showed that there were many common complements (by showing the choice of basis for the complements is dense in a certain space). Todd [2] obtained a similar result for any uncountable field. Using the techniques of [2], we obtain the following generalization.

THEOREM 5. *Let E^n be a finite dimensional vector space over the field K. Let $\underline{F}$ be a collection of k-dimensional subspaces of E, $0 < k < n$, with $|\underline{F}| < |K| + 1$. Then there are at least $|K| + 1 - |\underline{F}|$ common complements for $\underline{F}$ in E.*

Proof. Let $\underline{G}$ be the collection of all common complements for $\underline{F}$. Consider the collection $\underline{F} \cup \underline{G}$. Then $|\underline{F} \cup \underline{G}| = |\underline{F}| + |\underline{G}|$. Assume $|\underline{F} \cup \underline{G}| < |K| + 1$. Then $\underline{F} \cup \underline{G}$ does not cover E^n (Theorem 2), so let $v \in E^n \setminus \cup(\underline{F} \cup \underline{G})$. The collection $\{F \oplus \text{span}\,\{v\}\colon F \in \underline{F}\}$ satisfies the hypothesis of Theorem 3, so it has a common complement H in E. It follows that $H \oplus \text{span}\,\{v\}$ is a common complement for $\underline{F}$, so $H \oplus \text{span}\,\{v\} \in \underline{G}$, contradicting the choice of v. Thus $|\underline{F} \cup \underline{G}| \geqq |K| + 1$, so $|G| \geqq |K| + 1 - |F|$, proving the theorem.

For $|K|$ infinite (such as $\mathbb{R}$ or $\mathbb{C}$) $|\underline{G}| \geqq |K| + 1 - |\underline{F}|$ is equivalent to $|\underline{G}| \geqq |K|$. However, the cardinality of E^n is $|K|^n = |K|$. The dimension of each $G \in \underline{G}$ is $n - k$ so there are no more than $|K|^{n-k} = |K|$ choices of basis for common complements of $\underline{F}$. Thus we have the following:

COROLLARY. *With E^n and $\underline{F}$ as above and $|K|$ infinite, there are $|K|$ common complements for $\underline{F}$.*

For $|K|$ finite the problem of finding the exact number of common complements is more difficult (hence more interesting). Indeed, the next result shows (for instance, by letting $|\underline{F}| = 1$, $n = 4$, and $k = 2$) that there can be more than $|K| + 1 - |\underline{F}|$ common complements for $\underline{F}$. That result also shows (by letting $|\underline{F}| = 1$, $n = 2$ and $k = 1$) that the lower bound of $|K| + 1 - |\underline{F}|$ common complements can be obtained. (We suspect that this is the only situation where that lower bound is obtained.) In the final section, we consider some relatively simple collections to see just how complicated (at least using our techniques) the problem is.

THEOREM 6. *Let E^n be a linear space over the field K. Let F be any k-dimensional subspace of E^n. Then there are $|K|^{k(n-k)}$ distinct complements for F in E.*

Proof. Let G_0 be any complement for F. Then $\dim G_0 = n - k$; let $y_1, \ldots, y_{n-k}$ be a basis for G_0. Consider the $n - k$ translates of F given by $F + y_i$. Note that the sets $F + y_i$ are disjoint. FIGURE 2 illustrates this for $n = 3$, $k = 1$.

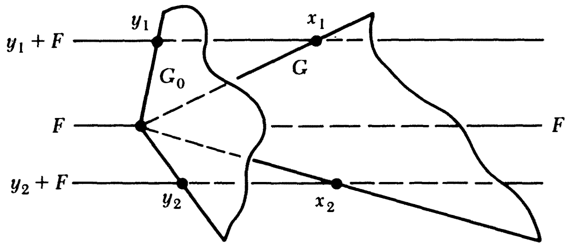


FIGURE 2

Let G be any complement for F. Then it is not hard to show that G intersects each $F + y_i$ in exactly one point. On the other hand, if $x_i \in F + y_i$, $i = 1, \ldots, n-k$, then it is not hard to show that span $\{x_1, \ldots, x_{n-k}\} \oplus F = E$, so span $\{x_1, \ldots, x_{n-k}\}$ is a complement for F.

Thus, there is a one-to-one correspondence between $n-k$ tuples $(x_1, \ldots, x_{n-k})$, $x_i \in F + y_i$ and complements for F. Since $|F + y_i| = |F| = |K|^k$ there are $|K|^{k(n-k)}$ complements for F.

The question of Lord's, referred to in §1, can now be answered.

THEOREM 7. *Let E^n be an n-dimensional linear space over the field K. The largest collection $\underline{F}$ of k-dimensional subspaces of E that has a common complement in E contains $|K|^{k(n-k)}$ subspaces.*

Proof. Let G be any $(n-k)$-dimensional subspace of E. The previous result shows that the collection of complements for G, call it $\underline{F}$, consists of $|K|^{k(n-k)}$ subspaces. Thus $\underline{F}$ has a common complement $- G$.

On the other hand, if a collection $\underline{F}$ of more than $|K|^{k(n-k)}$ k-dimensional subspaces of E has a common complement G, then G would have more than $|K|^{k(n-k)}$ complements. Since $\dim G = n - k$, this contradicts the previous theorem.

3. Examples We have not answered, in general, the question of the number of common complements for a given collection of subspaces. The corollary to Theorem 5 does provide an answer for $|K|$ infinite, and Theorems 4, 5 and 6 provide partial answers for $|K|$ finite. This would appear to be a difficult problem for, as illustrated by the following examples, the number will depend on how these subspaces are placed in the space—specifically on the dimension of the intersection of every possible subcollection.

We will only prove one of these examples. (It was the hardest for us.) The others can be proven using techniques found in that proof.

(1) Let F_1 and F_2 be 2-dimensional subspaces of E^n with $F_1 \cap F_2 = \{0\}$. Then there are $|K|^{2n-4} - |K|^{2n-5} - |K|^{2n-6} + |K|^{2n-7}$ common complements for F_1 and F_2.

Proof. By Theorem 7, there are $|K|^{2(n-2)}$ complements of F_1 in E^n. We must decide how many of these are not complements for F_2. If a complement of F_1 is not a complement of F_2, it either contains F_2 or intersects F_2 in a line.

We first count the number of complements of F_1 that contain F_2. An $(n-2)$-dimensional subspace H is such a complement if and only if there is an $(n-4)$-dimensional subspace G such that $H = F_2 \oplus G$ and $E^n = F_1 \oplus F_2 \oplus G$.

There are $|K|^{4(n-4)}$ complements for $F_1 \oplus F_2$ (i.e., choices for G) and for each choice of H there are $|K|^{2(n-4)}$ complements for F_2 as a subspace of H (i.e., choices for G such that $H = F_2 \oplus G$). It follows that there are $|K|^{2(n-4)}$ choices for H; i.e., there are $|K|^{2(n-4)}$ complements of F_1 that contain F_2.

Next we determine the number of complements for F_1 that contain a line l in F_2, but do not contain F_2. Fix a line l in F_2. Then H is a complement for F_1 that contains l if and only if there is an $(n-3)$-dimensional subspace G such that $H = l \oplus G$ and $E^n = F_1 \oplus l \oplus G$. There are $|K|^{3(n-3)}$ complements for $F_1 \oplus l$ and there are $|K|^{n-3}$ complements for l, as a subspace of H, for each choice of H. Thus, there are $|K|^{2(n-3)}$ choices for H (complements for F_1 that contain l). However, as we have shown, $|K|^{2(n-4)}$ of these contain F_2, so there are $|K|^{2(n-3)} - |K|^{2(n-4)}$ complements of F_1 that contain l, but not F_2. Since there are $|K| + 1$ lines through the origin in F_2, there are $(|K| + 1)(|K|^{2(n-3)} - |K|^{2(n-4)})$ complements of F_1 that intersect F_2 in a line. The result now follows easily.

(2) Let F_1 and F_2 be 2-dimensional subspaces of E^n that intersect in a line l. Then there are $|K|^{2n-4} - |K|^{2n-5}$ common complements for $\{F_1, F_2\}$.

(3) Let F_1 and F_2 be distinct lines or hyperplanes in E^n. Then there are $|K|^{n-1} - |K|^{n-2}$ distinct common complements for $\{F_1, F_2\}$.

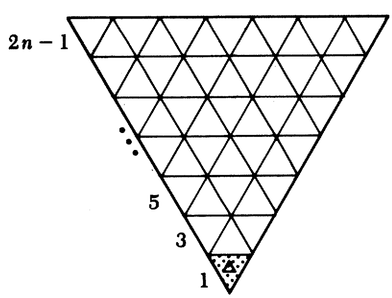
REFERENCES

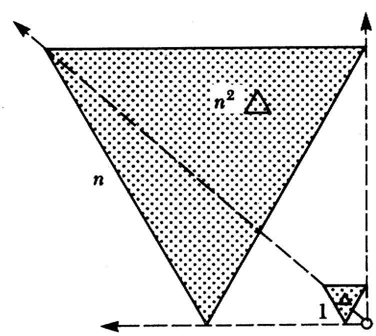
1. N. J. Lord, Simultaneous complements in finite-dimensional vector spaces, *Amer. Math. Monthly* 92 (1985), 492–493.
2. A. R. Todd, Covers by linear subspaces, this MAGAZINE 63 (1990), 339–342.

Proof without Words: The Sum of Odd Numbers

$$\Delta + 3 \cdot \Delta + \cdots + (2n-1) \cdot \Delta = A = n^2 \cdot \Delta$$

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





—JENŐ LEHEL
BUDAPEST, HUNGARY

There are $|K|^{4(n-4)}$ complements for $F_1 \oplus F_2$ (i.e., choices for G) and for each choice of H there are $|K|^{2(n-4)}$ complements for F_2 as a subspace of H (i.e., choices for G such that $H = F_2 \oplus G$). It follows that there are $|K|^{2(n-4)}$ choices for H; i.e., there are $|K|^{2(n-4)}$ complements of F_1 that contain F_2.

Next we determine the number of complements for F_1 that contain a line l in F_2, but do not contain F_2. Fix a line l in F_2. Then H is a complement for F_1 that contains l if and only if there is an $(n-3)$-dimensional subspace G such that $H = l \oplus G$ and $E^n = F_1 \oplus l \oplus G$. There are $|K|^{3(n-3)}$ complements for $F_1 \oplus l$ and there are $|K|^{n-3}$ complements for l, as a subspace of H, for each choice of H. Thus, there are $|K|^{2(n-3)}$ choices for H (complements for F_1 that contain l). However, as we have shown, $|K|^{2(n-4)}$ of these contain F_2, so there are $|K|^{2(n-3)} - |K|^{2(n-4)}$ complements of F_1 that contain l, but not F_2. Since there are $|K|+1$ lines through the origin in F_2, there are $(|K|+1)(|K|^{2(n-3)} - |K|^{2(n-4)})$ complements of F_1 that intersect F_2 in a line. The result now follows easily.

(2) Let F_1 and F_2 be 2-dimensional subspaces of E^n that intersect in a line l. Then there are $|K|^{2n-4} - |K|^{2n-5}$ common complements for $\{F_1, F_2\}$.

(3) Let F_1 and F_2 be distinct lines or hyperplanes in E^n. Then there are $|K|^{n-1} - |K|^{n-2}$ distinct common complements for $\{F_1, F_2\}$.

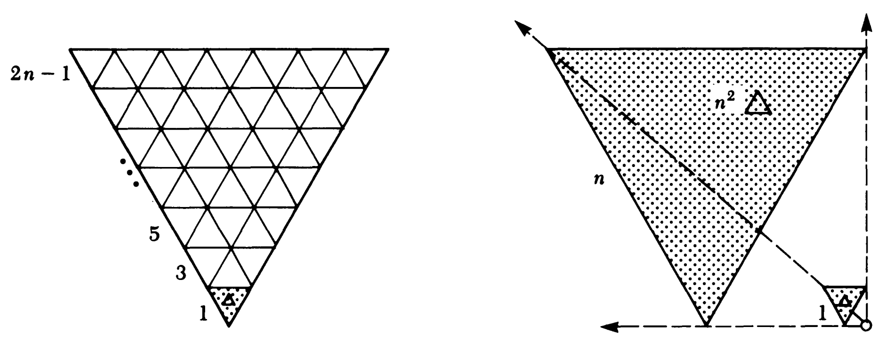
REFERENCES

1. N. J. Lord, Simultaneous complements in finite-dimensional vector spaces, *Amer. Math. Monthly* 92 (1985), 492–493.
2. A. R. Todd, Covers by linear subspaces, this MAGAZINE 63 (1990), 339–342.

Proof without Words: The Sum of Odd Numbers

$$\Delta + 3\cdot\Delta + \cdots + (2n-1)\cdot\Delta = A = n^2\cdot\Delta$$

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



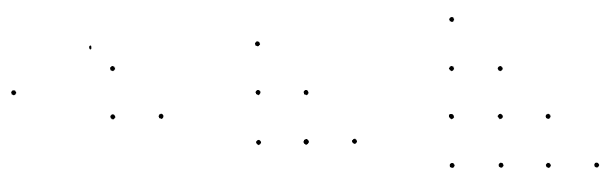
—JENŐ LEHEL
BUDAPEST, HUNGARY

Tetrahedral Numbers as Sums of Square Numbers

S. C. ALTHOEN
University of Michigan
Flint, MI 48502

C. B. LACAMPAGNE
Northern Illinois University
DeKalb, IL 60115

It is well known that the sum of the first consecutive positive integers 1 through n is the binomial coefficient $\binom{n+1}{2}$. This sum and its associated binomial coefficient may be expressed geometrically as a triangular number (FIGURE 1).

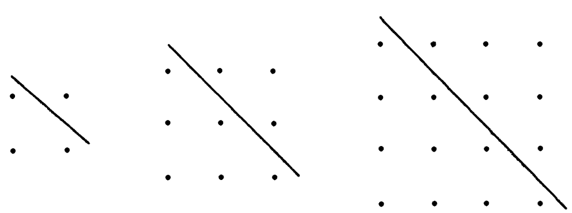


T_1 $\quad$ T_2 $\quad$ T_3 $\quad$ T_4

1 $\quad$ $1+2$ $\quad$ $1+2+3$ $\quad$ $1+2+3+4$

FIGURE 1

It is interesting to note as did Theon of Smyrna about 100 A.D. [1, p. 2] that the sum of two consecutive triangular numbers is a square number (FIGURE 2).



$T_1+T_2=2^2$ $\quad$ $T_2+T_3=3^2$ $\quad$ $T_3+T_4=4^2$

FIGURE 2

It seems remarkable that when we sum alternate squares starting with either 1 or 2 the result is the binomial coefficient $\binom{n+2}{3}$, where n^2 is the largest square in the sum. That is,

$$1^2+3^2+5^2+\cdots+n^2=\binom{n+2}{3} \tag{1}$$

and

$$2^2+4^2+6^2+\cdots+n^2=\binom{n+2}{3}, \tag{2}$$

where n is odd in (1) and even in (2). That sums (1) and (2) are related seems to have first been noted by Frederico Mariares [8]. (See [1, p. 35].) Mariares' article concludes with the question: "¿Las fórmulas ... serán realmente nuevas?"

The sum of alternating squares and related binomial coefficients may be expressed geometrically as a tetrahedral number. These are numbers 1, 4, 10, etc., associated with a tetrahedron as shown in FIGURE 3.

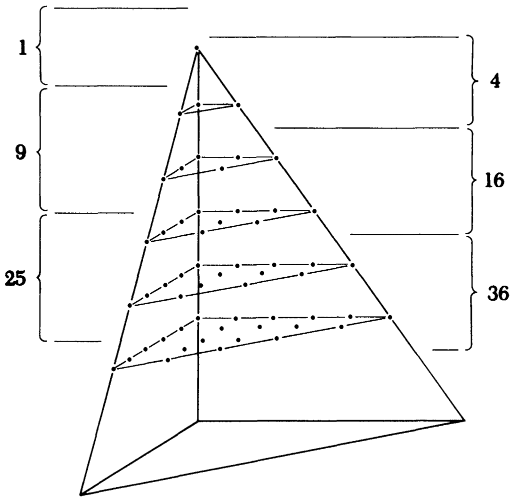


FIGURE 3

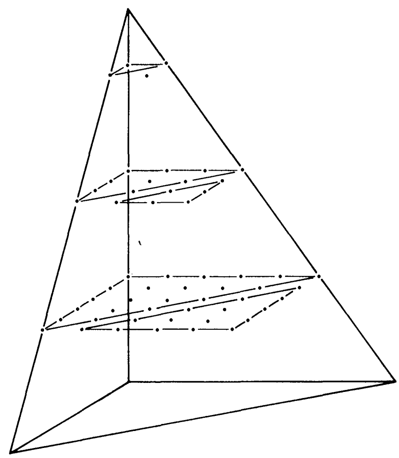


FIGURE 4

This is not so surprising in light of our consideration of the sum of two consecutive triangular numbers. That is, we can generate alternate square numbers in two ways from tetrahedral numbers. We may write (see FIGURE 4)

$$\begin{aligned} &1+3+6+10+15+21+\cdots \\ &\quad=(1+3)+(6+10)+(15+21)+\cdots \\ &\quad=2^2+4^2+6^2+\cdots \end{aligned}$$

or

$$\begin{aligned} &1+3+6+10+15+21+\cdots \\ &\quad=1+(3+6)+(10+15)+(21+28)+\cdots \\ &\quad=1^2+3^2+5^2+7^2+\cdots . \end{aligned}$$

Equations (1) and (2) make good exercises in proof by induction, in the use of generating functions, or in the manipulation of Σ-notation. For example, (1) may be obtained from $\Sigma_{i=0}^{(n-1)/2}(1+2i)^2$ by known results for sums of consecutive squares and first powers.

Equations (1) and (2) may be generalized slightly by multiplying by any square, say s^{2k}, where k is a nonnegative integer:

$$(s^k)^2+(3s^k)^2+\cdots+(ns^k)^2=s^{2k}\binom{n+2}{3}, \text{ for } n \text{ odd}$$

$$(2s^k)^2+(4s^k)^2+\cdots+(ns^k)^2=s^{2k}\binom{n+2}{3}, \text{ for } n \text{ even}.$$

In particular,

$$2^2+6^2+10^2+\cdots+(2n)^2=4\binom{n+2}{3}$$

and

$$4^2+8^2+12^2+\cdots+(2n)^2=4\binom{n+2}{3}.$$

The question remains: Which other arithmetic progressions yield sums of squares that may be represented by $\binom{m}{3}$?

Let

$$\sum_{i=0}^{l}(a+ki)^2=\binom{m}{3}, \tag{3}$$

where a, k, and m are positive integers. Note that the formula for the sum of consecutive squares suggests that the left side of (3) is a cubic in l. Thus, it is reasonable to expect (and, in fact, necessarily the case) that the right side of (3) is of the form $\binom{bl+c}{3}$. For example, in (1) we have

$$l=(n-1)/2,\quad a=1,\quad k=2,\quad b=2,\quad\text{and}\quad c=3.$$

So we assume that

$$\sum_{i=0}^{l}(a+ki)^2=\binom{bl+c}{3}. \tag{4}$$

Notice that for $l=0$

$$a^2=\binom{c}{3} \tag{5}$$

so c is a positive integer and it follows immediately that b is also a positive integer. It is easy to see that $(a,c)=(1,3)$ and $(a,c)=(2,4)$ are solutions to (5). It is surprising that $(a,c)=(140,50)$ is also a solution. According to Dickson [**1**, p. 25], this solution was noted by Moret-Blanc [**10**] in 1876. Moret-Blanc actually notes the corresponding solution to (11) below. However, later in 1876, in the same volume of *Nouvelles Annales de Mathématiques*, Lucas [**6**] asks for a proof that $c=3$, 4, and 50 yield the only solutions to (5).

To find all solutions to (4) that result from the three sets of values for a and c obtained from (5), we consider the equations generated by letting $l=1$ and $l=2$ in (4):

$$k^2+2ak+2a^2=\binom{b+c}{3} \tag{6}$$

$$5k^2+6ak+3a^2=\binom{2b+c}{3}. \tag{7}$$

Any solution to (4) must certainly be a set of positive integers $\{a,k,b,c\}$ that satisfies (6) and (7). In fact, from (6) and (7) we find that any solution to (4) must satisfy

$$4ak + 7a^2 = 5\binom{b+c}{3} - \binom{2b+c}{3}. \tag{8}$$

(The converse is false but irrelevant.) Below we find all solutions to (8) and show that they yield the known results (1) and (2), or else fail to satisfy (6) and hence fail to satisfy (4).

Case 1. $a = 1$ and $c = 3$.
Equation (8) becomes

$$8k = -b^3 + 2b^2 + 11b - 6.$$

Working modulo 8 it is easy to check that the cubic is congruent to 0 precisely when $b \equiv 2 \pmod 8$. Since k and b must be positive integers, $0 < b < 5$. Thus, the only possible solution is $b = 2$ and $k = 2$, which gives the familiar equation (1).

Case 2. $a = 2$ and $c = 4$.
Equation (8) becomes

$$16k = -b^3 + 3b^2 + 26b - 24.$$

Again, k and b are positive integers so $0 < b < 7$ and a quick check reveals two solutions, $b = 2$, $k = 2$ and $b = 4$, $k = 4$. The first solution yields the familiar equation (2). The second solution is extraneous; it does not satisfy (6).

Case 3. $a = 140$ and $c = 50$.
Equation (8) becomes

$$1120k = -b^3 + 49b^2 + 7202b - 117{,}600.$$

Reducing modulo 7 yields

$$0 \equiv -b^3 + 6b \pmod 7$$
$$0 \equiv b^3 + b = b(b^2 + 1) \pmod 7.$$

Thus, $b \equiv 0 \pmod 7$ and since k must be a positive integer, $15 < b < 107$. Now let $b = 7d$ in (8). Then

$$160k = -49d^3 + 7^3d^2 + 7202d - 16{,}800, \tag{9}$$

where $2 < d < 16$. Since in (9), $d \equiv 0 \pmod{32}$, the only solution is $d = 14$. Thus, in this case (8) has only one solution: $b = 98$, $k = 105$, which fails to satisfy (6).

Are there other pairs (a, c) that are solutions to (5)? This is equivalent to asking for solutions to the Diophantine equation

$$\binom{x+2}{3} = \frac{1}{6}x(x+1)(x+2) = y^2, \tag{10}$$

which, in turn, is equivalent to asking for solutions to the Diophantine equation

$$x(x+1)(2x+1) = 6y^2 \tag{11}$$

if x is even (replace x by $2x$ and y by $2y$ in (10)) or

$$x(2x-1)(2x+1) = 3y^2 \tag{12}$$

if x is odd (replace x by $2x - 1$ in (10)).

We mentioned above that in 1876 Lucas [6] asked for a proof that the only positive solutions to (10) are 1, 2, and 48, which correspond to $c = 3$, 4, and 50 in (5). In 1878 Meyl [9], who refers to himself as *ancien capitaine d'Artillerie, à La Haye*, showed that the only solutions to (10) are $x = -2, -1, 0, 1, 2$, and 48. These include our values of $c = 3, 4, 50$ along with three degenerate cases $c = 0$, 1, and 2, for which there really is no binomial coefficient. For x even, Meyl based his argument on the fact that (11) has solutions only for $x = -1, 0, 1$, and 24. Lucas believed that he had proved this in his "square pyramid equals a square" paper of 1877 [7]. However, Lucas's proof was flawed and a correct proof was first given by Watson in 1919 [12]. Ljunggren gave an arithmetic proof in 1952 [4]. For a history of this subject see Watson [12], Dickson [1], Guy [3], and Uchiyama [11], whose article was brought to our attention by Martin Gardner.

We thank Richard Guy for several helpful suggestions regarding this article.

REFERENCES

1. L. E. Dickson, *History of the Theory of Numbers*, Vol. 2, Diophantine Analysis, chapter I, Chelsea, New York, 1971, pp. 2, 25, 35.
2. M. Gardner, On the patterns and the unusual properties of figurate numbers, *Scientific American* 231 (1974), 116–120.
3. R. K. Guy, *Unsolved Problems in Number Theory*, Springer-Verlag, New York, 1980, p. 82.
4. W. Ljunggren, New solution of a problem proposed by E. Lucas, *Norsk Matematisk Tidsskrift* 34 (1952), 65–72.
5. E. Lucas, Question 1180, *Nouvelles Annales de Mathématiques* (2) 14 (1875), 336.
6. ______, Question 1194, *Nouvelles Annales de Mathématiques* (2) 15 (1876), 144.
7. ______, [Solution to] Question 1180, *Nouvelles Annales de Mathématiques* (2) 16 (1877), 429–432.
8. F. Mariares, Curiosidades arithméticas, *Revista de la Sociedad Mathemática Española* 2 (1913), 333–335.
9. A.-J.-J. Meyl, [Solution to] Question 1194, *Nouvelles Annales de Mathématiques* (2) 17 (1878), 464–467.
10. Moret-Blanc, [Solution to] Question 1180, *Nouvelles Annales de Mathématiques* (2) 15 (1876), 46–48.
11. S. Uchiyama, On some Diophantine equations, *Lecture Notes* (1984), Tokyo.
12. G. N. Watson, The problem of the square pyramid, *The Messenger of Mathematics* 48 (1919), 1–22.

Regular Polygons with Rational Area or Perimeter

R. B. KILLGROVE
108 N. 2nd Street, Unit C
Alhambra, CA 91801

D. W. KOSTER
University of Wisconsin
La Crosse, WI 54601

In this note we discuss two different approaches to the problem of determining which regular polygons, either inscribed within or circumscribed about the unit circle, possess rational area or rational perimeter. The first approach involves concepts usually encountered at the undergraduate level in a one-year course in modern algebra from texts such as [3] or [11], and as such it provides a nice application of abstract theory to a concrete problem involving the physical quantities of area and length. The second approach employs only material found in a traditional college algebra/trigonometry course and thus can be used to entice freshman students into actually doing some rigorous mathematics.

To facilitate the discussions we let $A_I(n)$ and $P_I(n)$ denote the area and perimeter of a regular n-gon inscribed within the unit circle and we let $A_c(n)$ and $P_c(n)$ denote the area and perimeter of a regular n-gon circumscribed about the unit circle. Using some elementary trigonometry, one quickly finds

$$A_I(n) = \frac{n}{2}\sin\left(\frac{2\pi}{n}\right) \qquad P_I(n) = 2n\sin\left(\frac{\pi}{n}\right)$$
$$A_c(n) = n\tan\left(\frac{\pi}{n}\right) \qquad P_c(n) = 2n\tan\left(\frac{\pi}{n}\right).$$

Note that $P_c(n) = 2A_c(n)$, and thus for circumscribed n-gons the area will be rational if, and only if, the perimeter is rational. Observe also that the requirement that the above formulas produce rational values leads one very naturally to the more general question of determining those rational multiples of π, at which some trigonometric function has a rational value. This topic has occurred frequently in the literature. For a purely geometrical approach (probably not accessible to undergraduates) see [4]. An elegant algebraic approach (again, probably not accessible to undergraduates) is outlined in problem 197.1 on page 144 of [9]. Our first approach is similar to that found in [6] and [7], while our second approach more closely follows that found in [1], [5], [8], and [10]. In what follows we let Q denote the field of rational numbers and we let $\omega = e^{2\pi i/n} = \cos(2\pi/n) + i\sin(2\pi/n)$ be a primitive nth root of unity. Letting $Q(\omega)$ denote the extension field of Q by ω, we will need in our first approach the elementary field theory result (see [11] pp. 289, 290 or [3] section 48) $[Q(\omega):Q] = \phi(n)$ where $\phi(n)$ is the value of the familiar Euler function giving the number of positive integers less than n that are relatively prime to n.

First approach. Let $n \geqslant 3$.

Consider first the case of an inscribed n-gon. Then, if $P_I(n) \in Q$, we have $\sin(\pi/n) \in Q$. So, $\cos(2\pi/n) = 1 - 2\sin^2(\pi/n) \in Q$ and hence $\sin^2(2\pi/n) = 1 - \cos^2(2\pi/n) \in Q$. Thus, both the real part and the square of the imaginary part of ω

lie in Q. It follows that $[Q(\omega):Q]=2$ and thus $\phi(n)=2$. In a similar vein, if $A_I(n)\in Q$, we have $\sin(2\pi/n)\in Q$. Hence $\cos^2(2\pi/n)=1-\sin^2(2\pi/n)\in Q$. Thus, the squares of both the real and the imaginary parts of ω lie in Q. It follows that $[Q(\omega):Q]=2$ or $[Q(\omega):Q]=4$ and thus $\phi(n)=2$ or $\phi(n)=4$.

Consider next the case of a circumscribed n-gon. Then, if $P_c(n)\in Q$ or if $A_c(n)\in Q$, we have $\tan(\pi/n)\in Q$. It follows that

$$\cos^2(\pi/n)=\frac{1}{\sec^2(\pi/n)}=\frac{1}{\tan^2(\pi/n)+1}\in Q.$$

So $\cos(2\pi/n)=2\cos^2(\pi/n)-1\in Q$ and hence $\sin^2(2\pi/n)=1-\cos^2(2\pi/n)\in Q$. Now, as above, since the real part and the square of the imaginary part of ω lie in Q, $[Q(\omega):Q]=2$ and thus $\phi(n)=2$.

In summary then, we need only consider circumscribed n-gons for which $\phi(n)=2$ and inscribed n-gons for which $\phi(n)=2$ or $\phi(n)=4$. However, these equations can be easily solved by making use of the familiar properties (see pp. 137–8 of [2]) that ϕ is a multiplicative function whose value at a prime power is given by $\phi(p^r)=p^{r-1}(p-1)$. To this end suppose that $\phi(n)=2$. Then no prime larger than 3 can be a divisor of n and thus $n=2^a3^b$. Further, if $b\neq 0$, then $b=1$, which forces $a=0$ or $a=1$ and thus $n=3$ or $n=6$. On the other hand, if $b=0$, then $a=2$ and thus $n=4$. Next suppose $\phi(n)=4$. Then no prime larger than 5 can be a divisor of n and thus $n=2^a3^b5^c$. Note that if $c\neq 0$, then $c=1$, forcing $b=0$, and $a=0$ or $a=1$. Thus, $n=5$ or $n=10$. Alternatively, suppose $c=0$. Then, if $b\neq 0$, we must have $b=1$ and $a=2$ leading us to $n=12$; while if $b=0$, we must have $a=3$ and thus $n=8$. To summarize, we have established that $\phi(n)=2\Leftrightarrow n\in\{3,4,6\}$ and $\phi(n)=4\Leftrightarrow n\in\{5,8,10,12\}$. Thus, the only possible inscribed n-gons with rational perimeter and the only possible circumscribed n-gons with either rational area or rational perimeter are those with $n\in\{3,4,6\}$. Similarly, the only possible inscribed n-gons with rational area must satisfy $n\in\{3,4,5,6,8,10,12\}$.

The table on p. 111 displays the areas and perimeters for these possibilities. Note that with the exceptions of $A_I(5)$ and $A_I(10)$ (whose squares are clearly irrational), the entries in the table require values of sine and tangent only at familiar arguments such as $\pi/6,\pi/4,\pi/3$, etc. The computation for the two "exceptional" entries is given below. From this table we see that the only circumscribed n-gon with rational area or perimeter is the square of area 4 and perimeter 8. With regard to inscribed n-gons we see that the square of area 2 and the dodecagon of area 3 are the only ones with rational area, while the only one with rational perimeter is the hexagon of perimeter 6. Another conclusion forced upon us by the table is that even though our question insisted only on rational values for area or perimeter, these rational values have turned out to be integers!

This pleasant "surprise" can be explained by delving a little more deeply into algebraic number theory than one usually does at the undergraduate level. To give some indication of the kind of argument used we will give an explanation for why the rational values produced by the formula for $A_I(n)$ must be integers and refer the interested reader to [6] and [7] for a more detailed discussion. Citing Lehmer [6], Niven ([7], Thm. 3.9) shows that for $n\neq 4$, $2\sin(2\pi/n)$ is an algebraic *integer* of degree $\phi(n)$, $\phi(n)/4$, or $\phi(n)/2$ according as $\gcd(n,8)<4$, $\gcd(n,8)=4$, or $\gcd(n,8)>4$. Thus, requiring

$$A_I(n)=\frac{n}{4}2\sin\left(\frac{2\pi}{n}\right)$$

to produce rational values amounts to requiring the algebraic integer $2\sin(2\pi/n)$ to be rational. But the only rational algebraic integers are in fact integers and thus $2\sin(2\pi/n)$ must be an integer. Since (as a rational number) $2\sin(2\pi/n)$ has degree one, and since (for polygons $n>2$) $\phi(n)\neq 1$, Lehmer's result forces the conclusion $\gcd(n,8)\geqslant 4$ and thus $n/4$ must be an integer also.

We now complete our discussion of this first approach by producing the values given in the table for $A_I(5)$ and $A_I(10)$. Consider the polynomial

$$x^5-1=(x-1)(x^4+x^3+x^2+x+1)$$
$$=(x-1)\left(x^2+\frac{1+\sqrt{5}}{2}x+1\right)\left(x^2+\frac{1-\sqrt{5}}{2}x+1\right)$$

whose roots are the five fifth roots of unity. Letting $\omega=e^{2\pi i/5}=\cos(2\pi/5)+i\sin(2\pi/5)$, we note that among the five roots $1,\omega,\omega^2,\omega^3,\omega^4$, ω is the only root having positive coefficients for both its real and imaginary parts. Thus, an application of the quadratic formula to the third factor above produces the value

$$\omega=\frac{\sqrt{5}-1}{4}+\frac{\sqrt{2\sqrt{5}-10}}{4}i.$$

Thus,

$$\sin(2\pi/5)=\frac{\sqrt{2\sqrt{5}+10}}{4},$$

$$\cos(2\pi/5)=\frac{\sqrt{5}-1}{4},\quad\text{and}\quad \sin(\pi/5)=\sqrt{\frac{1-\cos(2\pi/5)}{2}}=\frac{\sqrt{10-2\sqrt{5}}}{4}$$

yielding the values in the table.

		inscribed n-gon		circumscribed n-gon	
n	$\phi(n)$	$A_I(n)$	$P_I(n)$	$A_c(n)$	$P_c(n)$
3	2	$\frac{3\sqrt{3}}{4}$	$3\sqrt{3}$	$3\sqrt{3}$	$6\sqrt{3}$
4	2	2	$4\sqrt{2}$	4	8
6	2	$\frac{3\sqrt{3}}{2}$	6	$2\sqrt{3}$	$4\sqrt{3}$
5	4	$\frac{5\sqrt{10+2\sqrt{5}}}{8}$			
8	4	$2\sqrt{2}$			
10	4	$\frac{5\sqrt{10-2\sqrt{5}}}{4}$			
12	4	3			

Second approach. As stated above, this approach will employ only material from a traditional college algebra/trigonometry course. In particular we will need the familiar "Rational Root Theorem" to the effect that if b/c, in lowest terms, is a root of the polynomial

$$f(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$$

with integer coefficients, then b divides a_0 and c divides a_n. To set the stage further for using this type of elementary material to answer the question about the possible rational values of $A_I(n)$, $P_I(n)$, $A_c(n)$, or $P_c(n)$, we point out that in a first encounter with trigonometry, students often memorize (among other things) the "double angle" formula $\cos(2\theta) = 2\cos^2(\theta) - 1$. Then, in the exercises, they are usually asked to derive the additional formulas $\cos(3\theta) = 4\cos^3(\theta) - 3\cos(\theta)$ and $\cos(4\theta) = 8\cos^4(\theta) - 8\cos^2(\theta) + 1$. This naturally gives rise to the question about a similar formula expressing $\cos(n\theta)$ as a polynomial in $\cos(\theta)$. (Of course, those students who continue to pursue mathematics and study numerical analysis will see this again when confronted with Chebyshev polynomials.) Using the notation T_n to denote the desired polynomial satisfying $T_n(\cos(\theta)) = \cos(n\theta)$, we exhibit the first few polynomials in the table immediately below.

$\cos(0\theta) = 1$	$T_0(x) = 1$
$\cos(1\theta) = \cos(\theta)$	$T_1(x) = x$
$\cos(2\theta) = 2\cos^2(\theta) - 1$	$T_2(x) = 2x^2 - 1$
$\cos(3\theta) = 4\cos^3(\theta) - 3\cos(\theta)$	$T_3(x) = 4x^3 - 3x$
$\cos(4\theta) = 8\cos^4(\theta) - 8\cos^2(\theta) + 1$	$T_4(x) = 8x^4 - 8x^2 + 1$

A practical method for generating further rows in this table is within an easy reach for those students whose previous exposure to mathematical induction included a taste of recursion. For, by applying the formulas $\cos(\alpha \pm \beta) = \cos(\alpha)\cos(\beta) \mp \sin(\alpha)\sin(\beta)$, first with $\alpha = \theta$, $\beta = (n-1)\theta$ and then with $\alpha = (n-1)\theta$, $\beta = \theta$, we obtain

$$\cos(n\theta) = \cos(\theta + (n-1)\theta) = \cos(\theta)\cos((n-1)\theta) - \sin(\theta)\sin((n-1)\theta)$$

and

$$\cos((n-2)\theta) = \cos((n-1)\theta - \theta) = \cos(\theta)\cos((n-1)\theta) + \sin(\theta)\sin((n-1)\theta).$$

Adding and rearranging gives

$$\cos(n\theta) = 2\cos(\theta)\cos((n-1)\theta) - \cos((n-2)\theta). \tag{1}$$

Restating (1) in terms of the polynomials T_m gives

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x). \tag{2}$$

Keeping in mind that $T_n(\cos(\theta)) = \cos(n\theta)$, we see that one important property enjoyed by all these polynomials is

$$T_n\left(\cos\left(\frac{2m\pi}{n}\right)\right) = 1 \quad \text{for any integer } m. \tag{3}$$

Now, using (2) and the initial values $T_0(x) = 1$ and $T_1(x) = x$ from the table one can apply mathematical induction to verify the following additional properties of these polynomials:

For $n \geqslant 0$, T_n has degree n and integer coefficients. (4)

For $n \geqslant 1$, the leading coefficient of $2T_n$ is 2^n. (5)

For $n \geqslant 0$, the coefficient of x^k in $2T_n(x)$ is divisible by 2^k. (6)

Since (4) and (5) are immediate, we proceed to (6) and note that we can assume $k > 0$ since if $k = 0$, $2^k = 1$, and there is nothing to prove. Also, if $k = n$, we are done by (5). To begin the induction we glance at the table above and see that (6) is valid for $n < 5$. We now assume (6) for $m < n$ and using (2) we write $2T_n(x) = 2x(2T_{n-1}(x)) - 2T_{n-2}(x)$. From this we see that the coefficient of x^{n-1} in $2T_n(x)$ equals $2a$ where a is the coefficient of x^{n-2} in $2T_{n-1}(x)$. Since, by the inductive hypothesis, a is divisible by 2^{n-2}, we have $2a$ is divisible by 2^{n-1}. Now suppose $0 < k < n-1$. Looking again at the above expression for $2T_n(x)$, we see that the coefficient of x^k in $2T_n(x)$ equals $2a - b$ where a is the coefficient of x^{k-1} in $2T_{n-1}(x)$ and b is the coefficient of x^k in $2T_{n-2}(x)$. Since, by the inductive hypothesis, 2^{k-1} divides a and 2^k divides b, we have $2a - b$ is divisible by 2^k.

We are now in a position to use these properties to answer the question about the possible rational values of $A_I(n)$, $P_I(n)$, $A_c(n)$, and $P_c(n)$. First note that by (5) and (6) we can conclude $2T_n(x/2)$ is a *monic* polynomial with integer coefficients. Thus, letting $f_n(x) = 2T_n(x/2) - 2$, we apply (3) to obtain $f_n(2\cos(2m\pi/n)) = 0$ for any integer m. So, we have just shown that

$$\text{for any integer } m,\ 2\cos\left(\frac{2m\pi}{n}\right) \text{ is a root of a monic polynomial with integer coefficients.} \tag{7}$$

Consider first an inscribed n-gon. Then, as before, if $P_I(n) \in Q$, we have $\sin(\pi/n) \in Q$; and thus $\cos(2\pi/n) = 1 - 2\sin^2(\pi/n) \in Q$. Similarly, if $A_I(n) \in Q$, we have $\sin(2\pi/n) \in Q$ and hence $\cos(4\pi/n) = 1 - 2\sin^2(2\pi/n) \in Q$. Considering next the case of a circumscribed n-gon, we see that if $A_c(n) \in Q$ or if $P_c(n) \in Q$, we have $\tan(\pi/n) \in Q$. It follows that $\cos(2\pi/n) \in Q$ since

$$1 + \cos(2\pi/n) = 2\cos^2(\pi/n) = \frac{2}{\sec^2(\pi/n)}$$
$$= \frac{2}{1 + \tan^2(\pi/n)} \in Q.$$

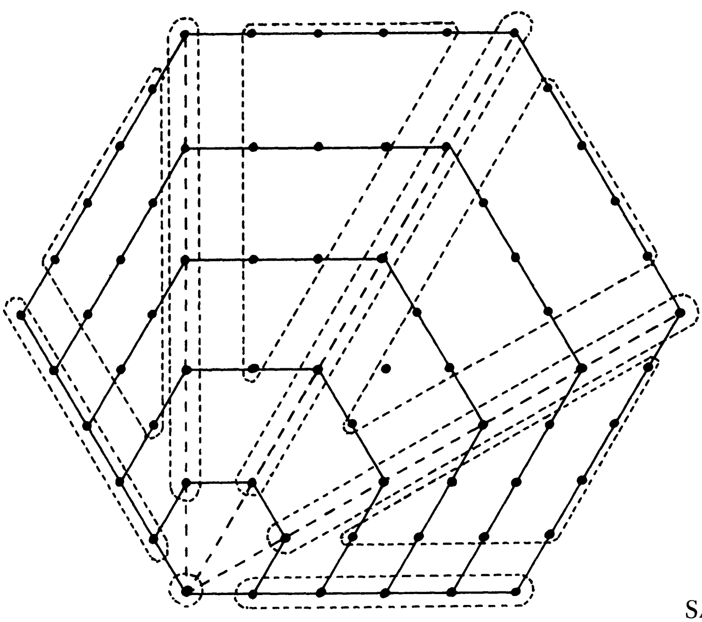
In summary then, we have learned that the only possible inscribed n-gons with rational perimeter and the only possible circumscribed n-gons with either rational perimeter or rational area are those satisfying $\cos(2\pi/n) \in Q$. Similarly, the only possible inscribed n-gons with rational area must satisfy $\cos(4\pi/n) \in Q$. However, by (7), $2\cos(2\pi/n)$ and $2\cos(4\pi/n)$ are roots of monic polynomials with integer coefficients. So, by the rational root theorem, requiring $\cos(2\pi/n)$ or $\cos(4\pi/n)$ to be rational forces (respectively) $2\cos(2\pi/n)$ or $2\cos(4\pi/n)$ to be an integer. Further, since $-1 \leqslant \cos(\theta) \leqslant 1$, these integer values must be found in the interval $[-2, 2]$. We are thus led to solving the equations $2\cos(2\pi/n) = 0, \pm 1, \pm 2$ and $2\cos(4\pi/n) = 0, \pm 1, \pm 2$; that is, $\cos(2\pi/n) = 0, \pm 1/2, \pm 1$ and $\cos(4\pi/n) = 0, \pm 1/2, \pm 1$. Because these are all familiar values one quickly finds $\cos(2\pi/n) = 0 \Rightarrow n = 4$, $\cos(2\pi/n) = -1/2 \Rightarrow n = 3$, $\cos(2\pi/n) = 1/2 \Rightarrow n = 6$, while $\cos(2\pi/n) = \pm 1$ has no solution since $n > 2$. Similarly, $\cos(4\pi/n) = 0 \Rightarrow n = 8$, $\cos(4\pi/n) = -1/2 \Rightarrow n = 6$, $\cos(4\pi/n) = 1/2 \Rightarrow n = 12$, $\cos(4\pi/n) = -1 \Rightarrow n = 4$, while $\cos(4\pi/n) = 1$ has no solution since $n > 2$. It follows that $n \in \{3, 4, 6\}$ is a necessary condition for any of $P_I(n)$, $A_c(n)$, or $P_c(n)$ to lie in Q and that $n \in \{4, 6, 8, 12\}$ is a necessary condition for $A_I(n) \in Q$. We can now obtain the same answer as we did previously by constructing a table like that at the end of the first approach. Note that this time the table will be smaller than before since this second approach is not only more elementary, but is actually more efficient in that the "troublesome" values $n=5$ and $n=10$ and the value $n=3$ have been eliminated from the necessary condition for $A_I(n) \in Q$.

REFERENCES

1. B. H. Arnold and H. Eves, A simple proof that, for odd $p>1$, $\arccos(1/p)$ and π are incommensurable. *Amer. Math. Monthly* 56 (1949), 20–21.
2. D. M. Burton, *Elementary Number Theory*, Allyn & Bacon, Needham Hts, MA, 1980.
3. J. B. Fraleigh, *A First Course in Abstract Algebra*, Addison-Wesley, Reading, MA, 1976.
4. H. Hadwiger, H. Debrunner, and V. Klee, *Combinatorial Geometry in the Plane*, Holt, Rinehart & Winston, New York, 1964.
5. R. W. Hamming, The transcendental character of $\cos(x)$, *Amer. Math. Monthly* 52 (1945), 336–337.
6. D. H. Lehmer, A note on trigonometric algebraic numbers, *Amer. Math. Monthly* 40 (1933), 165–166.
7. I. Niven, *Irrational Numbers*, Carus Math. Mono. 11, MAA, 1956, 28–41.
8. J. M. H. Olmsted, Rational values of trigonometric functions, *Amer. Math. Monthly*, 52 (1945), 507–508.
9. G. Pólya and G. Szegő, *Problems and Theorems in Analysis II*, Springer-Verlag, New York, 1976.
10. R. S. Underwood, On the irrationality of certain trigonometric functions, *Amer. Math. Monthly* 28 (1921), 374–376.
11. E. A. Walker, *Introduction to Abstract Algebra*, Random House, New York, 1987.

Proof without Words:

The kth n-gonal number is $1+(k-1)(n-1)+\frac{1}{2}(k-2)(k-1)(n-2)$.



—Dave Logothetti
Santa Clara University
Santa Clara, CA 95053

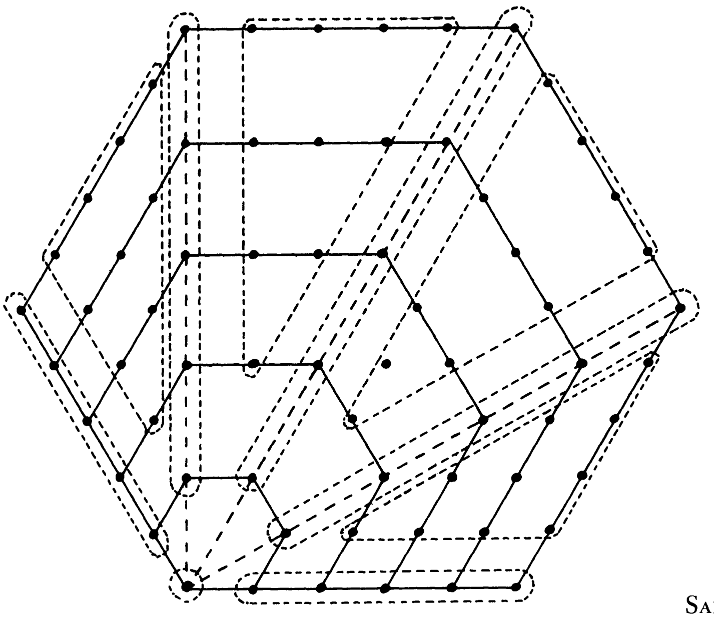
more elementary, but is actually more efficient in that the "troublesome" values $n = 5$ and $n = 10$ and the value $n = 3$ have been eliminated from the necessary condition for $A_I(n) \in Q$.

REFERENCES

1. B. H. Arnold and H. Eves, A simple proof that, for odd $p > 1$, $\arccos(1/p)$ and π are incommensurable. *Amer. Math. Monthly* 56 (1949), 20–21.
2. D. M. Burton, *Elementary Number Theory*, Allyn & Bacon, Needham Hts, MA, 1980.
3. J. B. Fraleigh, *A First Course in Abstract Algebra*, Addison-Wesley, Reading, MA, 1976.
4. H. Hadwiger, H. Debrunner, and V. Klee, *Combinatorial Geometry in the Plane*, Holt, Rinehart & Winston, New York, 1964.
5. R. W. Hamming, The transcendental character of $\cos(x)$, *Amer. Math. Monthly* 52 (1945), 336–337.
6. D. H. Lehmer, A note on trigonometric algebraic numbers, *Amer. Math. Monthly* 40 (1933), 165–166.
7. I. Niven, *Irrational Numbers*, Carus Math. Mono. 11, MAA, 1956, 28–41.
8. J. M. H. Olmsted, Rational values of trigonometric functions, *Amer. Math. Monthly*, 52 (1945), 507–508.
9. G. Pólya and G. Szegő, *Problems and Theorems in Analysis II*, Springer-Verlag, New York, 1976.
10. R. S. Underwood, On the irrationality of certain trigonometric functions, *Amer. Math. Monthly* 28 (1921), 374–376.
11. E. A. Walker, *Introduction to Abstract Algebra*, Random House, New York, 1987.

Proof without Words:

The kth n-gonal number is $1 + (k-1)(n-1) + \frac{1}{2}(k-2)(k-1)(n-2)$.



—DAVE LOGOTHETTI
SANTA CLARA UNIVERSITY
SANTA CLARA, CA 95053

Some Counting Problems Involving the Multinomial Expansion

FRED DODD
University of South Alabama
Mobile, AL 36688

RHODES PEELE
Auburn University at Montgomery
Montgomery, AL 36117

One of the loveliest and most surprising results of elementary number theory is Kummer's carry theorem. In the generalized form in which we will apply it, it says that for any nonnegative integers $e_1, e_2, \dots, e_t$ and any prime p, the exponent of the highest power of p that divides the multinomial coefficient $(e_1 + e_2 + \cdots + e_t)!/(e_1!e_2! \cdots e_t!)$ equals the *number of carries* that occur when the addition $e_1 + e_2 + \cdots + e_t$ is performed in the p-ary number system.

The research literature on Kummer's result and its various generalizations, such as the one in the previous sentence; analogs, such as in [4]; and applications, such as in [7] is quite large as can be seen from the bibliography of [12]. Our purpose here is less to prove new results than to develop some of the existing theory at an elementary level. (For a more advanced treatment, see [1] and [6]).

A good example of the sort of question we shall be concerned with is provided by *Mathematics Magazine* Problem 1255 [11]. We consider the multinomial expansion

$$(x_1 + x_2 + \cdots + x_t)^n = \sum \frac{n!}{e_1!e_2! \cdots e_t!} x_1^{e_1} x_2^{e_2} \cdots x_t^{e_t} \tag{1}$$

in which the summation is over all nonnegative integer vectors $(e_1, e_2, \dots, e_t)$ such that $e_1 + e_2 + \cdots + e_t = n$. The problem is to show that the number of terms on the right side of (1) whose coefficients are odd is t^d, where d is the number of ones in the binary representation of n. In fact, this problem can be thought of as a prototype in that by replacing "oddness" by other divisibility properties we can pose other interesting problems. In many cases Kummer's theorem can help us solve these problems by serving as a useful bridge from number theory to elementary combinatorics.

1. Preliminaries The reader will likely be familiar with most of the material in this section, but we thought it best to err on the side of caution.

Throughout, n and t are fixed positive integers with $t > 1$, p is a fixed prime; $e_1, e_2, \dots, e_t$ are nonnegative integers that sum to n, and $\binom{n}{e_1 e_2 \cdots e_t}$ denotes the multinomial coefficient $n!/(e_1!e_2! \cdots e_t!)$. The integer-valued function ord_p is defined for all positive integers m by $\mathrm{ord}_p(m) = \alpha$, where $p^\alpha | m$, $p^{\alpha+1} \nmid m$. We let $N(n,t,p^\alpha)$ be the number of multinomial coefficients C in (1) with $\mathrm{ord}_p(C) = \alpha$.* The

*There is a slight abuse of notation here in that $N(n,t,p^\alpha)$ may differ from $N(n,t,q^\alpha)$ when $\alpha = 0$, although p^0 and q^0 are numerically equal.

functions $\varepsilon(n,t,p)$ and $M(n,t,p)$ to be used in Sections 3, 5, and 6 are defined as follows: $\varepsilon(n,t,p) = \max\{\mathrm{ord}_p(C)\text{: } C \text{ is a multinomial coefficient in (1)}\}$; and $M(n,t,p) = N(n,t,p^{\varepsilon})$ where $\varepsilon = \varepsilon(n,t,p)$.

The p-ary representation of the nonnegative integer m is given by $m = \sum_{i=0}^{\infty} m_i p^i$, where the m_i are nonnegative integers less than p. If $m_i = 0$ for all $i > s$ then we denote the p-ary representation of m by $m = (m_s\ m_{s-1} \cdots m_0)_p$. It is sometimes convenient to allow leading zeros in this representation. For example, $(0\ 0\ 1\ 2\ 0)_3$ and $(1\ 2\ 0)_3$ are 3-ary representations of 15.

For $j = 1, 2, \ldots, t$, let $\sum_{i=0}^{\infty} e_i^{(j)} p^i$ be the p-ary representation of e_j. Then for each $i \geqslant 0$ there exist unique nonnegative integers c_i, n_i with $n_i < p$ such that

$$\begin{aligned} e_0^{(1)} + e_0^{(2)} + \cdots + e_0^{(t)} &= c_0 p + n_0, \text{ and} \\ e_i^{(1)} + e_i^{(2)} + \cdots + e_i^{(t)} + c_{i-1} &= c_i p + n_i \text{ for } i \geqslant 1. \end{aligned} \tag{2}$$

It is easy to check that $\sum_{i=0}^{\infty} n_i p^i$ is the p-ary representation of $n = e_1 + e_2 + \cdots + e_t$. We call c_i the carry in column i of the p-ary addition of $e_1, e_2, \ldots, e_t$; if $c_i > 0$ we call column i a carry column. Finally, $\sum_{i=0}^{\infty} c_i$ is the number of carries of the p-ary addition. Here's an example:

Example 1. For $e_1 = (2\ 2\ 0)_3$, $e_2 = (1\ 2\ 2\ 2)_3$, $e_3 = (2\ 2\ 2)_3$, $e_4 = (1\ 1\ 2\ 1)_3$, $e_5 = (2\ 2\ 1\ 2)_3$ we have $e_1 + e_2 + e_3 + e_4 + e_5 = (2\ 2\ 0\ 2\ 1)_3$, $c_0 = 2$, $c_1 = 3$, $c_2 = 4$, $c_3 = 2$ and $c_i = 0$ for $i \geqslant 4$. There are 11 carries in this 3-ary sum. The computations are best shown in the following format:

$$\begin{array}{r l} 2\,4\,3\,2 & \leftarrow \text{carries} \\ \hline 2\,2\,0 & \\ 1\,2\,2\,2 & \\ 2\,2\,2 & \\ 1\,1\,2\,1 & \\ 2\,2\,1\,2 & \\ \hline 2\,2\,0\,2\,1 & \end{array}$$

We close this section with several combinatorial items. For r any real number and k any nonnegative integer, we define the generalized binomial coefficient $\binom{r}{k}$ to be 1 if $k = 0$, and to equal

$$\frac{r(r-1)(r-2)\cdots(r-k+1)}{k!}$$

otherwise. We then have the reciprocity property

$$\binom{-r}{k} = (-1)^k \binom{r+k-1}{k}.$$

The generalized binomial theorem asserts that, for all y such that $|y| < 1$, $\sum_{k=0}^{\infty} \binom{r}{k} y^k$ converges to $(1+y)^r$. We will frequently use the well-known fact that the number of solutions in nonnegative integers to the equation $e_1 + e_2 + \cdots + e_t = n$ is $\binom{n+t-1}{t-1}$. Finally, if r and k are integers such that $r \geqslant k \geqslant 0$, then we have the symmetry property

$$\binom{r}{k} = \binom{r}{r-k} = \binom{r}{k\ \ (r-k)} = \frac{r!}{k!(r-k)!}.$$

2. Kummer's theorem for multinomial coefficients In 1852 Kummer showed that $\operatorname{ord}_p\binom{n}{k}$ is equal to the number of carries in the p-ary addition of k and $(n-k)$ ([2], [8]). In Theorem 1 we generalize this result to multinomial coefficients. Our proof parallels the proof given by Kummer and is based on the following lemma by Legendre ([9], [2]).

LEMMA. *If $\sum_{i=0}^{\infty} m_i p^i$ is the p-ary representation of the nonnegative integer m, then* $\operatorname{ord}_p(m!) = (m - \sum_{i=0}^{\infty} m_i)/(p-1)$.

Proof. For some s we have $m_i = 0$ for all $i > s$, so that $m = \sum_{i=0}^{s} m_i p^i$. From elementary number theory [10], $\operatorname{ord}_p(m!) = \sum_{j=1}^{s}[m/p^j]$ where [ ] is the greatest integer function. Since

$$\left[\frac{m}{p^j}\right] = \sum_{i=j}^{s} m_i p^{i-j}$$

for $j = 1, 2, \ldots, s$, we have

$$\begin{aligned}\operatorname{ord}_p(m!) &= \sum_{j=1}^{s}\sum_{i=j}^{s} m_i p^{i-j} = \sum_{i=1}^{s} m_i \sum_{j=1}^{i} p^{i-j} \\ &= \sum_{i=1}^{s} m_i\left(\frac{p^i-1}{p-1}\right) = \frac{1}{p-1}\sum_{i=0}^{s} m_i(p^i-1) \\ &= \frac{1}{p-1}\left(m - \sum_{i=0}^{s} m_i\right).\end{aligned}$$

THEOREM 1. *If γ is the number of carries in the p-ary addition of $e_1, e_2, \ldots, e_t$, then*

$$\operatorname{ord}_p\binom{n}{e_1\, e_2 \cdots e_t} = \gamma.$$

Proof. Set $\alpha = \operatorname{ord}_p\binom{n}{e_1\, e_2 \cdots e_t}$. Using the notation of Section 1, the equations given by (2), and the lemma, we obtain

$$\begin{aligned}(p-1)\alpha &= (p-1)\left(\operatorname{ord}_p(n!) - \sum_{j=1}^{t}\operatorname{ord}_p(e_j!)\right) \\ &= \left(n - \sum_{i=0}^{\infty} n_i\right) - \sum_{j=1}^{t}\left(e_j - \sum_{i=0}^{\infty} e_i^{(j)}\right) \\ &= \sum_{i=0}^{\infty}\left(\sum_{j=1}^{t} e_i^{(j)} - n_i\right) = c_0 p + \sum_{i=1}^{\infty}(c_i p - c_{i-1}) \\ &= (p-1)\sum_{i=0}^{\infty} c_i\end{aligned}$$

so that $\alpha = \sum_{i=0}^{\infty} c_i = \gamma$.

Example 2. The 3-ary representations of $e_1 = 24$, $e_2 = 53$, $e_3 = 26$, $e_4 = 43$, $e_5 = 77$ are given in Example 1. Thus $\operatorname{ord}_3\binom{223}{24\ 53\ 26\ 43\ 77} = 11$.

3. The number of multinomial coefficients not divisible by p Our first application of Theorem 1 is a generalization of the problem by Ruderman mentioned in the introduction.

THEOREM 2. *Let $n=(n_s\ n_{s-1}\ \cdots\ n_0)_p$. Then the number of multinomial coefficients in* (1) *that are not divisible by p is given by*

$$N(n,t,p^0)=\prod_{k=0}^{s}\binom{n_k+t-1}{t-1}. \tag{3}$$

Proof. The multinomial coefficient $n!/(e_1!e_2!\cdots e_t!)$ is not divisible by p if and only if there are no carries when we perform the p-ary summation of $e_1,e_2,\dots,e_t$. For $j=1,2,\dots,t$ let $e_j=(e_s^{(j)}\ e_{s-1}^{(j)}\ \cdots\ e_0^{(j)})_p$. Then there are no carries if and only if $e_k^{(1)}+e_k^{(2)}+\cdots+e_k^{(t)}=n_k$ for $k=0,1,\dots,s$. For fixed k there are $\binom{n_k+t-1}{t-1}$ solutions to the preceding equation. Formula (3) now follows from the multiplication principle for counting.

The following five statements are all consequences of Theorem 2. The verifications are left as exercises for the reader. In (4) and (5) we assume that $n_s>0$. Singmaster's published solution [13] of Problem 1255 establishes (8) by this approach.

$$t\leqslant N(n,t,p^0),\text{ with equality if and only if } n=p^s. \tag{4}$$

$$N(n,t,p^0)\leqslant\binom{p+t-2}{t-1}^{s+1},\text{ with equality if and only if } n=p^{s+1}-1. \tag{5}$$

$$\sum_{\alpha=1}^{\varepsilon(n,t,p)}N(n,t,p^\alpha)=\binom{n+t-1}{t-1}-\prod_{k=0}^{s}\binom{n_k+t-1}{t-1}. \tag{6}$$

$$N(n,2,p^0)=\prod_{k=0}^{s}(n_k+1). \tag{7}$$

$N(n,t,2^0)=t^d$, where d is the number of ones in the binary representation of n.. (8)

4. The number of singly even multinomial coefficients From statements (6) and (8), we see that there are $\binom{n+t-1}{t-1}-t^d$ even coefficients C in (1). It is natural to ask how many of these are singly even—i.e., satisfy $C\equiv 2 \pmod 4$. The next theorem gives the answer.

THEOREM 3. *Let $n=(n_s\ n_{s-1}\ \cdots\ n_0)_2$ and $n_s=1$. Suppose $(n_s\ n_{s-1}\ \cdots\ n_0)_2$ has c blocks of consecutive zeros, d ones, and e blocks of consecutive ones (a block may consist of one digit). Then the number of singly even multinomial coefficients in* (1) *is given by*

$$N(n,t,2^1)=c\binom{t}{2}t^{d-1}+(d-e)\binom{t}{3}t^{d-2}. \tag{9}$$

Example 3. For $t=4$ and $n=19$ we have $n=(1\ 0\ 0\ 1\ 1)_2$, so that $c=1$, $d=3$, $e=2$. Therefore, $N(19,4,2^1)=112$. Note that we get the same answer if we change n to 11, 25 or 13 (and in fact infinitely many other values) since the parameters c,d,e are the same.

Proof. The multinomial coefficient $n!/(e_1!e_2!\cdots e_t!)$ is singly even if and only if there is exactly one carry in the binary summation of $e_1, e_2, \ldots, e_t$. Let

$$\begin{array}{ccccccl} 0 & \cdots & 1 & 0 & \cdots & 0 & \leftarrow \text{carries} \\ \hdashline e_s^{(1)} & \cdots & e_{i+1}^{(1)} & e_i^{(1)} & \cdots & e_0^{(1)} & \\ e_s^{(2)} & \cdots & e_{i+1}^{(2)} & e_i^{(2)} & \cdots & e_0^{(2)} & \\ \vdots & & & & & \vdots & \\ e_s^{(t)} & \cdots & e_{i+1}^{(t)} & e_i^{(t)} & \cdots & e_0^{(t)} & \\ \hline n_s & \cdots & n_{i+1} & n_i & \cdots & n_0 & \end{array} \tag{10}$$

be such a one-carry binary summation, and suppose for the remainder of the proof that i is the carry column, so that $c_i = 1$ in the notation of Section 1.

Suppose first that $n_i = 0$. Then $n_{i+1} = 1$ since 1 is carried to column $i+1$, and column $i+1$ is not a carry column. Therefore n_i must be the left bit of a consecutive block of zeros. Thus there are only c ways to choose column i. Note that column i must contain exactly two ones since only column i is a carry column. We can choose these two ones in $\binom{t}{2}$ ways. All columns $j \neq i+1$ for which $n_j = 1$ have exactly one 1 and we can choose these 1's in t^{d-1} ways. Column $i+1$ is a column of zeros, as are all columns $j \neq i$ for which $n_j = 0$. We conclude by the multiplication principle for counting that there are $c\binom{t}{2}t^{d-1}$ one-carry binary summations (10) for which $n_i = 0$.

Next suppose that $n_i = 1$. As before, $n_{i+1} = 1$ and therefore n_i cannot be the left bit of a consecutive block of ones. Column i must contain three ones since only column i is a carry column. By an argument similar to the previous one, we conclude that there are $(d-e)\binom{t}{3}t^{d-2}$ one-carry binary summations (10) for which $n_i = 1$.

Any one-carry binary summation (10) obviously falls into exactly one of the two classes that we have enumerated ($n_i = 0$ or $n_i = 1$), so the theorem now follows from the addition principle for counting.

Note that by choosing $t = 2$ we obtain $N(n, 2, 2^1) = c \cdot 2^{d-1}$, a pendant for the more familiar formula $N(n, 2, 2^0) = 2^d$ deduced from (8). (See also [**5**].)

5. The number of binomial coefficients C for which $\mathrm{ord}_p(C)$ is maximum In this section we will evaluate $\varepsilon(n, t, p)$ and $M(n, t, p)$ for the special case $t = 2$. The following lemma is helpful.

LEMMA. *For nonnegative integers m, define $N(m)$ to be the number of integer solutions (x, y) to $x + y = m$ with $0 \leqslant x < p$, $0 \leqslant y < p$. Then*

$$N(m) = \begin{cases} m+1 & \text{if } 0 \leqslant m < p \\ 2p - m - 1 & \text{if } p \leqslant m \leqslant 2(p-1) \\ 0 & \text{if } 2(p-1) < m \end{cases}. \tag{11}$$

Proof. For $0 \leqslant m < p$ we see that $x = 0, 1, 2, \ldots, m$ yields the solutions so that $N(m) = m + 1$. For $p \leqslant m \leqslant 2(p-1)$ we see that $x = m - p + 1, m - p + 2, \ldots, p - 1$ yields the solutions so that $N(m) = 2p - m - 1$. It is clear that there are no solutions when $2(p-1) < m$.

THEOREM 4. *Let* $n=(n_s\ n_{s-1}\ \cdots\ n_0)_p$, $n_s>0$, *be the p-ary representation of n. If there is a subscript* $i\leqslant s$ *for which* $n_i<p-1$, *let r be the smallest such subscript; otherwise, let* $r=s$. *Then*

$$\varepsilon(n,2,p)=s-r; \tag{12}$$

$$M(n,2,p)=\begin{cases} n_s\cdot p^r\cdot(p-n_r-1)\cdot\prod_{i=r+1}^{s-1}(p-n_i) & \text{if } s>r+1;\\ n_s\cdot p^r\cdot(p-n_r-1) & \text{if } s=r+1;\\ n+1 & \text{if } s=r. \end{cases} \tag{13}$$

Proof. We will prove the theorem only for the case $s>r+1\geqslant 2$ as only minor modifications are needed for the remaining cases. When we add e_1 and e_2 in p-ary form to obtain n, the carry c_i in column i is either 0 or 1. Since $n_0=n_1=\cdots=n_{r-1}=p-1$, we cannot have a positive carry in columns $0,1,2,\ldots,r-1$. For otherwise, there would be a smallest subscript $i\leqslant r-1$ such that $c_i=1$. But then $e_i^{(1)}+e_i^{(2)}=p+n_i=2p-1$, which is impossible since $0\leqslant e_i^{(j)}<p$ for $j=1,2$. In addition, we cannot have a positive carry in column s. Hence, at most we can have a carry of 1 in each of columns r, $r+1,\ldots,s-1$; i.e., at most $s-r$ carries. Furthermore, we will have $s-r$ carries if, and only if,

$$\begin{cases} e_0^{(1)}+e_0^{(2)}=p-1\\ e_1^{(1)}+e_1^{(2)}=p-1\\ \quad\vdots \qquad\quad \vdots\\ e_{r-1}^{(1)}+e_{r-1}^{(2)}=p-1\\ e_r^{(1)}+e_r^{(2)}=p+n_r\\ e_{r+1}^{(1)}+e_{r+1}^{(2)}+1=p+n_{r+1}\\ \quad\vdots \qquad\quad \vdots\\ e_{s-1}^{(1)}+e_{s-1}^{(2)}+1=p+n_{s-1}\\ e_s^{(1)}+e_s^{(2)}+1=n_s \end{cases} \tag{14}$$

It follows from the lemma that each of the $s+1$ equations in (14) is solvable in nonnegative integers $e_i^{(1)}, e_i^{(2)}$ less than p and, with the aid of the multiplication principle for counting, that $M(n,2,p)$ has the value given in the theorem.

For integral $\varepsilon>0$ it is clear from Theorem 4 that $n=p^\varepsilon$ is the smallest n such that $\varepsilon(n,2,p)=\varepsilon$, and that $M(p^\varepsilon,2,p)=(p-1)p^{\varepsilon-1}$ so that exactly $(p-1)p^{\varepsilon-1}$ of the binomial coefficients $\binom{p^\varepsilon}{k}$, $k=0,1,\ldots,p^\varepsilon$ have $\operatorname{ord}_p\binom{p^\varepsilon}{k}=\varepsilon$. In another direction, the reader can use Theorem 4 to argue that, given integral $\varepsilon>0$, $n=2(p^\varepsilon-1)$ is the smallest n with $M(n,2,p)=1$ and $\varepsilon(n,2,p)=\varepsilon$. (The one binomial coefficient of course being $\binom{2(p^\varepsilon-1)}{p^\varepsilon-1}$.) Finally, in the spirit of the results $N(n,2,2^0)=2^d$ and $N(n,2,2^1)=c\cdot 2^{d-1}$, we deduce from Theorem 4 that for $s>r$, $M(n,2,2)=2^{r+b-1}$ where b is the number of zeros in the binary representation $(n_s\ n_{s-1}\ \cdots\ n_0)_2$, $n_s=1$, of n. (Of course, when $s=r$ then $n=2^{r+1}-1$ so that $M(n,2,2)=2^{r+1}$.)

6. The number of multinomial coefficients C for which $\operatorname{ord}_p(C)$ is maximum, given that $n\leqslant(p-1)t$ In this section we will evaluate $\varepsilon(n,t,p)$ and $M(n,t,p)$ when $n\leqslant(p-1)t$. We could base the proof on Kummer's theorem, but in this instance a more direct line of reasoning is possible, and (we believe) more transparent.

THEOREM 5. *If $n \leqslant (p-1)t$, then*

$$\varepsilon(n,t,p) = \operatorname{ord}_p(n!); \tag{15}$$

$M(n,t,p)$ is the coefficient of y^n in the expansion of $(1+y+y^2+\cdots+y^{p-1})^t$; and (16)

$$M(n,t,p) = \sum_{i=0}^{[n/p]} (-1)^i \binom{t}{i}\binom{t+n-1-ip}{t-1}. \tag{17}$$

Proof. The assumption that $(p-1)t \geqslant n$ implies that there exist nonnegative integers $e_1, e_2, \ldots, e_t$, each less than p, such that $e_1+e_2+\cdots+e_t=n$. It is clear that by (and only by) choosing such integers we maximize $\operatorname{ord}_p\binom{n}{e_1\, e_2 \cdots e_t}$ since p will not then divide the denominator of $n!/(e_1!e_2!\cdots e_t!)$, and therefore $\operatorname{ord}_p\binom{n}{e_1\, e_2 \cdots e_t} = \operatorname{ord}_p(n!)$ as claimed in (15).

Now we determine the number of multinomial coefficients C for which $\operatorname{ord}_p(C)$ equals this maximum value. Consider the ordinary generating function

$$g(y) = \underbrace{\begin{matrix}(1+y+y^2+\cdots+y^{p-1})(1+y+y^2+\cdots+y^{p-1})\\ \cdots(1+y+y^2+\cdots+y^{p-1})\end{matrix}}_{t\text{ factors}}. \tag{18}$$

The coefficient of y^n in the polynomial expansion of (18) is the number of ways that we can choose a term y^{e_1} from the first factor, a term y^{e_2} from the second factor,..., a term y^{e_t} from the last factor, so that $y^{e_1}y^{e_2}\cdots y^{e_t} = y^{e_1+e_2+\cdots+e_t} = y^n$. Thus the coefficient of y^n equals the number of integer solutions to $e_1+e_2+\cdots+e_t=n$ such that $0 \leqslant e_i < p$ for $i=1,2,\ldots,t$, and this number in turn equals $M(n,p,t)$ by the first paragraph of the proof.

Formula (17) is now a consequence of the following computation, where we assume that $|y|<1$ to guarantee convergence.

$$\begin{aligned} g(y) &= (1+y+y^2+\cdots+y^{p-1})^t = \left(\frac{1-y^p}{1-y}\right)^t = (1-y^p)^t(1-y)^{-t} \\ &= \sum_{j=0}^{t}\binom{t}{j}(-y^p)^j \cdot \sum_{k=0}^{\infty}\binom{-t}{k}(-y)^k \\ &= \sum_{j=0}^{t}(-1)^j\binom{t}{j}y^{pj} \cdot \sum_{k=0}^{\infty}\binom{t+k-1}{t-1}y^k. \end{aligned}$$

Collecting the coefficients of y^n in the last double summation we deduce (17).

7. Concluding remarks We have shown Kummer's theorem to be an important tool in the solution of divisibility problems concerning multinomial coefficients. Since the theorem and its applications are easily accessible for introductory courses in number theory and combinatorics, it is rather surprising that it is not a better-known result. We hope this article will encourage its inclusion in the aforementioned courses.

We conclude with a recommendation that the computer-minded reader employ Kummer's theorem to construct a program to compute $\operatorname{ord}_p\binom{n}{e_1\, e_2 \cdots e_t}$. The program used in our investigations was based on Kummer's theorem for binomial coefficients and the result

$$\binom{e_1+e_2+\cdots+e_t}{e_1\,e_2\cdots e_t}=\prod_{i=1}^{t-1}\binom{e_1+e_2+\cdots+e_{i+1}}{e_{i+1}}.$$

By melding this program with one that generates all solutions in nonnegative integers of $e_1+e_2+\cdots+e_t=n$, the reader can verify, as we did, the results of our article for specific values of n, t, and p.

REFERENCES

1. L. Carlitz, The number of binomial coefficients divisible by a fixed power of a prime, *Rend. Circ. Mat. Palermo* 16 (1967), 299–320.
2. L. E. Dickson, *History of the Theory of Numbers*, Carnegie Institution of Washington; reprinted Chelsea Publishing Co., New York, 1966, Vol. 1, pp. 263–278.
3. ______, Theorems on the residues of multinomial coefficients with respect to a prime modulus, *Quart. J. Math.* 33 (1902), 378–384.
4. R. Fray, Congruence properties of ordinary and q-binomial coefficients, *Duke Math J.* 34 (1967), 467–480.
5. Ross Honsberger, *Mathematical Gems II*, MAA, 1976, pp. 1–9.
6. F. T. Howard, The number of multinomial coefficients divisible by a fixed power of a prime, *Pacific J. Math.* 50 (1974), 99–108.
7. R. E. Jamison, Dimensions of hyperplane spaces over finite fields, *Math. Z.* 162 (1978), 101–111.
8. E. E. Kummer, Über die ergänzungssätze zu den allgemeinen reciprocitätsgesetzen, *J. für Math.* 44 (1852), 115–116.
9. A. M. Legendre, *Théorie des Nombres*, 3rd edition, Paris, 1830, Vol. 1, p. 10.
10. I. Niven and H. S. Zuckerman, *An Introduction to the Theory of Numbers*, 4th edition, John Wiley and Sons, New York, 1980, pp. 100–103.
11. H. D. Ruderman, Problem 1255, this MAGAZINE 59 (1986), 297.
12. David Singmaster, Divisibility of binomial and multinomial coefficients by primes and prime powers, *18th Anniversary Volume of the Fibonacci Association*, 1980, pp. 98–113.
13. ______, Solution III of Problem 1255, this MAGAZINE 61 (1988), 53–54.

Volumes of Cones, Paraboloids, and Other "Vertex Solids"

PAUL B. MASSELL
United States Naval Academy
Annapolis, MD 21402

While performing some calculations involving the volume of a solid circular paraboloid $z(r)=h(1-(r/a)^2)$ (with $h>0$) as illustrations of the Divergence Theorem in vector calculus, the author noticed that the ratio of the volume of the portion of the solid paraboloid above the polar plane to that of the solid cylinder with the same base and height h (its associated solid cylinder) is equal to $1/2$ for all values of the radius a. A natural question is whether this ratio holds for elliptical paraboloids or for

$$\binom{e_1+e_2+\cdots+e_t}{e_1\, e_2 \cdots e_t} = \prod_{i=1}^{t-1} \binom{e_1+e_2+\cdots+e_{i+1}}{e_{i+1}}.$$

By melding this program with one that generates all solutions in nonnegative integers of $e_1+e_2+\cdots+e_t=n$, the reader can verify, as we did, the results of our article for specific values of n, t, and p.

REFERENCES

1. L. Carlitz, The number of binomial coefficients divisible by a fixed power of a prime, *Rend. Circ. Mat. Palermo* 16 (1967), 299–320.
2. L. E. Dickson, *History of the Theory of Numbers*, Carnegie Institution of Washington; reprinted Chelsea Publishing Co., New York, 1966, Vol. 1, pp. 263–278.
3. ______, Theorems on the residues of multinomial coefficients with respect to a prime modulus, *Quart. J. Math.* 33 (1902), 378–384.
4. R. Fray, Congruence properties of ordinary and q-binomial coefficients, *Duke Math J.* 34 (1967), 467–480.
5. Ross Honsberger, *Mathematical Gems II*, MAA, 1976, pp. 1–9.
6. F. T. Howard, The number of multinomial coefficients divisible by a fixed power of a prime, *Pacific J. Math.* 50 (1974), 99–108.
7. R. E. Jamison, Dimensions of hyperplane spaces over finite fields, *Math. Z.* 162 (1978), 101–111.
8. E. E. Kummer, Über die ergänzungssätze zu den allgemeinen reciprocitätsgesetzen, *J. für Math.* 44 (1852), 115–116.
9. A. M. Legendre, *Théorie des Nombres*, 3rd edition, Paris, 1830, Vol. 1, p. 10.
10. I. Niven and H. S. Zuckerman, *An Introduction to the Theory of Numbers*, 4th edition, John Wiley and Sons, New York, 1980, pp. 100–103.
11. H. D. Ruderman, Problem 1255, this MAGAZINE 59 (1986), 297.
12. David Singmaster, Divisibility of binomial and multinomial coefficients by primes and prime powers, *18th Anniversary Volume of the Fibonacci Association*, 1980, pp. 98–113.
13. ______, Solution III of Problem 1255, this MAGAZINE 61 (1988), 53–54.

Volumes of Cones, Paraboloids, and Other "Vertex Solids"

PAUL B. MASSELL
United States Naval Academy
Annapolis, MD 21402

While performing some calculations involving the volume of a solid circular paraboloid $z(r)=h(1-(r/a)^2)$ (with $h>0$) as illustrations of the Divergence Theorem in vector calculus, the author noticed that the ratio of the volume of the portion of the solid paraboloid above the polar plane to that of the solid cylinder with the same base and height h (its associated solid cylinder) is equal to $1/2$ for all values of the radius a. A natural question is whether this ratio holds for elliptical paraboloids or for paraboloids with any simple curve as a base. Another question is whether there is a similar ratio that is independent of the shape and size of the base for exponents other than 2 in the formula for $z(r)$. Our theorem answers these questions for a class of solids we call vertex solids; a class that includes cones and paraboloids.

We will now define vertex solids. Let $r = g(\theta)$ describe a simple closed curve in the polar plane such that $0 \leqslant g(\theta)$ for $0 \leqslant \theta \leqslant 2\pi$. Let the point V be on the positive z-axis with $z = h$ (this will be the top vertex for the vertex solid). For each fixed θ in $[0, 2\pi]$, consider the curves $z_k(r) = h(1 - (r/g(\theta))^k)$ where k is a positive constant (if $g(\theta) = 0$, let $z_k(r) = h$). If k is an integer, then $z_k(r)$ is clearly the unique curve of kth degree that goes through V and the point $(\theta, g(\theta), 0)$ in the polar plane and that has the property $d^i z/dr^i = 0$ at V for $i = 1, 2, \dots, k-1$ (for fixed θ). Now consider r and θ as independent variables and view the above expression for $z_k = z_k(r, \theta)$ as representing a surface. If $r = g(\theta)$ describes an ellipse, then z_1 represents an elliptical cone and z_2 represents an elliptical paraboloid. For all $k > 0$, we call the solid defined by the set of points (r, θ, z) satisfying $0 \leqslant r \leqslant g(\theta)$, $0 \leqslant \theta \leqslant 2\pi$, $0 \leqslant z \leqslant z_k(r, \theta)$ a *vertex solid*. Its *associated solid cylinder* is the set of points (r, θ, z) satisfying $0 \leqslant r \leqslant g(\theta)$, $0 \leqslant \theta \leqslant 2\pi$, $0 \leqslant z \leqslant h$.

THEOREM. *The ratio of the volume of the vertex solid of degree k to that of its associated solid cylinder is $k/(k+2)$. (Thus for an elliptical cone the volume is $1/3$ that of its associated cylinder's volume πabh; for an elliptical paraboloid its volume is $1/2$ of πabh. Here, a and b are the minor and major radii of the ellipse. The ratios $1/3$ and $1/2$ hold for cones and paraboloids (respectively) with any base that is describable by a simple closed curve.)*

Proof.

$$\text{vol(vertex solid)} = \int_0^{2\pi} \int_0^{g(\theta)} z_k(r, \theta) \cdot r \, dr \, d\theta$$

$$\text{vol(solid cylinder)} = \int_0^{2\pi} \int_0^{g(\theta)} h \cdot r \, dr \, d\theta = A.$$

Substitution of the expression for z_k and a fairly simple integration reveals that the ratio of the volume of the vertex solid to that of its associated solid cylinder is

$$\frac{A - (2/(k+2))A}{A} = \frac{k}{k+2}.$$

Notes:

(1) This result easily can be extended to the case where the base of the vertex solid does not lie below the vertex V. In this case, the vertex solid is not entirely contained in its solid cylinder.

(2) As k increases, the vertex solid occupies more and more of its associated solid cylinder, and in the limit occupies all of it.

(3) Consider the cross sections of the vertex solid and its solid cylinder generated by the plane $\theta = c$ (constant). (Assume $g(c) > 0$.) Denoting them by C_v and C_c, it's easy to see that $\text{area}(C_v)/\text{area}(C_c) = k/(k+1)$.

I would like to thank my colleague Tom Mahar for a very helpful discussion of this result and Bruce Richter for encouraging me to publish it. Thanks are also due to two referees for several helpful suggestions for improving the readability of the paper.

The Complementary Function and the General Solution

P. RAMANKUTTY
University of Auckland
New Zealand

In a first course in differential equations the basic theme consists in obtaining explicitly the complete (i.e., general) solutions of well-known types of ordinary differential equations. Among these types one would certainly like to include the linear nth-order equation with constant coefficients not only because of its simplicity but also because of its common occurrence in engineering and other sciences. However, in asserting that the so called *complementary function* (CF) is indeed the *general solution* of the homogeneous equation, some textbooks either appear to pay insufficient attention to the significance of the statement or tend to confuse the issue and often simply point to the fact that the CF contains n arbitrary constants [see, e.g., **5**, **9**, **10**]. Some authors who recognize this as inadequate even in a first course include in their books a statement of the basic uniqueness theorem for the associated initial value problem with or without a proof in a later chapter or an apology that the proof is beyond the scope of a first course [**1**, **2**, **3**, **4**, **6**, **7**, **8**]. Either way, the situation is not very satisfactory.

One purpose of this note is to indicate a simple and elementary approach by which the problem can be overcome rather effortlessly. This is done in Part I, which is presented without skipping any of the details of calculation in order to show that it is not only complete and rigorous but is even within the reach of students who have no familiarity whatever with differential equations. Another matter this note deals with concerns the question of how to obtain a particular integral of $L(y)=f$ where L is a linear differential operator. Here the standard methods, available after determining the CF, are

(1) the *annihilator method of undetermined coefficients*, which succeeds if L has constant coefficients and f is such that there can be found a linear differential operator M with constant coefficients satisfying $M(f)=0$, and

(2) Lagrange's *method of variation of parameters* whose applicability is not limited as in (1).

Part II describes a method somewhat different from these two that applies regardless of whether or not L has constant coefficients and irrespective of the nature of f. More importantly, this method does not depend on the predetermination of the CF. However, the applicability of this method is limited to those equations in which L can be expressed as a product of first order factors raised to appropriate powers.

First I prove a simple basic lemma regarding certain indefinite integrals and then proceed directly to the theorem that gives the general solution of the homogeneous equation with constant coefficients. Only mathematical induction is used in the proofs in addition to the fundamental theorem of algebra that of course is, as always, assumed. Next I indicate an alternative and completely direct approach to the solution of the nonhomogeneous equation by repeated application of the "exponential shift technique." When it applies, this method gives not merely a particular integral but in fact the general solution without any need to rely upon the fundamental uniqueness theorem for its justification.

Part I Constant Coefficients

1. LEMMA. *Let $\lambda \in \mathbb{C}$ and $P: \mathbb{R} \to \mathbb{C}$ be a polynomial. Then $\int e^{\lambda x} P(x)\,dx = e^{\lambda x} Q(x) + C$ where $C \in \mathbb{C}$ and $Q: \mathbb{R} \to \mathbb{C}$ is a polynomial with*

$$\text{degree } Q = \begin{cases} \text{degree } P & \text{if } \lambda \neq 0 \\ 1 + \text{degree } P & \text{if } \lambda = 0. \end{cases}$$

Proof. If $\lambda = 0$, the result is immediate. So suppose $\lambda \neq 0$; let n be a nonnegative integer and assume the induction hypothesis that the result holds for all polynomials of degree n. Let $P(x)$ be a polynomial of degree $n+1$. Then,

$$\int e^{\lambda x} P(x)\,dx = \frac{e^{\lambda x}}{\lambda} P(x) - \int \frac{e^{\lambda x}}{\lambda} P'(x)\,dx$$

$$= \frac{e^{\lambda x}}{\lambda} P(x) - \frac{1}{\lambda}\left(e^{\lambda x} Q_1(x) + C\right),$$

where $Q_1(x)$ is a polynomial with degree $Q_1 = \text{degree } P' = n$. This gives

$$\int e^{\lambda x} P(x)\,dx = e^{\lambda x} Q(x) - \frac{C}{\lambda},$$

where $Q(x) = (1/\lambda)(P(x) - Q_1(x))$ is a polynomial of degree $n+1$; thus the result holds for all polynomials of degree $n+1$. Since the induction hypothesis obviously holds for $n = 0$, the induction is complete.

2. THEOREM. *Let $N \in \mathbb{N}$, $a_0, a_1, \ldots, a_N$ be complex numbers with $a_0 \neq 0$ and $y: \mathbb{R} \to \mathbb{C}$ be N times differentiable. Then $\sum_{k=0}^{N} a_k (d^{N-k} y / dx^{N-k}) = 0$ iff $y(x) = \sum_{k=1}^{n} e^{\lambda_k x} P_k(x)$ where $\lambda_1, \ldots, \lambda_n$ are the distinct roots of the polynomial $P(\lambda) = \sum_{k=0}^{N} a_k \lambda^{N-k}$ with multiplicities $m_1, \ldots, m_n$ and $P_k: \mathbb{R} \to \mathbb{C}$ is a polynomial of degree $m_k - 1$.*

Proof. Let $D = d/dx$ so that $\sum_{k=0}^{N} a_k (d^{N-k} y / dx^{N-k}) = P(D)y$. If $N = 1$, then $P(\lambda) = a_0 \lambda + a_1$ so that $\lambda_1 = -a_1/a_0$ is the root of $P(\lambda)$ with $m_1 = 1$. Now, with I denoting the identity operator, $P(D)y = 0 \Leftrightarrow (D - \lambda_1 I)y = 0 \Leftrightarrow D(e^{-\lambda_1 x} y) = 0 \Leftrightarrow e^{-\lambda_1 x} y = c$ for some $c \in \mathbb{C} \Leftrightarrow y(x) = e^{\lambda_1 x} P_1(x)$ where $P_1(x) = c$ is a polynomial of degree $0 = m_1 - 1$. This proves the assertion of the theorem for $N = 1$. Now let $N \in \mathbb{N}$ with $N \geqslant 2$ and assume for induction that the theorem is valid for differential equations of order $k < N$. Since λ_1 is a root of $P(\lambda)$ with multiplicity m_1, we have $P(\lambda) = Q(\lambda)(\lambda - \lambda_1)$ where $Q(\lambda)$ is a polynomial of degree $N - 1$ whose roots are $\lambda_1, \ldots, \lambda_n$ with multiplicities $m_1 - 1, m_2, \ldots, m_n$. [A complex number z is called a root of $Q(\lambda)$ with multiplicity zero iff $Q(z) \neq 0$. Also, a polynomial of degree -1 is the zero function.] Hence, $P(D)y = 0 \Leftrightarrow Q(D)(D - \lambda_1 I)y = 0 \Leftrightarrow (D - \lambda_1 I)y = \sum_{k=1}^{n} e^{\lambda_k x} Q_k(x)$ by the induction hypothesis where $Q_k(x)$ is a polynomial in x with degree $Q_k = m_k - 1$ if $k > 1$ and degree $Q_1 = m_1 - 2$.

But

$$(D - \lambda_1 I)y = \sum_{k=1}^{n} e^{\lambda_k x} Q_k(x) \Leftrightarrow D(e^{-\lambda_1 x} y) = \sum_{k=1}^{n} e^{(\lambda_k - \lambda_1)x} Q_k(x) \Leftrightarrow e^{-\lambda_1 x} y$$

$$= \sum_{k=1}^{n} \left(e^{(\lambda_k - \lambda_1)x} P_k(x) + C_k\right),$$

by Lemma 1, where $C_k \in \mathbb{C}$ and $P_k(x)$ is a polynomial in x with degree $P_k = m_k - 1$ if $k > 1$ and degree $P_1 = 1 + \text{degree } Q_1 = m_1 - 1$. Thus $P(D)y = 0 \Leftrightarrow e^{-\lambda_1 x}y \cdot = \sum_{k=1}^{n} e^{(\lambda_k - \lambda_1)x} P_k(x)$ where $(\sum_{k=1}^{n} C_k)$ has been absorbed into the polynomial $P_1(x)$ $\Leftrightarrow y(x) = \sum_{k=1}^{n} e^{\lambda_k x} P_k(x)$. This completes the induction and proves the theorem.

It may be noted that the proof given above is essentially a repeated application of the method of solving first-order linear equations by multiplication by an integrating factor.

The standard treatment of the nonhomogeneous equation is reproduced below for the sake of completeness.

3. THEOREM. *Let $N \in \mathbb{N}$, $a_0, a_1, \dots, a_N$ be complex numbers with $a_0 \neq 0$ and let $f\colon \mathbb{R} \to \mathbb{C}$ be a given function. If $y_0\colon \mathbb{R} \to \mathbb{C}$ is an N times differentiable function such that*

$$\sum_{k=0}^{N} a_k \frac{d^{N-k} y_0}{dx^{N-k}} = f$$

and $y\colon \mathbb{R} \to \mathbb{C}$ is an N times differentiable function then $\sum_{k=0}^{N} a_k (d^{N-k}y/dx^{N-k}) = f$ iff $y(x) = y_0(x) + \sum_{k=1}^{n} e^{\lambda_k x} P_k(x)$ where $\lambda_1, \dots, \lambda_n$ are the distinct roots of the polynomial $P(\lambda) = \sum_{k=0}^{N} a_k \lambda^{N-k}$ with multiplicities $m_1, \dots, m_n$ and $P_k\colon \mathbb{R} \to \mathbb{C}$ is a polynomial of degree $m_k - 1$.

Proof. With $D = d/dx$, $P(D)y = f \Leftrightarrow P(D)y = P(D)y_0 \Leftrightarrow P(D)(y - y_0) = 0$ by linearity of $P(D) \Leftrightarrow y(x) - y_0(x) = \sum_{k=1}^{n} e^{\lambda_k x} P_k(x)$ by Theorem 1.

Part II Variable Coefficients The approach indicated in Part I fails in the case of equations with variable coefficients since the differential polynomial $P(D)$ may not, in this case, be so readily factored as $Q(D)(D - \lambda I)$ even if λ and the coefficients of Q are permitted to be variable. However, an exponential shift technique can be used to solve $P(x, D)y = f(x)$ directly where $P(x, D)$ is a linear differential operator whose coefficients may not be constants provided that $P(x, D)$ may be factored into the form

$$P(x, D) = a_0(x)\left(D - b_1(x)I\right)^{m_1}\left(D - b_2(x)I\right)^{m_2} \cdots \left(D - b_p(x)I\right)^{m_p}$$

where $m_1, \dots, m_p$ are positive integers. This procedure is indicated below.

First, if ϕ and V are n times differentiable functions then the elementary "product rule" $D(e^{\phi(x)}V(x)) = e^{\phi(x)}(D + \phi'(x)I)V(x)$ leads by trivial induction to $D^k(e^{\phi(x)}V(x)) = e^{\phi(x)}(D + \phi'(x)I)^k V(x)$ for $k = 0, 1, \dots, n$. Consequently the following result is immediate.

4. LEMMA (*The exponential shift*). *Let $P(x, D) = \sum_{k=0}^{N} a_k(x) D^{N-k}$, where $D = d/dx$, be a linear ordinary differential operator of order N with continuous coefficients $\{a_k\}_{k=0}^{N}$ defined on an interval J of $\mathbb{R}$ and let $\phi\colon J \to \mathbb{C}$ be an N times differentiable function. Then, for each N times differentiable function $V\colon J \to \mathbb{C}$ there holds*

$$P(x, D)\left(e^{\phi(x)}V(x)\right) = e^{\phi(x)} P(x, D + \phi'(x)I)V(x).$$

Next suppose the differential operator $P(x, D)$ can be factored into the form

$$P(x, D) = a_0(x)\left(D - b_1(x)I\right)^{m_1}\left(D - b_2(x)I\right)^{m_2} \cdots \left(D - b_p(x)I\right)^{m_p},$$

where $m_1, \dots, m_p$ are positive integers with $m_1 + \cdots + m_p = N$ and for each k,

$b_k\colon J \to \mathbb{C}$ is $m_1 + \cdots + m_k - 1$ times differentiable. Also, without loss of generality, assume that $a_0(x) \neq 0$ for any $x \in J$. Then the differential equation $P(x, D)y = f(x)$ may be written $(D - b(x)I)^m Q(x, D)y = g(x)$ where $g(x) = f(x)/a_0(x)$, $b = b_1$, $m = m_1$ and $Q(x, D) = (D - b_2(x)I)^{m_2} \cdots (D - b_p(x)I)^{m_p}$ is a differential operator of the same form as $P(x, D)$. Hence, if for some fixed $c \in J$, $B\colon J \to \mathbb{C}$ is defined by $B(x) = \int_c^x b(t)\,dt$, then

$$\begin{aligned} P(x, D)y = f(x) &\Leftrightarrow (D - b(x)I)^m Q(x, D)y = g(x) \\ &\Leftrightarrow (D - b(x)I)^m \big(e^{B(x)} e^{-B(x)} Q(x, D)y\big) = g(x) \\ &\Leftrightarrow e^{B(x)} D^m \big(e^{-B(x)} Q(x, D)y\big) = g(x) \text{ by Lemma 4} \\ &\Leftrightarrow D^m \big(e^{-B(x)} Q(x, D)y\big) = g(x)e^{-B(x)} \\ &\Leftrightarrow e^{-B(x)} Q(x, D)y = p(x) + g_m(x), \end{aligned}$$

where $p(x)$ is a polynomial in x of degree $m - 1$ and g_m is the function defined inductively by $g_0(x) = g(x)e^{-B(x)}$ and $g_k(x) = \int_c^x g_{k-1}(t)\,dt$ so that $D^k g_k(x) = g_0(x)$ for $k = 0, 1, 2, \ldots, m$. Thus we have $P(x, D)y = f(x) \Leftrightarrow Q(x, D)y = (p(x) + g_m(x))e^{B(x)}$. By straightforward induction and reversal of order of integration in a repeated integral, it is easy to show that

$$g_m(x) = 1/(m-1)! \int_c^x (x-t)^{m-1} \big(f(t)/a_0(t)\big) e^{-B(t)}\,dt.$$

Altogether, therefore, we have established the following result.

5. Reduction Theorem. *Let J be an interval of the real line, $c \in J$ and let $P(x, D)$ be the linear differential operator defined by*

$$P(x, D) = a_0(x)(D - b_1(x)I)^{m_1}(D - b_2(x)I)^{m_2} \cdots (D - b_p(x)I)^{m_p}$$

where I is the identity, $D = d/dx$, $m_1, \ldots, m_p$ are positive integers, $a_0\colon J \to \mathbb{C}\setminus\{0\}$ is continuous and $\{b_k\}_1^p$ are continuous complex-valued functions defined on J with b_k differentiable $m_1 + \cdots + m_k - 1$ times. If $f\colon J \to \mathbb{C}$ is a continuous function, then

$$P(x, D)y = f(x) \Leftrightarrow Q(x, D)y = \big(p_1(x) + g_{m_1}(x)\big)e^{B_1(x)}$$

where $Q(x, D) = (D - b_2(x)I)^{m_2} \cdots (D - b_p(x)I)^{m_p}$, $p_1(x)$ is a polynomial in x of degree $m_1 - 1$, $B_1(x) = \int_c^x b_1(t)\,dt$ and g_{m_1} is the function defined by

$$g_{m_1}(x) = 1/(m_1-1)! \int_c^x (x-t)^{m_1-1} \big(f(t)/a_0(t)\big) e^{-B_1(t)}\,dt.$$

6. *Remark*. The reduction theorem can now be applied to the equation $Q(x, D)y = (p_1(x) + g_{m_1}(x))e^{B_1(x)}$ since this equation is of the same form as the equation $P(x, D)y = f(x)$. As the order of $Q(x, D)$ is less than that of $P(x, D)$, it is clear that repeated application of the theorem gives rise to the general solution of $P(x, D)y = f(x)$.

7. *Examples*

(a) Consider the reasonably simple problem:

$$y''' - 3y'' + 3y' - y = \frac{e^x}{x} \quad \text{on } 0 < x < \infty.$$

In trying to solve this by the method of variation of parameters, it will be observed that while it is quite easy to write down the complementary function, it takes quite a while to get at a particular integral. And it must of course be borne in mind that the solution so obtained is *not justified* to be *the general solution* unless the uniqueness theorem is invoked. In contrast, the *fully justified general solution* by the method (not the content) of the Reduction Theorem is obtained (in all its details) thus

$$\begin{aligned}(D-I)^3y=\frac{e^x}{x}&\Leftrightarrow (D-I)^3(e^xye^{-x})=\frac{e^x}{x}\\ &\Leftrightarrow e^xD^3(ye^{-x})=\frac{e^x}{x}\\ &\Leftrightarrow D^3(ye^{-x})=\frac{1}{x}\\ &\Leftrightarrow ye^{-x}=Ax^2+Bx+C+\frac{x^2}{2}\log x,\end{aligned}$$

where, in the last step, a few integrations by parts have been done; but that's all!

(b) Consider the more complicated problem:

$$x^3y''+(x^4+1)y'+x(x^2+1)y=2x^4+x^2+1 \quad \text{on } 0<x<\infty.$$

In this example there is no standard procedure to find the complementary function and hence the usual methods of variation of parameters and of undetermined coefficients are unavailable. However, the method of the reduction theorem succeeds as shown below.

First note that the differential operator $x^3D^2+(x^4+1)D+x(x^2+1)I$ can be factored as $(x^3D+I)(D+xI)$. The rest is easy enough.

$$\begin{aligned}&x^3y''+(x^4+1)y'+x(x^2+1)y=2x^4+x^2+1\\ &\Leftrightarrow (x^3D+I)(D+xI)y=2x^4+x^2+1\\ &\Leftrightarrow \left(D+\frac{1}{x^3}I\right)(D+xI)y=2x+\frac{1}{x}+\frac{1}{x^3}\\ &\Leftrightarrow \left(D+\frac{1}{x^3}I\right)\left(e^{\frac{1}{2x^2}}e^{-\frac{1}{2x^2}}(D+xI)y\right)=2x+\frac{1}{x}+\frac{1}{x^3}\\ &\Leftrightarrow e^{\frac{1}{2x^2}}D\left(e^{-\frac{1}{2x^2}}(D+xI)y\right)=2x+\frac{1}{x}+\frac{1}{x^3}\\ &\Leftrightarrow D\left(e^{-\frac{1}{2x^2}}(D+xI)y\right)=e^{-\frac{1}{2x^2}}\left(2x+\frac{1}{x}+\frac{1}{x^3}\right)\\ &\qquad\qquad\qquad\qquad\qquad\;=D\left(e^{-\frac{1}{2x^2}}(1+x^2)\right)\\ &\Leftrightarrow e^{-\frac{1}{2x^2}}(D+xI)y=B+e^{-\frac{1}{2x^2}}(1+x^2)\\ &\Leftrightarrow (D+xI)y=Be^{\frac{1}{2x^2}}+1+x^2\\ &\Leftrightarrow D\left(e^{\frac{x^2}{2}}y\right)=Be^{\frac{1+x^4}{2x^2}}+(1+x^2)e^{\frac{x^2}{2}}\end{aligned}$$

$$\Leftrightarrow e^{\frac{x^2}{2}} y = A + B\int_1^x e^{\frac{1+t^4}{2t^2}}\,dt + xe^{\frac{x^2}{2}}$$

$$\Leftrightarrow y = \left(A + B\int_1^x e^{\frac{1+t^4}{2t^2}}\,dt\right)e^{-\frac{x^2}{2}} + x.$$

8. *Remark*. The procedure of repeated application of the Reduction Theorem described above is applicable *mutatis mutandis* to the case in which $P(x, D)$ can be factored into the form

$$P(x, D) = \left(a_1(x)D - b_1(x)I\right)^{m_1} \cdots \left(a_p(x)D - b_p(x)I\right)^{m_p},$$

where $a_1, \dots, a_p$; $b_1, \dots, b_p$ are sufficiently differentiable complex-valued functions defined on an interval J of $\mathbb{R}$ with $a_k(x) \neq 0$ for each $x \in J$ and for each k and, as before, I denotes the identity operator. The details sketched below are analogous although slightly more involved.

9. Lemma. *Let J be an interval of the real line and let $c \in J$. If a: $J \to \mathbb{C}\setminus\{0\}$ is $n-1$ times differentiable and f: $J \to \mathbb{C}$ is continuous, then with $D = d/dx$,*

$$\left(a(x)D\right)^n y = f(x) \quad \text{iff } y(x) = \sum_{k=0}^{n-1} A_k C_k(x) + f_n(x)$$

where $A_0, A_1, \dots, A_{n-1}$ are arbitrary constants, $C_0(x) = 1$, $f_0(x) = f(x)$ and the functions $\{C_k\}_{k=1}^{n-1}$ and $\{f_k\}_{k=1}^{n}$ are defined inductively by

$$C_k(x) = \int_c^x \frac{C_{k-1}(t)}{a(t)}\,dt,\ f_k(x) = \int_c^x \frac{f_{k-1}(t)}{a(t)}\,dt$$

so that

$$C_k(x) = \frac{1}{(k-1)!}\int_c^x \frac{(x-t)^{k-1}}{a(t)}\,dt$$

and

$$f_k(x) = 1/(k-1)!\int_c^x (x-t)^{k-1}\left(f(t)/a(t)\right)dt.$$

10. Theorem. *Let J be an interval of the real line, $c \in J$ and let $P(x, D)$ be the linear differential operator defined by*

$$P(x, D) = \left(a_1(x)D - b_1(x)I\right)^{m_1} \cdots \left(a_p(x)D - b_p(x)I\right)^{m_p},$$

where $D = d/dx$, $m_1, \dots, m_p$ are positive integers, $a_1, \dots, a_p$; $b_1, \dots, b_p$ are sufficiently differentiable complex-valued functions defined on J with $a_k(x) \neq 0$ for each $x \in J$ and for each k, and I denotes the identity operator. If f: $J \to \mathbb{C}$ is continuous and

$$Q(x, D) = \left(a_2(x)D - b_2(x)I\right)^{m_2} \cdots \left(a_p(x)D - b_p(x)I\right)^{m_p},$$

then

$$P(x, D)y = f(x) \Leftrightarrow Q(x, D)y = e^{\phi_1(x)}\left(f_{m_1}(x) + \sum_{k=0}^{m_1-1} A_k C_k(x)\right)$$

where $A_0, \ldots, A_{m_1-1}$ *are arbitrary constants and the functions* ϕ_1, $\{C_k\}_{k=0}^{m_1-1}$ *and* f_{m_1} *are defined by the formulas*

$$\phi_1(x) = \int_c^x \frac{b_1(t)}{a_1(t)}\,dt,\; C_0(x) = 1,\; f_{m_1}(x) = \frac{1}{(m_1-1)!}\int_c^x (x-t)^{m_1-1}\frac{f(t)}{a_1(t)}\,dt$$

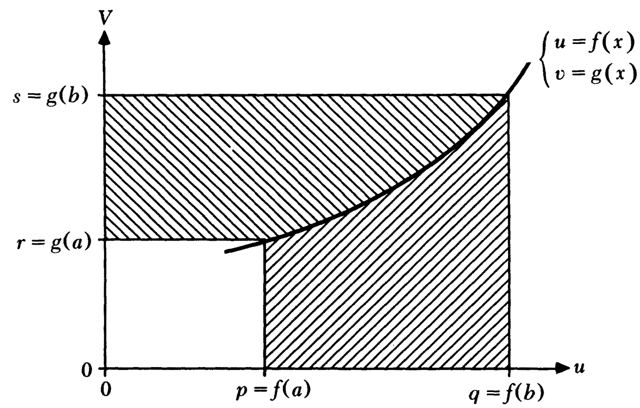
and for $k \geqslant 1$,

$$C_k(x) = \frac{1}{(k-1)!}\int_c^x \frac{(x-t)^{k-1}}{a_1(t)}\,dt.$$

REFERENCES

1. R. C. Buck and E. F. Buck, *Introduction to Differential Equations*, Houghton-Mifflin, Boston, 1976.
2. R. W. Hunt, *Differential Equations*, Brooks/Cole, Belmont, CA, 1973.
3. D. L. Kreider, R. G. Kuller, D. R. Ostberg and F. W. Perkins, *An Introduction to Linear Analysis*, Addison-Wesley, Reading, MA, 1966.
4. R. H. Martin, *Ordinary Differential Equations*, McGraw-Hill, Tokyo, 1983.
5. R. C. McCann, *Introduction to Ordinary Differential Equations*, Harcourt Brace Jovanovich, New York, 1982.
6. E. D. Rainville and P. E. Bedient, *Elementary Differential Equations*, Macmillan, New York, 1981.
7. S. L. Ross, *Introduction to Ordinary Differential Equations*, John Wiley and Sons, Inc., New York, 1974.
8. G. F. Simmons, *Differential Equations*, Tata-McGraw-Hill, New Delhi, 1974.
9. B. Spain, *Ordinary Differential Equations*, Van-Nostrand Reinhold, London, 1969.
10. M. R. Spiegel, *Applied Differential Equations*, Prentice Hall, Englewood Cliffs, NJ, 1981.

Proof without Words: Integration by Parts



Area + Area $= qs - pr$

$$\int_r^s u\,dv + \int_p^q v\,du = uv\Big|_{(p,r)}^{(q,s)}$$

$$\int_a^b f(x)g'(x)\,dx = f(x)g(x)\Big|_a^b - \int_a^b g(x)f'(x)\,dx$$

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

where $A_0, \ldots, A_{m_1-1}$ *are arbitrary constants and the functions* ϕ_1, $\{C_k\}_{k=0}^{m_1-1}$ *and* f_{m_1} *are defined by the formulas*

$$\phi_1(x) = \int_c^x \frac{b_1(t)}{a_1(t)}\,dt,\ C_0(x) = 1,\ f_{m_1}(x) = \frac{1}{(m_1-1)!}\int_c^x (x-t)^{m_1-1}\frac{f(t)}{a_1(t)}\,dt$$

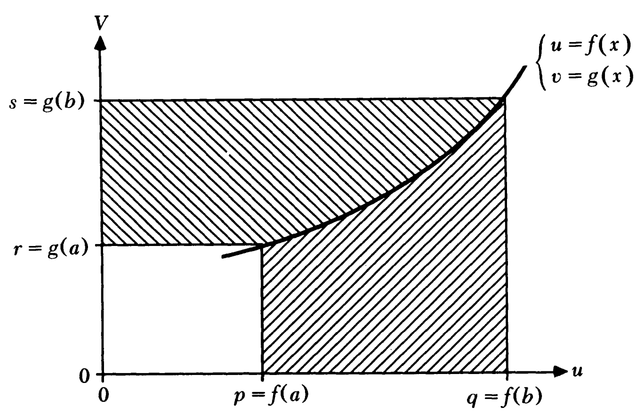
and for $k \geqslant 1$,

$$C_k(x) = \frac{1}{(k-1)!}\int_c^x \frac{(x-t)^{k-1}}{a_1(t)}\,dt.$$

REFERENCES

1. R. C. Buck and E. F. Buck, *Introduction to Differential Equations*, Houghton-Mifflin, Boston, 1976.
2. R. W. Hunt, *Differential Equations*, Brooks/Cole, Belmont, CA, 1973.
3. D. L. Kreider, R. G. Kuller, D. R. Ostberg and F. W. Perkins, *An Introduction to Linear Analysis*, Addison-Wesley, Reading, MA, 1966.
4. R. H. Martin, *Ordinary Differential Equations*, McGraw-Hill, Tokyo, 1983.
5. R. C. McCann, *Introduction to Ordinary Differential Equations*, Harcourt Brace Jovanovich, New York, 1982.
6. E. D. Rainville and P. E. Bedient, *Elementary Differential Equations*, Macmillan, New York, 1981.
7. S. L. Ross, *Introduction to Ordinary Differential Equations*, John Wiley and Sons, Inc., New York, 1974.
8. G. F. Simmons, *Differential Equations*, Tata-McGraw-Hill, New Delhi, 1974.
9. B. Spain, *Ordinary Differential Equations*, Van-Nostrand Reinhold, London, 1969.
10. M. R. Spiegel, *Applied Differential Equations*, Prentice Hall, Englewood Cliffs, NJ, 1981.

Proof without Words: Integration by Parts



Area [back-hatched] + Area [hatched] $= qs - pr$

$$\int_r^s u\,dv + \int_p^q v\,du = uv\Big|_{(p,r)}^{(q,s)}$$

$$\int_a^b f(x)g'(x)\,dx = f(x)g(x)\Big|_a^b - \int_a^b g(x)f'(x)\,dx$$

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

PROBLEMS

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

GEORGE GILBERT, *associate editor*
Texas Christian University

Proposals

To be considered for publication, solutions should be received by September 1, 1991.

1368. *Proposed by Allen J. Schwenk, Western Michigan University, Kalamazoo, Michigan.*

In my kitchen I have a set of three measuring cups with capacities c, $1-c$, and 1 where $0<c<1/2$. Starting with the largest cup full, I can measure several additional fractions of a cup by sequentially pouring from one cup to another. I always pour until either the receiving cup is full or until the dispensing cup is empty. I never spill, waste, or consume any of the liquid. Define the accuracy of this set of cups to be the length ε of the longest interval $(a, a+\varepsilon)$ in $[0,1]$ for which no $x \in (a, a+\varepsilon)$ can be measured. For example, when $c = 1/n$ I can measure $2/n$ by filling the smallest cup, transferring this $1/n$ to the middle cup, refilling the smallest, and transferring again. Clearly each fraction i/n can be measured in this way, so for $c = 1/n$ the accuracy is also $1/n$.

For each value of c, find the accuracy of the set.

1369. *Proposed by Mihály Bencze, Braşov, Romania.*

a. Find all natural numbers x, y, z such that $3^x + 4^y = 5^z$.

b.* Given natural numbers $A > B$ and C, find all natural numbers x, y, z such that

$$\left((A^2-B^2)C\right)^x + (2ABC)^y = \left((A^2+B^2)C\right)^z.$$

ASSISTANT EDITORS: CLIFTON CORZATT, BRUCE HANSON, RICHARD KLEBER, KAY SMITH *and* THEODORE VESSEY, *St. Olaf College and* MARK KRUSEMEYER, *Carleton College. We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals should be accompanied by solutions, if at all possible, and by any other information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution. An asterisk (*) next to a problem number indicates that neither the proposer nor the editors supplied a solution.*

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

Solutions and new proposals should be mailed in duplicate to Loren C. Larson, Department of Mathematics, St. Olaf College, Northfield, MN 55057.

1370. *Proposed by R. S. Luthar, University of Wisconsin Center, Janesville, Wisconsin.*

a. If x, y, and z are positive real numbers with $x+y+z=1$, prove that

$$\frac{2}{3} \leqslant \frac{\ln(x^5+y^5+z^5)}{\ln(x^7+y^7+z^7)} \leqslant \frac{5}{7}.$$

b. Generalize the result of part a.

1371. *Proposed by Hüseyin Demir, Middle East Technical University, Ankara, Turkey.*

Let A, B, and C be vertices of a triangle and let D, E, and F be points on the sides BC, AC, and AB, respectively. Let U, X, V, Y, W, Z be the midpoints of, respectively, BD, DC, CE, EA, AF, FB. Prove that

$$\text{Area}(\triangle UVW) + \text{Area}(\triangle XYZ) - \frac{1}{2}\text{Area}(\triangle DEF)$$

is a constant independent of D, E, and F.

1372. *Proposed by Nick Lord, Tonbridge School, Tonbridge, Kent, England.*

For which angles θ, a rational number of degrees, is it the case that $\tan^2\theta + \tan^2 2\theta$ is rational?

Quickies

Answers to the Quickies are on page 137.

Q775. *Proposed by Neil Eklund, Centre College, Danville, Kentucky.*

Evaluate the integral

$$\int_0^1 \frac{x^k-1}{\ln x}\,dx.$$

Q776. *Proposed by Chris Caldwell, University of Tennessee at Martin, Martin, Tennessee.*

Show that if $P(x)$ is a real polynomial with real zeros, then $P'(x)^2 \geqslant P(x)P''(x)$ for all real x.

Q777. *Proposed by Murray S. Klamkin and Andy Liu, University of Alberta, Edmonton, Alberta, Canada.*

T_1 and T_2 are two acute triangles inscribed in the same circle. If the perimeter of T_1 is greater than the perimeter of T_2, must the area of T_1 also be greater than the area of T_2?

Solutions

A Curious Recursion **April 1990**

1343. *Proposed by Ronald L. Graham, AT&T Bell Labs, Murray Hill, New Jersey.*

a. What is the behavior of the recursive sequence defined by $x_{n+2} = (1 + x_{n+1})/x_n$, with x_0, x_1 arbitrary positive numbers?

b. Answer the same question for $x_{n+3} = (1 + x_{n+1} + x_{n+2})/x_n$.

c.* Generalize (a) and (b).

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

a. The sequence may be traced backwards using $x_n = (1 + x_{n+1})/x_{n+2}$, and there is no harm in extending this to negative indices. We then have

$$x_2 = \frac{1+x_1}{x_0}, \qquad x_3 = \frac{1+(1+x_1)/x_0}{x_1} = \frac{1+x_0+x_1}{x_0x_1}$$

as well as

$$x_{-1} = \frac{1+x_0}{x_1}, \qquad x_{-2} = \frac{1+(1+x_0)/x_1}{x_0} = \frac{1+x_0+x_1}{x_0x_1}.$$

So $x_{-2} = x_3$. Since it is obviously immaterial which two successive terms are called x_0 and x_1 we conclude that for all integers n, $x_n = x_{n+5}$. That is, the sequence is periodic with period 5.

b. We proceed as in part a.

$$x_3 = \frac{1+x_1+x_2}{x_0}, \quad x_4 = \frac{1+x_2+((1+x_1+x_2)/x_0)}{x_1} = \frac{1+x_0+x_1+x_2+x_0x_2}{x_0x_1};$$

$$x_{-1} = \frac{1+x_0+x_1}{x_2}, \quad x_{-2} = \frac{1+x_0+((1+x_0+x_1)/x_2)}{x_1} = \frac{1+x_0+x_1+x_2+x_0x_2}{x_1x_2}$$

So $x_{-2}x_2 = x_0x_4$. As above we conclude that $x_nx_{n+4} = x_{n+2}x_{n+6}$ for all integers n. But we then also have $x_{n+2}x_{n+6} = x_{n+4}x_{n+8}$, so $x_nx_{n+4} = x_{n+4}x_{n+8}$, whence $x_n = x_{n+8}$ (since x_{n+4} is positive). That is, the sequence is periodic with period 8.

Also solved, in some cases using Mathematica *or* MAPLE, *by Charles Ashbacher, Duane Broline, Francis M. Henderson, R. High, Kee-Wai Lau (Hong Kong), Stanley Wu-Wei Liu, Reiner Martin (Germany), Mike Pinter, Robert L. Raymond et al., Volkhard Schindler (Germany), Heinz-Jürgen Seiffert (Germany), Stephen L. Snover, John S. Sumner, University of Wyoming Problem Circle, Ravi Vakil (student, Canada), Michael Vowe (Switzerland), XAMPLE (Xavier Mathematics Problem League), and the proposer. There was one partially correct solution.*

Con Amore Problem Group, Henderson, and *Sumner* investigated the possibility of the sequence having a shorter period (1 in part a and 1, 2, or 4 in part b) for specific initial values. For example, in part b, period 2 occurs when $x_2 = x_0 > 1$, $x_1 = (x_0 + 1)/(x_0 - 1)$.

Several solvers investigated possible generalizations, ranging from

$$x_{n+k} = (1 + x_{n+1} + \cdots + x_{n+k-1})/x_n,$$

through similar expressions in which the x_i are moved to the denominator. *Con Amore Problem Group, High, Liu, Raymond et al.,* and *Snover* reported "pseudoperiodicity," usually on the basis of computer experiments. Specifically, for $k = 4$ and starting values such as $x_0 = x_1 = x_2 = x_3 = 1$, the sequence "almost" has period 11. *Liu* presented an especially detailed discussion with various observations, conjectures, and numerical evidence about the behavior of recurrences such as $x_{n+4} = (1 + x_{n+3} + x_{n+1})/(x_{n+2} + x_n)$ (and many others). However, no solver found a generalization of the *results* of parts a and b, and several speculated that no such generalization might exist.

Generalization of gcd-lcm Formula **April 1990**

1344. *Proposed by Keith Loseke, student, Buhler High School, Buhler, Kansas.*

Prove the following generalization of the formula $\gcd(a, b) = ab/\operatorname{lcm}(a, b)$:

$$\gcd(a_1, a_2, \dots, a_k) = \frac{P_0}{P_E},$$

where $a_1, a_2, \dots, a_k$ are positive integers, "gcd" and "lcm" are abbreviations for greatest common divisor and least common multiple, P_0 is the product of the lcm's of all subsets of $\{a_1, a_2, \dots, a_k\}$ with an odd number of elements, and P_E is the similar product of the non-empty subsets of even numbers of elements.

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

For $1 \leqslant i \leqslant k$ we can let $a_i = \prod_{j=1}^{n} p_j^{c_{ij}}$, where the p_j's are distinct primes and $c_{ij} \geqslant 0$. Then for I a subset of $\{1, 2, \dots, k\}$,

$$\operatorname{lcm}\{a_i\}_{i \in I} = \prod_{j=1}^{n} p_j^{\max_{i \in I}\{c_{ij}\}} \qquad \text{and} \qquad \gcd\{a_i\}_{i \in I} = \prod_{j=1}^{n} p_j^{\min_{i \in I}\{c_{ij}\}}.$$

Fix j and denote c_{ij} by c_i. Without loss of generality, assume $c_1 \leqslant c_2 \leqslant \cdots \leqslant c_n$. Then we must show that

$$\sum_{I:\,|I|\text{ odd}} \max_{i \in I}\{c_i\} - \sum_{I:\,|I|\text{ even}} \max_{i \in I}\{c_i\} = c_1.$$

Clearly c_1 occurs once in the first sum on the left side and not at all in the second. For $i \geqslant 1$, c_{i+1} occurs $\binom{i}{0} + \binom{i}{2} + \binom{i}{4} + \cdots = 2^{i-1}$ times in the first sum and $\binom{i}{1} + \binom{i}{3} + \cdots = 2^{i-1}$ times in the second, and hence cancels. The result follows.

Also solved by Duane M. Broline, Con Amore Problem Group (Denmark), David Doster, Newcomb Greenleaf, Russell Jay Hendel, R. High, Kee-Wai Lau (Hong Kong), Sandor Lehoczky (student), N. J. Lord (England), David E. Manes, Jean-Marie Monier (France), Allan Pedersen (Denmark), Vivek Prabhakaran (student), Brian Ruder (student), Heinz-Jürgen Seiffert (Germany), John S. Sumner, Ravi Vakil (student, Canada), Ellen Veed, University of Wyoming Problem Circle, XAMPLE (Xavier Mathematics Problem League), Paul J. Zwier, and the proposer.

Manes notes that this problem is Theorem 7.4, p. 11, in L. K. Hua's book *Introduction to Number Theory*, Springer-Verlag, New York, 1982. Notice also the related result, Theorem 7.3: lcm $(a_1, a_2, \dots, a_k) = G_O/G_E$, where G_O is the product of the gcd's of all subsets of $\{a_1, a_2, \dots, a_k\}$ with an odd number of elements and G_E is the similar product of the nonempty subsets of even numbers of elements. This related result also appears as Problem 29, p. 117, and p. 308, in G. Pólya and G. Szegö, *Problems and Theorems in Analysis II*, Springer-Verlag, New York, 1976, where the proof is based on the inclusion-exclusion principle.

Limit of an Inverse Image **April 1990**

1345. *Proposed by the Fullerton Problem Group, California State University, Fullerton, California.*

Let $A(\varepsilon)$ be the union of all maximal bounded closed intervals of $(0,1)$ such that $-\varepsilon \leqslant \sin(1/x) \leqslant \varepsilon$ for every $x \in A(\varepsilon)$. Denote by $S(\varepsilon)$ the sum of the lengths of all these intervals. Prove that

$$\lim_{\varepsilon \to 0} \frac{S(\varepsilon)}{\varepsilon} = \frac{1}{3}.$$

Solution by the University of Wyoming Problem Circle, University of Wyoming, Laramie, Wyoming.

If $0<\varepsilon<1$, then the set of y's for which $-\varepsilon \leqslant \sin y \leqslant \varepsilon$ is the union of the intervals $[k\pi-a(\varepsilon), k\pi+a(\varepsilon)]$ as k ranges over the set of integers. Here, $a(\varepsilon) \equiv \sin^{-1}\varepsilon$. It follows that

$$A(\varepsilon)=\bigcup_{k=1}^{\infty}\left[\frac{1}{k\pi+a(\varepsilon)}, \frac{1}{k\pi-a(\varepsilon)}\right]$$

so that

$$\begin{aligned}\frac{S(\varepsilon)}{\varepsilon} &= \frac{1}{\varepsilon}\sum_{k=1}^{\infty}\left(\frac{1}{k\pi-a(\varepsilon)}-\frac{1}{k\pi+a(\varepsilon)}\right)\\ &= \frac{1}{\varepsilon}\sum_{k=1}^{\infty}\frac{2a(\varepsilon)}{k^2\pi^2-a^2(\varepsilon)}\\ &= \frac{a(\varepsilon)}{\varepsilon}\cdot\frac{2}{\pi^2}\cdot\sum_{k=1}^{\infty}\frac{1}{k^2-\dfrac{a^2(\varepsilon)}{\pi^2}}.\end{aligned}$$

Now, by L'Hôpital's Rule,

$$\lim_{\varepsilon\to 0}\frac{a(\varepsilon)}{\varepsilon}=\lim_{\varepsilon\to 0}\frac{\sin^{-1}\varepsilon}{\varepsilon}=\lim_{\varepsilon\to 0}\frac{(1-\varepsilon^2)^{-1/2}}{1}=1.$$

Also, the series defines a continuous function of ε on the interval $[0,1]$ since it is a uniformly convergent series of continuous functions. In fact, the Weierstrass M-Test can be applied since

$$\frac{1}{k^2-\dfrac{a^2(\varepsilon)}{\pi^2}} \leqslant \frac{1}{k^2-\dfrac{1}{4}},$$

and the series $\Sigma(k^2-(1/4))^{-1}$ converges. Consequently,

$$\lim_{\varepsilon\to 0^+}\frac{S(\varepsilon)}{\varepsilon}=1\cdot\frac{2}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2}=\frac{2}{\pi^2}\frac{\pi^2}{6}=\frac{1}{3}.$$

Also solved by Seung-Jin Bang (Korea), S. F. Barger, Larry Blaine and T. A. Giebutowski, Paul Bracken (Canada), David Callan, Con Amore Problem Group (Denmark), Jesse Deutsch, David Doster, Robert Doucette, Mordechai Falkowitz (Israel), H. Guggenheimer, Kee-Wai Lau (Hong Kong), Sandor Lehoczky (student), Peter W. Lindstrom, N. J. Lord (England), Reiner Martin, Jean-Marie Monier (France), William A. Newcomb, Stephen Noltie, Stephen G. Penrice, C. Ray Rosentrater, Heinz-Jürgen Seiffert (Germany), John S. Sumner, Gan Wee Teck (student, Singapore), Western Maryland College Problems Group, Michael Vowe, and the proposer.

Solitaire Equivalence **April 1990**

1346. *Proposed by Jerry M. Metzger, University of North Dakota, Grand Forks, North Dakota.*

Consider a solitaire game played as follows. From an ordinary deck of 52 cards, lay out the top 8 cards face up in a row. If two or more cards have the same rank (for example, two eights or two jacks), pick two of them and cover them by two cards dealt face up from the top of the deck. Continue in this way as long as at least two of the cards show the same rank. The game ends if one cannot play, and is a win if the deck is exhausted. Prove that the game is a win working from the top to the bottom if and only if it is a win with the deck turned over (that is, working from the bottom to the top).

Solution by Ravi Vakil, student, University of Toronto, Toronto, Canada.

The game is a win working from the top of the deck if and only if, for all even n, there are at most 7 unmatched cards in the top n cards of the deck. But then this means that there are at most 7 unmatched cards in the bottom $52-n$ cards for all even n. (Each unmatched card in the top n cards gives rise to an unmatched card of the same rank in the bottom $52-n$ cards.) A game is a win if and only if the game with the deck turned over is a win.

Also solved by Larry Blaine, Con Amore Problem Group (Denmark), Roger B. Eggleton (Brunei Darussalam), R. High, John and Libby Krussel, Mary Ann Martin (student), William J. Metzger, Adam Riese, Harry Sedinger, Paul J. Zwier, and the proposer.

Center of the General Linear Group **April 1990**

1347. *Proposed by James O. Chilaka, Long Island University, Greenvale, New York.*

Let $G = GL(n, \mathbf{R})$ be the group under multiplication of $n \times n$ matrices with real entries and non-zero determinant. Find the least number k of elements $g_1, g_2, \ldots, g_k$ in G such that if a is in G and $a \cdot g_i = g_i \cdot a$, $i = 1, 2, \ldots, k$, then $a \cdot g = g \cdot a$ for all g in G.

Solution by Reiner Martin, student, University of California, Los Angeles, California.

In the non-trivial case, $n \geqslant 2$, the answer is $k = 2$.

Let $g_1 \in G$ be the diagonal matrix with entries $1, 2, \ldots, n$ and let $g_2 = (\delta_{i\sigma(j)}) \in G$ be the permutation matrix with $\sigma(k) = k+1$ for $k = 1, \ldots, n-1$ and $\sigma(n) = 1$. Now let $a = (a_{ij}) \in G$ be arbitrary. Then $a \cdot g_1 = g_1 \cdot a$ implies $a_{ij} = 0$ for $i \neq j$ and $a \cdot g_2 = g_2 \cdot a$ implies $a_{11} = \cdots = a_{nn}$, so both conditions together imply that a is a scalar multiple of the identity matrix, hence $a \cdot g = g \cdot a$ for all $g \in G$.

It remains to show that k cannot be improved to 1. Therefore, we assume that there exists a $g_1 \in G$ with the property that if $a \in G$ commutes with g_1 then a commutes with every element of G. Choosing $a = g_1$, we see that an arbitrary element of G commutes with g_1. So every element of G commutes with every element of G, a contradiction to the fact that G is not commutative.

Also solved by Ricardo Alfaro, S. F. Barger, Duane M. Broline, David Callan, Con Amore Problem Group (Denmark), David W. Koster, Marvin Marcus and Susan Franklin, Daniel B. Shapiro, John S. Sumner, Western Maryland College Problems Group, and the proposer.

Answers

Solutions to the Quickies on p. 132.

A775. *Solution* 1. Note that for $x > 0$,

$$\int_0^k e^{t\ln x}\,dt = \frac{1}{\ln x}e^{t\ln x}\Big|_{t=0}^{t=k} = \frac{1}{\ln x}\left(e^{k\ln x}-1\right) = \frac{x^k-1}{\ln x}.$$

Therefore,

$$I \equiv \int_0^1 \frac{x^k-1}{\ln x}\,dx = \int_0^1\int_0^k e^{t\ln x}\,dt\,dx.$$

If we interchange the order of integration we get

$$I = \int_0^k\int_0^1 x^t\,dx\,dt = \int_0^k \frac{1}{t+1}\,dt = \ln(k+1).$$

Solution 2. Let

$$g(k) = \int_0^1 \frac{x^k-1}{\ln x}\,dx.$$

Then

$$g'(k) = \int_0^1 x^k\,dx = \frac{1}{k+1}.$$

Integrating back, we find that $g(k) = \ln(k+1)$.

Comment. This integral can be found in various integral handbooks, and the case $k = 2$ appears in *Mathematical Spectrum*, Vol. 22, no. 1, problem 22.4, 1989–90.

A776. If x is one of the roots $r_1, r_2, \ldots, r_n$ of $P(x)$, then the result is obvious. Otherwise, we have

$$\frac{P'(x)}{P(x)} = \sum_{i=1}^{n}\frac{1}{x-r_i},$$

and differentiating each side,

$$\frac{P(x)P''(x)-\left(P'(x)\right)^2}{P(x)^2} = \sum_{i=1}^{n}\frac{-1}{\left(x-r_i\right)^2} < 0,$$

which completes the proof. (Notice that equality holds if and only if x is a multiple root of P.)

A777. By considering two triangles with angles $(80°, 50°, 50°)$ and $(70°, 70°, 40°)$ the answer is in the negative.

The result would be valid for two general triangles if the angles of T_2 majorized those of T_1, that is, if $A_1 \geqslant B_1 \geqslant C_1$, $A_2 \geqslant B_2 \geqslant C_2$, then $A_2 \geqslant A_1$ and $A_2 + B_2 \geqslant A_1 + B_1$. Then by the majorization inequality,

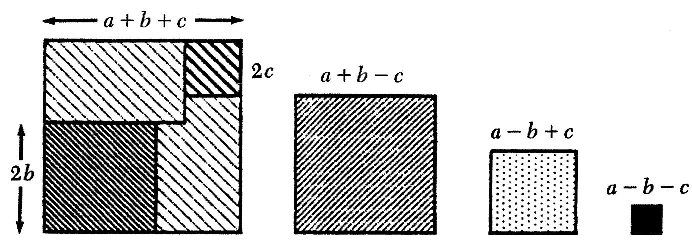
$$F(A_1) + F(B_1) + F(C_1) \geqslant F(A_2) + F(B_2) + F(C_2) \qquad (1)$$

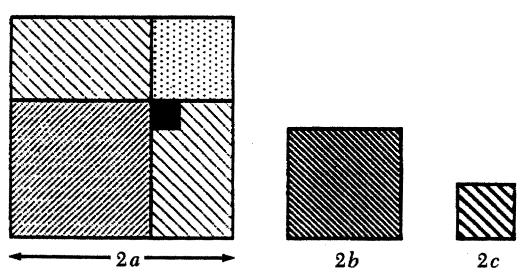
for concave functions F. The rest follows since the perimeter and area of a triangle ABC is given by $2R(\sin A + \sin B + \sin C)$ and $2R^2 \sin A \sin B \sin C$, respectively, and $\sin x$ and $\ln \sin x$ are concave on $(0, \pi)$.

Comments. For the special case when T_1 and T_2 have a common angle (or equivalently a common side), then the angles of T_2 majorize those of T_1. It would be of interest to give an elementary geometric proof of (1) for $F(x) = \sin x$ or $\ln \sin x$.

Proof without Words: Algebraic Areas

$$(a+b+c)^2 + (a+b-c)^2 + (a-b+c)^2 + (a-b-c)^2 = 2^2(a^2+b^2+c^2)$$





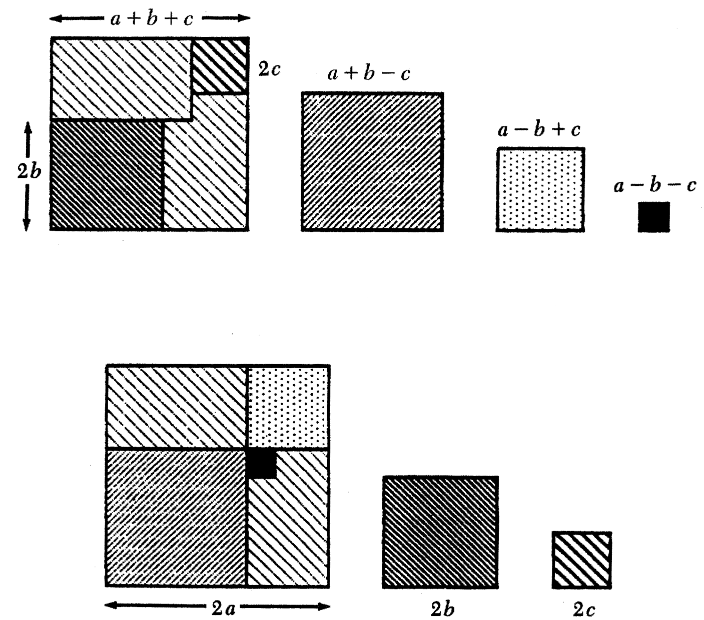
—Sam Pooley
K. Ann Drude
Millsaps College
Jackson, MS 39210

for concave functions F. The rest follows since the perimeter and area of a triangle ABC is given by $2R(\sin A + \sin B + \sin C)$ and $2R^2 \sin A \sin B \sin C$, respectively, and $\sin x$ and $\ln \sin x$ are concave on $(0, \pi)$.

Comments. For the special case when T_1 and T_2 have a common angle (or equivalently a common side), then the angles of T_2 majorize those of T_1. It would be of interest to give an elementary geometric proof of (1) for $F(x) = \sin x$ or $\ln \sin x$.

Proof without Words: Algebraic Areas

$$(a+b+c)^2 + (a+b-c)^2 + (a-b+c)^2 + (a-b-c)^2 = 2^2(a^2+b^2+c^2)$$



—Sam Pooley
K. Ann Drude
Millsaps College
Jackson, MS 39210

REVIEWS

PAUL J. CAMPBELL, *editor*
Beloit College

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

Bower, B., College classes spur lifelong math memory, *Science News* (15 December 1990) 375.

"People who take college mathematics courses at or above the level of calculus retain most of their knowledge of high school algebra or geometry up to 50 years later, while those who take no college mathematics courses suffer steep declines in algebra and geometry knowledge during adulthood." Are you surprised? The author of the study, Harry P. Bahrick (Ohio Wesleyan), also makes a case for promoting retention by spacing out learning sessions and making all examinations cumulative, principles which are also at the heart of John Saxon's controversial texts and learning philosophy. For details, see *Journal of Experimental Psychology: General* (March 1990).

Daviss, Bennett, Laid-back computers, *Discover* 12:1 (January 1991) 60–61.

The fuzzy set theory of the 1960s, in which set membership is conceived of as a matter of degree, has reemerged as the backdrop behind fuzzy logic, now being incorporated into controls of consumer products. Fuzzy logic, touted as the "first widely commercial form of artificial intelligence," allows finer, smoother, fast, and more flexible control of an appliance, based on the input of sensors to a fuzzy-logic computer chip. Matsushita offers a fuzzy washing machine (sensors judge how dirty the clothes are and how big the load is, then choose one of 600 cleaning cycles) and a fuzzy vacuum cleaner (which adapts its suction to the floor surface and the volume of dirt), and has designed a fuzzy showerhead (which keeps the water temperature constant). Subway trains in Sendai, Japan, run under the control of computers employing fuzzy logic; Nissan offers a car with a fuzzy automatic transmission and may be working on fuzzy antiskid brakes. But fuzzy logic has failed to catch on with American engineers. Could the reason be cultural differences? terminology (what do the Japanese call fuzzy logic?)? the Japanese government's $36 million subsidy of a laboratory for fuzzy engineering research? a greater Japanese propensity for techno-gadgets?

Taubes, Gary, Great balls of carbon, *Discover* 11(9) (September 1990) 52–59. Ross, Philip E., Buckeyballs: Fullerenes open new vistas in chemistry, *Scientific American* (January 1991) 114–116.

Chemists are now making macro quantities of fullerenes, molecules made of large numbers of carbon atoms. These new molecules containing only carbon, misfortunately named after Buckminster Fuller, take the shapes of polyhedra well-known to mathematicians (but not necessarily to chemists, some of whom have been rediscovering the relevant geometry). For example, buckminsterfullerene, the molecule with 60 carbon atoms, denoted C_{60}, forms a truncated icosahedron. The molecules C_{70} and C_{84} have also been discovered; like C_{60}, they are easily produced in soot. Fullerenes with 240, 540, 960, or more atoms may be lurking there, too; and chemists can knock atoms off to make fullerenes with any even number of atoms in between. How much do mathematicians know about possible minimal-energy polyhedra with so many vertices?

Levi, Barbara Goss, Are fractures fractal or quakes chaotic?, *Physics Today* 93(11) (November 1990) 17–19.

Are earthquakes fractal in nature? (This, of course, is the question that immediately came to mind when a temblor shook my bed this morning.) One of the early quantitative discoveries about earthquakes is that the frequency of a quake of magnitude m or greater is 10^{c-bm}, with b and c constants and b near 1. "Such a relationship indicates that small events follow the same laws as large events, and thus implies that all events must be governed by a scale-invariant mechanism." Thus, earthquakes exhibit the self-similarity at a change of scale that is characteristic of fractals. Simulations of earthquakes by systems of blocks and springs exhibit a similar power law and an earthquake-like "self-organized criticality" (which characterizes marginally stable systems that, when perturbed, tend to return to marginal stability rather than move to greater stability). With differential friction forces on as few as two blocks, chaotic trajectories result; whether chaos characterizes real fault systems is not known (and may be unknowable).

Krasner, Saul, ed., *The Ubiquity of Chaos*, American Association for the Advancement of Science, 1990; viii + 247 pp, $31.50 (P). ISBN 0-87168-350-4

Papers presented at a special session at the 1989 AAAS Annual Meeting, noting applications of chaos in many fields, including business cycles, bird flocks, and childhood epidemics, as well as non-applications: sudden death from heart arrhythmia does *not* result from chaotic heartbeat, though "the electrical activity of the heart under healthy conditions appears to be a fractal."

Nievergelt, Yves, *Fractals on Hewlett-Packard Supercalculators*, Preliminary Edition, Wadsworth, 1991; vi + 74 pp, $8 (stapled). ISBN 0-534-15037-3

At the AMS Centennial banquet in Atlanta in 1988, almost 2,000 mathematicians took delivery of HP-28 programmable calculators capable of symbolic computation, graphics, and lots more. How many of those calculators are in daily use now? (I confess mine isn't; I prefer my surviving HP-25, which fits in my left palm and I can operate with my left thumb, though I rarely program it.) Well, you may not have thought about using an HP-28 or an HP-48 to graph fractals; but you can, and this booklet shows you how. The booklet's purpose is to convince students of the relevance of the "abstract" topics of introductory analysis and topology to the "concrete" pictures of fractals. The booklet, which could serve as either a supplement in one of several courses or as a text in a seminar on fractals, primarily develops the theory behind fractals (including proofs), secondarily devises algorithms to display them, and then uses the HP calculators to implement the algorithms. Potential users of the booklet should be aware that HP now markets a software development system for these HP calculators on the Macintosh platform. The place for graphics and symbolic manipulation calculators is still undetermined, as one might expect in an educational system that still makes virtually no use of simpler calculators in elementary school; this booklet helps demonstrate that short programs on sophisticated calculators can do mathematically interesting computations.

Schoenfeld, Alan (ed.), *A Source Book for College Mathematics Teaching: A Report from the MAA Committee on the Teaching of Undergraduate Mathematics*, MAA, 1990; ii + 68 pp (P). ISBN 0-88385-068-0

A valuable handbook of resources and perspectives for improving college mathematics learning, with sections on goals, curriculum, resources, advising, evaluation, computers, and instructional techniques.

NEWS AND LETTERS

51st ANNUAL WILLIAM LOWELL PUTNAM MATHEMATICAL COMPETITION

These solutions have been contributed by Loren Larson, St. Olaf College.

A-1. Let

$$T_0 = 2,\ T_1 = 3,\ T_2 = 6,$$

and for $n \geq 3$,

$$T_n = (n+4)T_{n-1} - 4nT_{n-2} + (4n-8)T_{n-3}.$$

The first few terms are

$$2,\ 3,\ 6,\ 14,\ 40,\ 152,\ 784,\ 5168,\ 40576.$$

Find, with proof, a formula for T_n of the form $T_n = A_n + B_n$, where (A_n) and (B_n) are well-known sequences.

Solution. The formula for T_n is $T_n = n! + 2^n$.

Let $P(n)$ be the statement that $T_n = n! + 2^n$. $P(0), P(1), P(2)$ are true; assume that $P(n)$ is true for $n = 0, 1, 2, \ldots, k-1$. Then

$$\begin{aligned} T_k = (k+4)&\left[(k-1)! + 2^{k-1}\right] \\ &- 4k\left[(k-2)! + 2^{k-2}\right] \\ &+ (4k-8)\left[(k-3)! + 2^{k-3}\right], \end{aligned}$$

and the right side reduces to $k! + 2^k$. Thus, $P(k)$ is true, so by induction, $P(n)$ holds for all non-negative integers n.

A-2. Is $\sqrt{2}$ the limit of a sequence of numbers of the form $\sqrt[3]{n} - \sqrt[3]{m}$ $(n, m = 0, 1, 2, \ldots)$?

Solution. Let $n = m + \lfloor 3\sqrt{2}m^{2/3} \rfloor$. Then $\lim_{m\to\infty}\left(n^{1/3} - m^{1/3}\right) = \sqrt{2}$.

To see this, apply the Mean Value Theorem to $x^{1/3}$ over the interval $[m, n]$. Thus, there is a number $c = c(n)$ between m and n such that

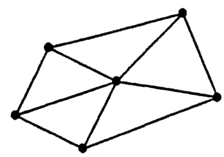
$$\begin{aligned} n^{1/3} - m^{1/3} &= \frac{1}{3c^{2/3}}(n-m) \\ &= \sqrt{2}\left(\frac{m}{c}\right)^{2/3}\left(\frac{\lfloor 3\sqrt{2}m^{2/3}\rfloor}{3\sqrt{2}m^{2/3}}\right). \end{aligned}$$

The result follows from this, since the last factor on the right clearly approaches 1 as $m \to \infty$, and because $m/n < m/c < 1$ and $m/n \to 1$ as $m \to \infty$.

A-3. Prove that any convex pentagon whose vertices (no three of which are collinear) have integer coordinates must have area $\geq 5/2$.

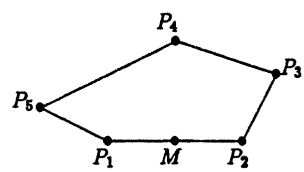
Solution. Two of the five vertices must have coordinates that are pairwise congruent modulo 2 (since there are only four classes of integer points modulo 2: $(0,0), (0,1), (1,0)$, and $(1,1)$.) The mid-point, M, between these two points therefore also has integer coordinates. Since the pentagon is convex, M is either in the interior or on the boundary.

Case 1. M is in the interior.



In this case there are 5 triangles formed by M and the sides of the pentagon. By the standard formula for the area of a triangle in terms of the coordinates of its vertices, we see that each of these triangles has area $\geq 1/2$. Hence the total area is $\geq 5/2$.

Case 2. M is on the boundary.



Since the pentagon is convex, the segment $\overline{P_1P_2}$ cannot be parallel to both $\overline{P_4P_5}$ and $\overline{P_3P_4}$. Say it is not parallel to $\overline{P_3P_4}$. Then $\triangle P_2P_3P_4, \triangle MP_3P_4$, and $\triangle P_1P_3P_4$ must all have distinct areas since their heights relative to $\overline{P_3P_4}$ form a strictly monotonic sequence. Thus one of these areas is $\geq 3/2$, say $\triangle P_1P_3P_4$. This area plus the remaining two triangles made by the sides yield total area $\geq 5/2$.

A-4. Consider a paper punch that can be centered at any point of the plane and that, when operated, removes from the plane precisely those points whose distance from the center is irrational. How many punches are needed to remove every point?

Solution. It's easy to see that two punches are not enough. Many examples can be given to show that three punches suffice. For example, punch at $A(0,0), B(d,0), C(2d,0)$, where d is yet to be determined. Let P be any point in the plane, and let a,b,c denote distances PA, PB, PC respectively. By appropriately applying the Law of Cosines to triangles ABP and BCP and adding, we find that $a^2+c^2 = 2d^2+2d^2$, or equivalently, $a^2-2b^2+c^2 = 2d^2$. Thus, if $d=\sqrt[4]{2}$, the right side of the last equation is $2\sqrt{2}$, and it follows that not all of a,b,c are rational.

A-5. If **A** and **B** are square matrices of the same size such that **ABAB** = **0**, does it follow that **BABA** = **0** ?

Solution. The answer is yes for 2×2 matrices and no in all other cases. Indeed: since $\mathbf{ABAB}=\mathbf{0}$, it follows that $\mathbf{B}(\mathbf{ABAB})\mathbf{A}=\mathbf{0}$, and hence that **BA** is nilpotent. If a 2×2 matrix **M** is nilpotent, then $\mathbf{M}^2=\mathbf{0}$ (because the characteristic equation of **M** has degree 2 or less).

A counterexample for 3×3 (and therefore, just by enlargement by 0's, for any size) is to take

$$\mathbf{A}=\begin{pmatrix}0&0&1\\0&0&0\\0&1&0\end{pmatrix},\quad \mathbf{B}=\begin{pmatrix}0&0&1\\1&0&0\\0&0&0\end{pmatrix}.$$

A-6. If X is a finite set, let $|X|$ denote the number of elements in X. Call an ordered pair (S,T) of subsets of $\{1,2,\ldots,n\}$ admissible if $s>|T|$ for each $s\in S$, and $t>|S|$ for each $t\in T$. How many admissible ordered pairs of subsets of $\{1,2,\ldots,10\}$ are there? Prove your answer.

Solution. Let A_n denote the number of admissible ordered pairs of subsets of $\{1,2,\ldots,n\}$. It follows that

$$A_n=\sum_{i=0}^{n}\sum_{j=0}^{n}\binom{n-i}{j}\binom{n-j}{i}$$
$$=\sum_{i=0}^{n}\sum_{j=0}^{n-i}\binom{n-i}{j}\binom{n-j}{i}$$

Let $k=i+j$. Note that the conditions of the problem imply that $i+j\le n$ because the smallest values the i elements from S can be is $j+1, j+2,\ldots$. Thus, we have

$$A_n=\sum_{i=0}^{n}\sum_{k=i}^{n}\binom{n-i}{k-i}\binom{n+i-k}{i},$$

which, on changing order of summation, gives

$$A_n=\sum_{k=0}^{n}\sum_{i=0}^{k}\binom{n+i-k}{i}\binom{n-i}{k-i}$$
$$=\sum_{k=0}^{n}\binom{2n-k+1}{k}$$

the last step by a known identity, or from a combinatorial argument. This last sum is equal to F_{2n+2}, the $(2n+2)$-nd Fibonacci number defined by $F_0=0, F_1=1$ and $F_n=F_{n-1}+F_{n-2}$ for $n\ge 2$. In our problem, $A_{10}=F_{22}=17711$.

B-1. Find all real-valued continuously differentiable functions f on the real line such that for all x

$$(f(x))^2=\int_0^x\left[(f(t))^2+(f'(t))^2\right]dt+1990.$$

Solution. There are two such functions, namely $f(x)=\sqrt{1990}\,e^x$, and $f(x)=-\sqrt{1990}\,e^x$.

To see this, suppose that the identity holds. Differentiating each side gives

$$2f(x)f'(x)=(f(x))^2+(f'(x))^2,$$

or equivalently,

$$(f(x)-f'(x))^2=0,$$
$$f'(x)=f(x),$$
$$\log|f(x)|=x+C,$$
$$|f(x)|=e^Ce^x.$$

But f is continuous and $f(0)=\pm\sqrt{1990}$, implying that $f(x)=\pm\sqrt{1990}e^x$.

B-2. Prove that for $|x|<1, |z|>1$,

$$1+\sum_{j=1}^{\infty}(1+x^j)P_j=0,$$

where P_j is

$$\frac{(1-z)(1-zx)(1-zx^2)\cdots(1-zx^{j-1})}{(z-x)(z-x^2)(x-x^3)\cdots(z-x^j)}.$$

Solution. Let $S_0=1$, and for $n\ge 1$, let

$$S_n=1+\sum_{j=1}^{n}(1+x^j)P_j$$

It is easy to check that $S_1=\dfrac{1-zx}{z-x}$, $S_2=\dfrac{(1-zx)(1-zx^2)}{(z-x)(z-x^2)}$, and by induction,

$$S_n = \frac{(1-zx)(1-zx^2)\cdots(1-zx^n)}{(z-x)(z-x^2)\cdots(z-x^n)}.$$

To complete the proof, we need to prove that $\lim_{n\to\infty} S_n = 0$. To see this, we note that

$$S_{n+1} = \left(\frac{1-zx^{n+1}}{z-x^{n+1}}\right) S_n.$$

As $n \to \infty$, $1-zx^{n+1}$ goes to 1 and $z-x^{n+1}$ goes to z. Thus, there exist positive numbers N and ε such that

$$\left|\frac{1-zx^{n+1}}{z-x^{n+1}}\right| < \frac{1}{|z|} + \varepsilon < 1$$

for all integers $n > N$. It follows that

$$|S_{n+1}| < \left(\frac{1}{|z|} + \varepsilon\right) |S_n|$$

and therefore

$$|S_{n+k}| < \left(\frac{1}{|z|} + \varepsilon\right)^k |S_n|$$

for all positive integers k. The result follows.

B-3. Let S be a set of 2×2 integer matrices whose entries a_{ij} (1) are all squares of integers, and, (2) satisfy $a_{ij} \le 200$. Show that if S has more than 50387 $(= 15^4 - 15^2 - 15 + 2)$ elements, then it has two elements that commute.

Solution. Let U be all such 2×2 matrices, D the diagonal ones, R those whose only nonzero element is in the upper-right corner, and L those whose only nonzero element is in the lower-left corner. Note that any two matrices from D commute, any two from R commute, and any two from L commute. There are 15^2 matrices in D, and 14 matrices in each of R and L. The three sets D, R, L, are mutually disjoint. Thus, there are $15^4 - 15^2 - 28$ elements in $U - (D \cup R \cup L)$, so if 4 more are added to this complement, at least two will be in D or R or L. That is, if we have more than $15^4 - 15^2 - 28 + 3 = 50375$ (< 50387) then two of them will commute.

The number 50387 is far from the best possible and there are many potential solutions. *Jiuqiang Liu* and *Allen J. Schwenk* from *Western Michigan University* have shown, using an inclusion-exclusion argument, that the maximum number of elements in U in which no two elements commute is 32390.

B-4. Let G be a finite group of order n generated by a and b. Prove or disprove: there is a sequence

$$g_1, g_2, g_3, \ldots, g_{2n}$$

such that

(1) every element of G occurs exactly twice, and

(2) g_{i+1} equals $g_i a$ or $g_i b$, for $i = 1, 2, \ldots, 2n$. (Interpret g_{2n+1} as g_1.)

Solution. Construct a directed graph whose vertices are labeled by the elements of G, so that for each vertex g, there are two "out" arcs, one to vertex ga and one to vertex gb (and consequently, each vertex g has two "in" arcs coming to it, one from ga^{-1} and one from gb^{-1}). The resulting graph is connected and each vertex has outdegree 2 and indegree 2. Therefore there is an Eulerian path which traverses the arcs, once and only once, and returns to the beginning. We get the desired sequence by listing the group elements associated with the vertices as we follow this path.

B-5. Is there an infinite sequence $a_0, a_1, a_2, \ldots$ of nonzero real numbers such that for

$n = 1, 2, 3, \ldots$ the polynomial

$$p_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

has exactly n distinct real roots?

Solution. Take $a_0 = 1, a_1 = -1$, and proceed by induction. Say $p_n(x)$ has the property, and also $p_n(x) \to \infty$ or $-\infty$ as $x \to \infty$ depending upon whether n is even or odd. Then

$$p_{n+1}(x) \equiv p_n(x) + \frac{(-x)^{n+1}}{M}$$

has a sign change arbitrarily close to every root of $p_n(x)$ for M sufficiently large, and also the same sign as $p_n(x)$ at $x_n^* + 1$ where x_n^* is the largest root of $p_n(x)$. But now (M is already fixed) for x sufficiently large $p_{n+1}(x)$ has another sign change. Since $p_{n+1}(x)$ has at most $n+1$ roots, the result follows.

B-6. Let S be a nonempty closed bounded convex set in the plane. Let K be a line and t a positive number. Let L_1 and L_2 be support lines for S parallel to K, and let $\overline{L}$ be the line parallel to K and midway between L_1 and L_2. Let $B_S(K,t)$ be the band of points whose distance from $\overline{L}$ is at most $(t/2)w$, where w

is the distance between L_1 and L_2. What is the smallest t such that

$$S \cap \bigcap_K B_S(K,t) \neq \emptyset$$

for all S? (K runs over all lines in the plane.)

Solution. (*taken from a student paper*) The smallest t is $t = 1/3$.

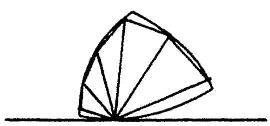
To show that smaller t will not do, examine the triangle with vertices $(0,0),(0,2),(2,0)$, and support lines of the form $x + y = k, x = k, y = k$.

In the three cases it is easy to see that $\overline{L}$ is given by $x + y = 1, x = 1$, and $y = 1$ respectively, and $B_S(K,t)$ in each case is $|x+y-1| \leq t, |x-1| \leq t, |y-1| \leq t$. If $t < 1/3$ and $(x,y) \in S \cap \bigcap_K B_S(K,t)$, then $y \geq 1 - t > 1 - 1/3 = 2/3$, $x \geq 1 - t > 1 - 1/3 = 2/3$, so $x + y - 1 > 1/3 > t$, and this contradicts $|x + y - 1| \leq t$.

We now show that $t = 1/3$ suffices by showing that the centroid of S is in $B_S(K,t)$ for all K, so $S \cap \bigcap_K B_S(K, 1/3) \neq \emptyset$.

Proof. Rotate, rescale, and translate the plane so that L_2 lies on the x-axis and L_1 lies along the line $y = 3$, and S touches L_2 at the point P. We show that the centroid (x,y) satisfies $y \leq 2$ (and then by symmetry it must satisfy $y \geq 1$).

Let A be the area of S. Cover S with triangles with one vertex at P, such that at most ϵA remains uncovered.



P

If a triangle has vertex y-coordinates $0, y_1, y_2$ $(0 \leq y_1, y_2 \leq 3)$ then the centroid of that triangle has y-coordinate ≤ 2. Thus each triangle has centroid of height at most 2 from the x-axis. Let $\overline{y}_\Delta$ be the y-coordinate of the centroid of the triangle-tiled portion of S, $(\overline{y}_\Delta \leq 2)$, and $\overline{y}_\epsilon$ the y-coordinate of the centroid of the remainder, and $\overline{y}_S$ the y-coordinate of the centroid of S. Then

$$\overline{y}_S A = \epsilon A \overline{y}_\epsilon + (1-\epsilon) A \overline{y}_\Delta,$$
$$0 \leq \overline{y}_S \leq \epsilon 3 + (1-\epsilon)2 = 2 + \epsilon,$$

Since $\overline{y}_\Delta \leq 2 + \epsilon$ for all $\epsilon > 0$, it must be the case that $\overline{y}_S \leq 2$, and the proof is complete.

WINNING TEAMS

Teams from 289 institutions participated in the 1990 competition. The top five winning teams, in descending rank, are:

HARVARD UNIVERSITY: Jordan S. Ellenberg, Raymond M. Sidney, Eric K. Wepsic

DUKE UNIVERSITY: Jeanne A. Nielsen, Will A. Schneeberger, Jeffrey M. Vanderkam

UNIVERSITY OF WATERLOO: Dorian Birsan, Daniel R. L. Brown, Colin M. Springer

YALE UNIVERSITY: Thomas Zuwei Feng, Andrew H. Kresch, Zhaoliang Zhu

WASHINGTON UNIVERSITY: William Chen, Adam M. Costello, Jordan A. Samuels

PUTNAM FELLOWS

A total of 2347 students participated. The five highest-ranking individuals, in alphabetical order, are:

JORDAN S. ELLENBERG, Harvard University

JORDAN LAMPE, University of California, Berkeley

RAYMOND M. SIDNEY, Harvard University

RAVI D. VAKIL, University of Toronto

ERIC K. WEPSIC, Harvard University

LETTER TO THE EDITOR

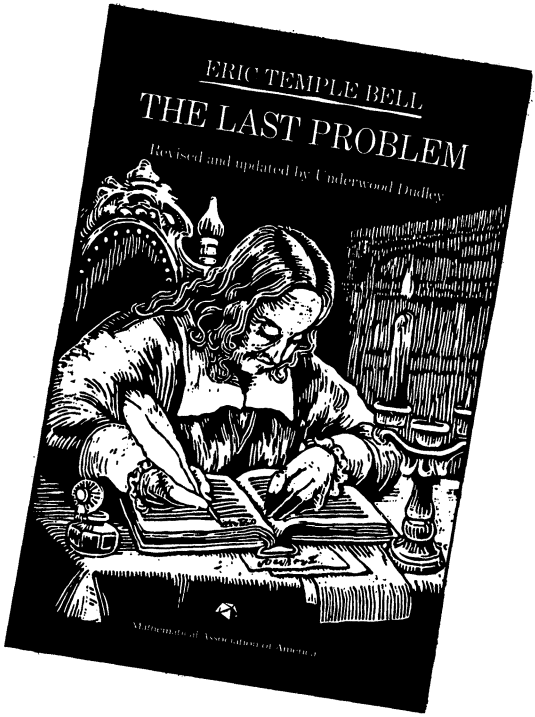
Dear Editor:

In the note "Why December 21 is the Longest Day of the Year," December 1990, Stan Wagon has the analemma in Figure 8 flipped horizontally from its correct representation. The position of the sun "in the sky" at noon should be farthest to the left, not right (as asserted in the caption), on Feb. 9. This correction is evident from the rest of the text and Reference 2.

Stephen J. Whitsitt
Manhattan Beach, CA

In the figure the legend should have been on the left side of the analemma as opposed to the right, and in the caption the word "right" should have been replaced by "left." By the way, you cannot see the asymmetry in the analemma, but look at Figure 2. If you draw a horizontal line through y=0, you will see the small max and min are not symmetric about this line. This corresponds to some asymmetry in the small loop of the analemma. There also is some asymmetry in the large loop, but it is less.—Stan Wagon

Revised and Updated



THE LAST PROBLEM

E. T. Bell

Revised and updated by Underwood Dudley

What Eric Temple Bell calls the last problem is the problem of showing that Pierre Fermat was not mistaken when he wrote in the margin of a book, almost 350 years ago, that $x^n + y^n = z^n$ has no solution in positive integers when $n \geq 3$. The original text of THE LAST PROBLEM traced the problem from Babylonia in 2000 B.C. to seventeenth-century France. Along the way we learn quite a bit about history, and just as much about mathematics. Underwood Dudley's notes bring us up-to-date on recent attempts to solve the problem.

The book is unique in that it is a biography of a famous problem. The book fits no categories. It is not a book of mathematics. Pages go by without an equation appearing. It is not a history of number theory because it includes too much about the history of the western world, and it is not a history of western civilization because its focus is on mathematics. It is too entertaining to be scholarly and contains too much mathematics to be widely popular. It is an unusual book.

What T.A.A. Broadbent said about Bell's work applies to THE LAST PROBLEM.

His style is clear and exuberant, his opinions, whether we agree with them or not, are expressed forcefully, often with humor and a little gentle malice. He was no uncritical hero-worshipper, being as quick to mark the opportunity lost as the ground gained, so that from his books we get a vision of mathematics as a high activity of the questing human mind, often fallible, but always pressing on the neverending search for mathematical truth.

This is a rich and varied, wide-ranging book, written with force and vigor by someone with a distinctive style and point of view. It will provide hours of enjoyable reading for anyone interested in mathematics.

328 pp., Paperbound, 1990

ISBN-0-88385-451-1

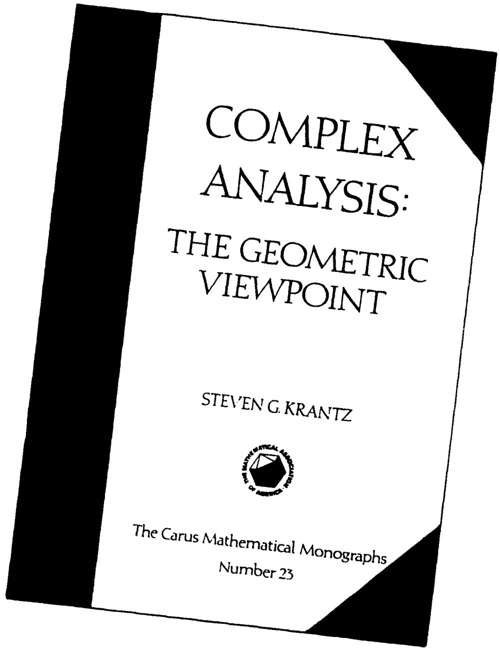
List: $17.50 MAA Member: $13.50

Catalog Number TLP

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

New From Steven G. Krantz



COMPLEX ANALYSIS:
THE GEOMETRIC VIEWPOINT

Steven G. Krantz

Geometric methods have been used in complex analysis since the 1930s when Lars Ahlfors discovered that they give a nice way to look at the Schwarz lemma. Since that time they have become a central part of the research activities of complex analysis. However these important techniques have never found their way into a text accessible to a broad audience.

Steven G. Krantz, a leading worker in complex analysis and a well-known mathematical expositor, has written the first book explaining how complex analysis can be studied using methods of geometry. Assuming no background in Riemannian geometry, and only one semester of complex analysis, Krantz explains the role of Hermitian metrics and of curvature in understanding the Schwarz lemma, normal families, Picard's theorems, conformal mappings, and many other topics. A minimum of geometric formalism is used to gain a maximum of geometric and analytic insight. The climax of the book is an introduction to several complex variables from the geometric viewpoint. Poincaré's theorem, that the ball and bidisc are biholomorphically inequivalent, is discussed and proved.

Except for the minimal background requirements, the book is self-contained. A review of relevant topics in the classical theory of one complex variable is provided. The style is light and inviting. The book is a must for anyone with an interest in complex analysis. Take a glance at the main chapter headings and order your copy today

- Principal Ideas of Classical Function Theory,
- Basic Notions of Differential Geometry,
- Curvature and Applications,
- Some New Invariant Metrics,
- A Glimpse of Several Complex Variables

210 pp., 1990, Hardbound,
ISBN 0-88385-026-5

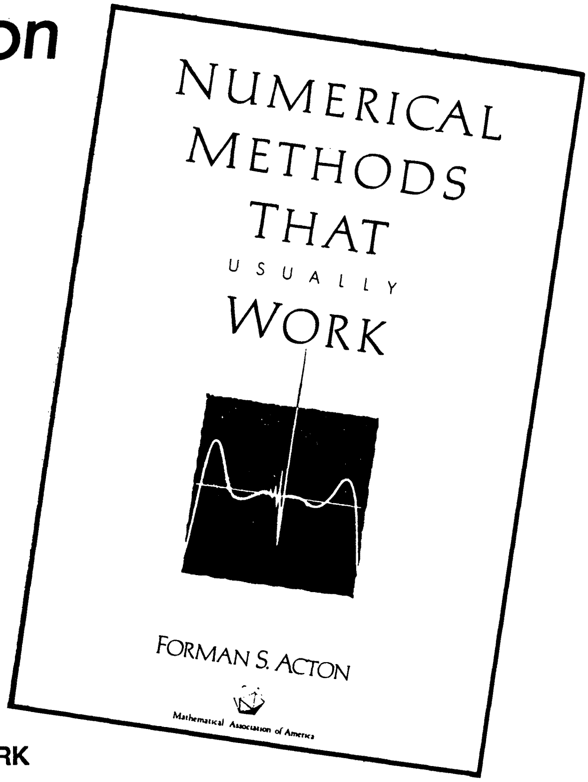
List $22.00 MAA Member $18.50

Catalog Number CAM-23

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

New From Forman S. Acton



NUMERICAL METHODS THAT WORK

Forman Acton

Numerical Methods That Work, originally published in 1970, has been reissued by the MAA with a new preface and some additional problems. Acton deals with a common sense approach to numerical algorithms for the solution of equations: algebraic, transcendental, and differential. It assumes that a computer is available for performing the bulk of the arithmetic. The book is divided into two parts, either of which could form the basis of a one-semester course in numerical methods. Part I discusses most of the standard techniques; roots of transcendental equations, roots of polynomials, eigenvalues of symmetric matrices, and so on. Part II cuts across the basic tools, stressing such common problems as instabilities in extrapolation, removal of singularities, and loss of significant figures. The book is written with clarity and precision, intended for practical rather than theoretical use.

An eminently readable and very well movitivated introductory text.
Mathematics of Computation

A first-rate book which can be used either as a text or reference.
Choice

560 pp., Paperbound, 1990
ISBN 0-88385-450-3

List $16.50 MAA Member $13.00

Catalog Number NMTW

ORDER FROM

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D. C. 20036

CONTENTS

ARTICLES

75 Plane Curves, Polar Coordinates, and Winding Numbers, *by John A. Baker.*

NOTES

91 Limerick, *by Arthur White.*

92 Packing a Box with Bricks, *by Charles H. Jepsen.*

97 Proof without Words: $\sin(x+y)=\sin x \cos y + \cos x \sin y$ for $x+y<\pi$, *by Sidney H. Kung.*

98 Complements of Linear Subspaces, *by Daryl Tingley.*

103 Proof without Words: The Sum of Odd Numbers, *by Jenő Lehel.*

104 Tetrahedral Numbers as Sums of Square Numbers, *by S. C. Althoen and C. B. Lacampagne.*

109 Regular Polygons with Rational Area or Perimeter, *by R. B. Killgrove and D. W. Koster.*

114 Proof without Words, *by Dave Logothetti.*

115 Some Counting Problems Involving the Multinomial Expansion, *by Fred Dodd and Rhodes Peele.*

122 Volumes of Cones, Paraboloids, and Other "Vertex Solids," *by Paul B. Massell.*

124 The Complementary Function and the General Solution, *by P. Ramankutty.*

130 Proof without Words: Integration by Parts, *by Roger B. Nelsen.*

PROBLEMS

131 Proposals 1368–1372.

132 Quickies 775–777.

133 Solutions 1343–1347.

137 Answers 775–777.

138 Proof without Words: Algebraic Areas, *by Sam Pooley and K. Ann Drude.*

REVIEWS

139 Reviews of recent books and expository articles.

NEWS AND LETTERS

141 1990 Putnam Competition.

144 Letter to the Editor.

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