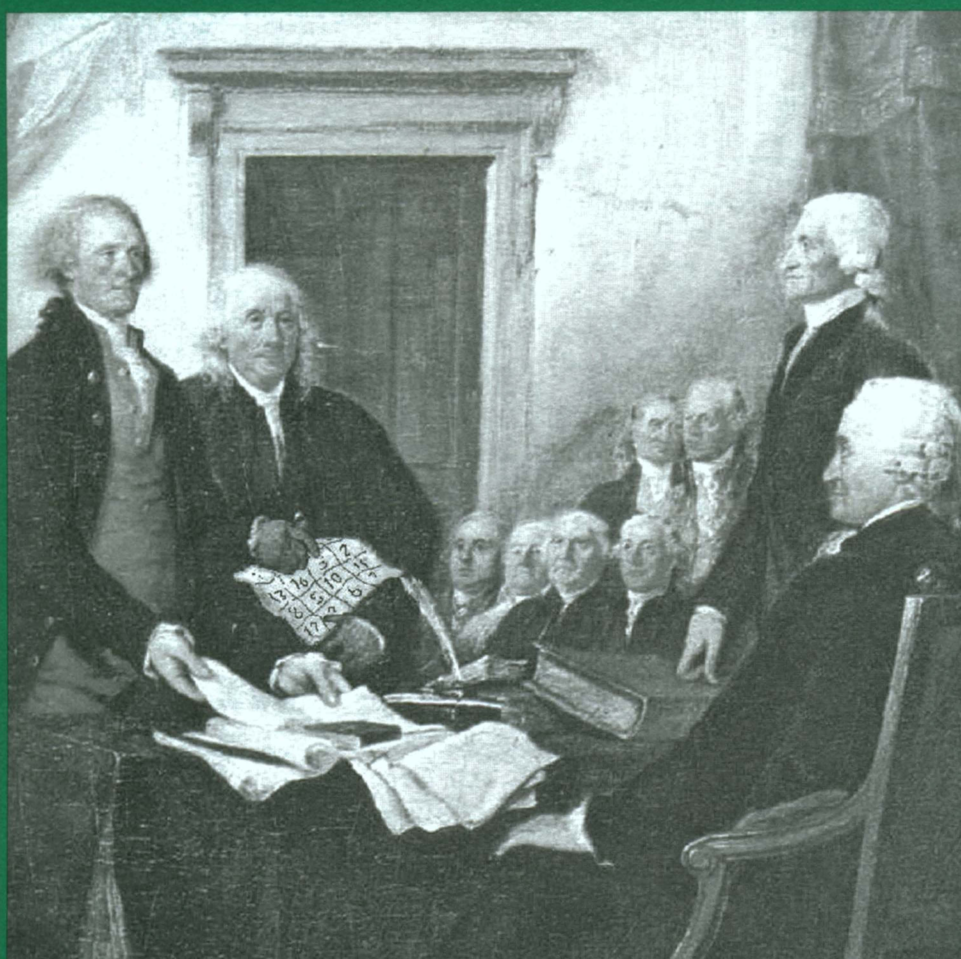
Vol. 79, No. 1, February 2006

MATHEMATICS MAGAZINE



- A Bent for Magic
- Packing Boxes with Bricks
- A New Model for Ribbons in $\mathbb{R}^3$

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

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

A more detailed statement of author guidelines appears in this *Magazine*, Vol. 74, pp. 75–76, and is available from the Editor or at www.maa.org/pubs/mathmag.html. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, or published by another journal or publisher.

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

The cover image, courtesy of author Paul Pasles, provides a peek at how the founding fathers may have filled their spare time during breaks in Philadelphia.

AUTHORS

Paul C. Pasles is an associate professor in the Department of Mathematical Sciences at Villanova University. He holds a B.A. from the University of Pennsylvania, and he received M.A. and Ph.D. degrees from Temple University, where he studied modular forms and integrals. He is proud of the fact that 43% of all people who share his surname earn their living from mathematics. (14% are journalists, 29% are retired, and the rest are below the age of six.) At the latest MathFest, he led a short course in honor of Benjamin Franklin's tercentenary. Paul is an MAA Special Interest Group webmaster and a semiprofessional virtuoso on the greek *bouzouki*.

Richard Bower graduated with honors from the U.S. Naval Academy in 1999 with a B.S. in mathematics. His undergraduate research project on tilings and packings led to a publication in a combinatorial journal as well as the survey in this MAGAZINE. He is now a logistics officer in the Marine Corps. He enjoys marathons and hopes to attend graduate school someday.

T. S. Michael received his B.S. in mathematics from Caltech in 1983 and his Ph.D. in combinatorics from Wisconsin under Richard Brualdi in 1988. He has been on the faculty at the Naval Academy since 1990. Time at the pool allows him to enjoy his interest in tiling and swimming simultaneously.

Tom Farmer received his Ph.D. in mathematics at the University of Minnesota in 1976, working in classical Lie groups and Lie algebras. He has been on the faculty at Miami University in Oxford, Ohio ever since. Recent fascination with digital photography, and with the idea of using mathematics to manipulate photographic images, led to this study of ribbons.

Vol. 79, No. 1, February 2006

MATHEMATICS MAGAZINE

MATHEMATICS MAGAZINE (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.
The annual subscription price for *MATHEMATICS MAGAZINE* to an individual member of the Association is $131. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 20% dues discount for the first two years of membership.)

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

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

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

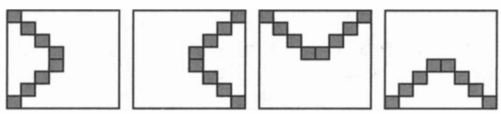
Postmaster: Send address changes to Membership/ Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036-1385.

Printed in the United States of America

A Bent for Magic

PAUL C. PASLES
Villanova University
Villanova, PA 19085
paul.pasles@villanova.edu

A *semi-magic square* of order n is an $n \times n$ matrix with constant sum along each of its n rows and n columns. When the two main diagonals also share that same sum with the rows and columns, the square is called *fully magic*. A *Franklin magic square*, on the other hand, is a semi-magic square whose four main "bent rows" share the same sum as the rows and columns, as in the figure below. Each of these three concepts (semi-, fully, and Franklin) has been abbreviated as *magic square* at various times by different authors, but the meaning is usually clear from context. Finally, a single row, column, diagonal or bent row of a square matrix is called *magic* if its sum is equal to $S = nc$, where c is the average of all entries in the matrix and n is the number of rows.



Of course things get trickier when we require that entries be distinct, in particular when these are the first n^2 positive integers. One could use instead some other set, like a subset of the primes, and still other variations are possible [**12, 14, 15**]. A silly but popular diversion long ago was to find all fully magic squares of consecutive integers for which the magic sum matched the current year. More serious research has tied magic squares to geometry, vector analysis, group theory, and experimental design.

The Franklin magic squares are named for Benjamin Franklin, who coined the term *bent row* and who was the first to intentionally incorporate bent rows into his squares. We address some questions raised by a recent historical article [**7**]: When are the various types of squares possible, how prevalent are they, and what are their mathematical properties? Are there any new clues as to how Ben Franklin went about his constructions? In the process we will consider his only known rough notes on the subject, just recently discovered.

With a little effort the current author concocted Franklin magic examples for $n = 4, 5, 6$ (see FIGURE 1, below). These are *natural* squares, that is, the entries are $1, 2, \ldots, n^2$. Two of them (a and c) are also *vertically symmetric*, meaning $a_{i,j} + a_{n+1-i,j}$ is constant. FIGURE 1(b), meanwhile, proves the existence of odd-order Franklin magic squares.

8	1	12	13
14	2	11	7
3	15	6	10
9	16	5	4

(a)

4	7	10	19	25
20	1	14	22	8
12	15	13	16	9
6	24	11	3	21
23	18	17	5	2

(b)

1	14	25	8	28	35
24	33	11	30	3	10
5	21	20	19	15	31
32	16	17	18	22	6
13	4	26	7	34	27
36	23	12	29	9	2

(c)

Figure 1 Some Franklin magic squares

Benjamin Franklin's own magical creations far surpassed these simple examples in many ways. Some of his squares were largely hidden from public view until a recent paper reintroduced them [**7**]. That article focused on the history of these squares, and now we consider some of the mathematics behind them.

FIGURE 2 shows three of Ben Franklin's magic squares, of orders 6, 8, and 16. Other examples can be found at http://www.pasles.org/Franklin.html and in [**7**]. In FIGURE 2(c), a 16-square that was published on Franklin's behalf by an English acquaintance in 1767 (and which has appeared before in this MAGAZINE [**1**]), the chevron-shaped bent row is shown along with 3 of its 15 vertical translates. In general, if every bent row of a Franklin magic square can be translated in the direction of its vertex without changing its sum, the square is called *panfranklin*. That is, the bent rows shaped like $\vee$ and $\wedge$ can be shifted up and down, while those oriented as $>$ and $<$ can be shifted to the left and right, for a total of $2n$ bent rows all sharing the same sum. This is analogous to the *pandiagonal* magic square, wherein the two main diagonals of a fully magic square can be translated up or down to $2n$ positions without affecting the "magic." FIGURE 2(b) and (c) are panfranklin, since each $n \times n$ example has $4n$ bent rows that sum equally; other properties are considered in separate articles [**3, 5, 6, 7**]. FIGURE 2(a) is not panfranklin, because some of the $4n$ bent rows fail to be magic.

We will investigate the structure of the Franklin magic and panfranklin squares, further illuminating the work of the master. We begin by showing that his examples were the best ones possible!

2	9	4	29	36	31
34	32	30	7	5	3
6	1	8	33	28	35
20	27	22	11	18	13
25	23	21	16	14	12
24	19	26	15	10	17

(a)

17	47	30	36	21	43	26	40
32	34	19	45	28	38	23	41
33	31	46	20	37	27	42	24
48	18	35	29	44	22	39	25
49	15	62	4	53	11	58	8
64	2	51	13	60	6	55	9
1	63	14	52	5	59	10	56
16	50	3	61	12	54	7	57

(b)

200	217	232	249	8	25	40	57	72	89	104	121	136	153	168	185
58	39	26	7	250	231	218	199	186	167	154	135	122	103	90	71
198	219	230	251	6	27	38	59	70	91	102	123	134	155	166	187
60	37	28	5	252	229	220	197	188	165	156	133	124	101	92	69
201	216	233	248	9	24	41	56	73	88	105	120	137	152	169	184
55	42	23	10	247	234	215	202	183	170	151	138	119	106	87	74
203	214	235	246	11	22	43	54	75	86	107	118	139	150	171	182
53	44	21	12	245	236	213	204	181	172	149	140	117	108	85	76
205	212	237	244	13	20	45	52	77	84	109	116	141	148	173	180
51	46	19	14	243	238	211	206	179	174	147	142	115	110	83	78
207	210	239	242	15	18	47	50	79	82	111	114	143	146	175	178
49	48	17	16	241	240	209	208	177	176	145	144	113	112	81	80
196	221	228	253	4	29	36	61	68	93	100	125	132	157	164	189
62	35	30	3	254	227	222	195	190	163	158	131	126	99	94	67
194	223	226	255	2	31	34	63	66	95	98	127	130	159	162	191
64	33	32	1	256	225	224	193	192	161	160	129	128	97	96	65

(c)

Figure 2 Vintage 18th century magic squares with constant sum (or *magic sum*) of (a) 111 (b) 260 (c) 2056

The panfranklin squares

Fact Franklin himself actually found a *minimal* example, in the following sense: As far as we can tell, he was interested exclusively in squares of even order, and the smallest even order for which natural panfranklin squares exist is 8. Let's see why. Write A^H for the horizontal reflection of A, meaning that the columns of A have been reordered last-to-first; A^V for the vertical reflection; and A^T for the transpose. Note that $A^V = ((A^T)^H)^T$ and put $A^{HV} = (A^H)^V$. Finally, J_n denotes the $n \times n$ matrix of ones. Using this notation, we can show that for small orders, panfranklin squares can be decomposed into building blocks in such a way that all of the entries must repeat, and that they do so in a very regular and obvious fashion. Therefore no natural example is possible in those orders, and the reason can be seen with the naked eye.

THEOREM. *Let $2 \le n \le 6$ and let M be an $n \times n$ matrix whose average entry is c. M is panfranklin of order n if and only if there exists an $\lfloor n/2 \rfloor \times \lfloor n/2 \rfloor$ matrix A such that*

$$M - cJ_n = \left[\begin{array}{c|c} A & -A^H \\ \hline -A^V & A^{HV} \end{array}\right] \quad or \quad \left[\begin{array}{c|c|c} A & 0 & -A^H \\ \hline 0\cdots & \vdots & \cdots 0 \\ \hline -A^V & 0 & A^{HV} \end{array}\right],$$

depending on the parity of n. Thus the smallest even order in which panfranklin squares without repeated entries can exist is $n = 8$.

The "if" direction is easy to prove: since J_n and the partitioned matrix have the desired summatory properties, so does their linear combination M. The "only if" part is a little bit tougher. For $n = 2$ and 3 the proof is still a straightforward exercise by hand. For $n = 4$, 5, and 6 the panfranklin properties can be viewed as $6n$ equations in n^2 variables, where the variables are the entries of M. (Each equation has n nonzero coefficients, and those coefficients are all equal to 1.) The system of equations can be solved using technology.

An immediate consequence of the theorem is that for small orders, any panfranklin examples necessarily are vertically and horizontally symmetric. Thus every entry appears at least twice! This establishes the minimality of Franklin's results: since $n = 2$, 4, and 6 are not options, he had to start with $n = 8$ to get a panfranklin square that used the entries $1, 2, \ldots, n^2$.

The theorem's characterization fails in general for higher orders than 6—more precisely, the theorem holds in the "if" direction only—as can be seen by the panfranklin 7-square we have devised below. This, then, is the smallest order for which examples with a distinct entry set can be constructed. Whether *natural* examples exist in order 7 is unknown; if not, then Franklin's 8-squares are minimal in an even stronger sense!

430	541	411	512	658	102	342
387	275	619	215	130	688	682
519	435	257	664	576	444	101
381	441	440	428	129	702	475
324	443	574	479	305	390	481
482	559	262	354	701	190	448
473	302	433	344	497	480	467

By the theorem, then, Benjamin Franklin did find panfranklin squares of minimal even order. Was he aware of the fact? As with so much of what he did, we cannot know for sure. He may have known that smaller orders would not work, or there might have been some trial-and-error, or it may simply be a coincidence that he came up with examples of order 8. There is, however, one tantalizing bit of evidence to suggest that the theorem was indeed known to him, namely this: By altering just slightly the format in the theorem (a partitioned matrix added to the appropriate multiple of J_n), it is possible to describe every single one of his six known examples!

Incidentally, it also follows from the theorem that for orders up to 6, the matrix product of panfranklin squares is again panfranklin. Suppose that the order is even. Then we can multiply using the form given in the theorem, as follows.

$$\left\{\begin{bmatrix} A & -A^H \\ -A^V & A^{HV} \end{bmatrix} + kJ_{2p}\right\}\left\{\begin{bmatrix} B & -B^H \\ -B^V & B^{HV} \end{bmatrix} + \ell J_{2p}\right\}$$
$$= \begin{bmatrix} A & -A^H \\ -A^V & A^{HV} \end{bmatrix}\begin{bmatrix} B & -B^H \\ -B^V & B^{HV} \end{bmatrix} + \begin{bmatrix} A & -A^H \\ -A^V & A^{HV} \end{bmatrix}(\ell J_{2p}) +$$
$$(kJ_{2p})\begin{bmatrix} B & -B^H \\ -B^V & B^{HV} \end{bmatrix} + (kJ_{2p})(\ell J_{2p})$$
$$= \begin{bmatrix} AB + A^H B^V & -AB^H - A^H\left(B^H\right)^V \\ -A^V B - (A^V)^H B^V & A^V B^H + A^{HV} B^{HV} \end{bmatrix} + k\ell J_{2p}^2$$
$$= \begin{bmatrix} 2AB & -2AB^H \\ -2A^V B & 2A^V B^H \end{bmatrix} + 2k\ell p J_{2p}$$
$$= 2\begin{bmatrix} AB & -(AB)^H \\ -(AB)^V & (AB)^{HV} \end{bmatrix} + 2k\ell p J_{2p},$$

where we have used the identities $A^H B^V = AB$, $AB^H = (AB)^H$, and $A^V B = (AB)^V$, as well as the fact that $JC = CJ = O$ (the zero matrix) for any semi-magic C with zero sum. The proof for odd order is quite similar. ■

Here is another corollary, of an algebraic nature: If P_n is the vector space of real panfranklin squares, allowing repeated entries, then $P_2 \simeq P_3$ and $P_4 \simeq P_5$ but $P_6 \not\simeq P_7$. Observe also that for $2 \le n \le 6$, not only do we have the usual vector space structure common to every type of 'magic' square, but we have also proven closure under multiplication. Thus P_n is an associative algebra.

The case $n = 4$ as a model

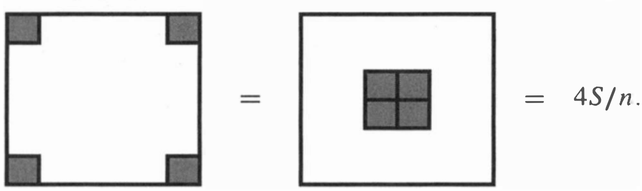
Franklin seems to have followed the problem-solver's dictum that one should approach a difficult problem by looking first at a special case. As Pólya wrote, one can use "the less difficult, less ambitious, special, auxiliary problem as a *stepping stone* in solving the more difficult, more ambitious, general, original problem." [**11**, p. 196, emphasis his.] This strategy had served the earlier square-makers well; to take just one example, the fully magic 3-square was easily generalized to produce a method which works in any odd order. Likewise, I believe that Franklin used the $n = 4$ case as a template.

Admittedly, there is little *direct* evidence to support that claim. Indeed, not a single completed example survives to indicate that Franklin ever drew a bent-row square of order 4. But there is considerable circumstantial evidence in favor of this hypothesis.

Of an 8-square [7], Franklin once wrote that the "four corner numbers, with the four middle numbers" add up to the magic sum. It's curious that he commented only on those particular eight cells (shaded in the diagram below) and not on the much stronger property of his 8-squares, that every 2×2 *block*, or square submatrix, is equal to half the magic sum. Of the 16-square in FIGURE 2(c), he observed likewise "that a four-square hole being cut in a piece of paper" placed atop it would show a 4×4 submatrix that possesses the magic sum [7]. Again, why did he not remark instead on its more general 2×2 block property, in this case that every such block totals to one-quarter the magic sum? We might conjecture that this is because not all of his squares satisfied the stronger conditions. Very few examples by Franklin do survive, but these include FIGURE 2(a), a magic square that may indicate what weaker properties he was willing to accept. Here is a bit of block magic common to *all* of Franklin's surviving examples:

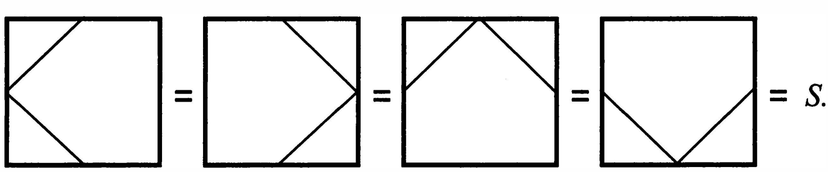
Property A. *An even order square, be it Franklin magic or fully magic, natural or not, is said to have Property A if the four corner cells and the four middle cells both sum to* $4S/n$*:*

$$\sum_{i,j\in\{1,n\}} a_{ij} = \sum_{2i,2j\in\{n,n+2\}} a_{ij} = 4S/n.$$



That is, each set of four cells is proportionally magic: its sum is equal to $4c$, where c is the average value of all entries in the matrix. Another commonly satisfied condition involves bent rows:

Property B. *A Franklin magic square of order n (natural or not, with n even or odd) is said to have Property B if these four parallel bent rows are magic:*



As with the main bent rows, one of these is redundant. That's because $B_4 = B_1 + B_2 - B_3$.

And now for the promised circumstantial evidence. We have claimed that the 4-square served as Franklin's model. Are there good reasons for believing so? Yes, here are seven of my favorites:

- Property A holds automatically in every Franklin (or fully) magic square of order 4, but it does not always hold in higher orders. Nevertheless, all of Franklin's own examples satisfy this condition. While most of his squares possess the stronger block property—every 2×2 block is proportionally magic—Franklin's 6- and 4-squares do not, yet Property A is still common to *all* of them. Indeed, even when the more impressive block property is present, Franklin only bothers to mention explicitly the weaker Property A. Clearly he thought it important. Perhaps this was an integral part of all of his standard techniques, though with such a small portion of the corpus surviving it is impossible to say for sure.

- Although Property B holds for every Franklin magic square of order 4, it is hardly automatic in higher orders, for example in FIGURE 1(b,c). Nevertheless, all of Franklin's own examples satisfy this condition, with a single exception that is not Franklin magic anyway. Indeed, the *only* magic parallel bent rows common to *every* Franklin square drawn by the man himself are the main four and the ones described in Property B. Again, it's as if the 4-case, of which we have no surviving Franklin magic examples, was a template on which he based all of his squares.
- Four is the smallest order for which Franklin magic squares with distinct entries exist, so their construction would be an obvious first step.
- As far as we can tell from the surviving papers, Franklin worked in even orders only.
- Most of Franklin's squares have an additional property: bent rows can be translated "against the grain." That is, one might shift the $\vee$-shaped and $\wedge$-shaped bent rows right or left (instead of up and down) by an *even* number of cells and still obtain the magic sum. (When a pattern is shifted over one edge of the square, you are to assume it continues from the opposite edge. In this sense, that magic *square* is really a *torus*.) This is true of 2(c), for example, and it also works for the transpose of (b). It is only absent from one of Franklin's bent-row squares, namely, 2(a). This strange attribute seems to be part and parcel of whatever general methods Franklin had at his disposal. And again, it is automatic in the 4-case; there, it's just another way to describe Property B. I have conjectured that such shift-invariance was motivated by the fact that Franklin also drew magic circles [**5, 6, 7**].
- Some of Franklin's magic squares can be partitioned into 4-squares that themselves satisfy most of the Franklin magic properties. Once again, a 4×4 matrix seems to have been in the back of his mind. Franklin's own comments bolster this argument; recall the "four-square hole."
- Franklin seems to have spent considerable time experimenting with squares of order 4 at various times in his life. The construction of one particular 4-square, he wrote in 1765, was an arduous task, the implication being that the creative process took some time [**7**]. And there is another instance of experimentation with the 4-case that I discovered recently—the image itself to be unveiled in my forthcoming book on Franklin and his squares. Aside from supporting the 4-square hypothesis, this new example (an incomplete rough sketch) also indicates that Benjamin Franklin was still writing magic squares much later in life than could be proven previously. (Also to be revealed: what really inspired him to draw squares in the first place.)

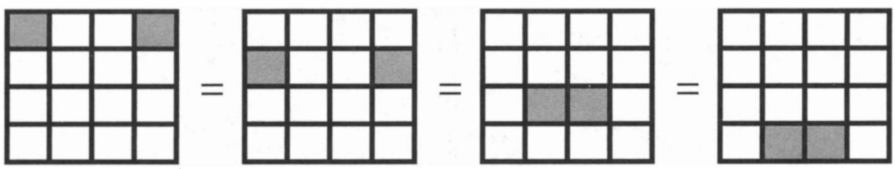
Though Franklin's general method remains a mystery, the $n = 4$ case was truly a "stepping stone" to higher orders. Suffice to say, order 4 seems to have been very important.

How hard is it to build a Franklin square? That question is ill-defined, but we should have better luck with a related one: How many Franklin squares are there of given order? In particular, are they more numerous than their fully magic brethren, or less so? The answer, in general, is neither. Intuitively, two diagonal conditions should impose less restriction than three (independent) bent-row conditions, but as we shall see, that isn't necessarily the case.

Order four: combinatorics and algebra

The $n = 4$ case is notable for other reasons as well.

Property C. *A square matrix of order 4 is said to have Property C if the four cell-pairs below share the same sum.*



Every Franklin magic square of order 4, natural or not, automatically has Property C. This can be seen by comparing the four straight rows with the top and bottom bent rows, and then applying Property A. As an amusing exercise, you can prove our first theorem for the case $n = 4$ using Property C.

One property oft-prized by mathemagicians is the quality of being *associated*: The sum of the (i, j) and $(n + 1 - i, n + 1 - j)$ cells is constant. Natural Franklin magic 4-squares are never associated, but they do display pretty symmetries of another sort. These come in three flavors.

Trichotomy law. *Every natural Franklin 4-square displays one of the symmetries shown in* FIGURE 3, *allowing for transposition. Here x^c denotes the complement of x, namely, $x^c = 17 - x$.*

a	b	c	d
e	f	g	h
e^c	f^c	g^c	h^c
a^c	b^c	c^c	d^c

Type I: vertically symmetric

a	b	c	d
e	f	g	h
a^c	b^c	c^c	d^c
e^c	f^c	g^c	h^c

Type II: shift-symmetric

a	b	c	d
b^c	a^c	d^c	c^c
f^c	e^c	h^c	g^c
e	f	g	h

Type III: fully magic

Figure 3 Symmetries of the natural Franklin magic squares of order 4

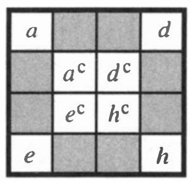
This distinction into types has some algebraic significance. Interestingly, in higher orders such symmetries are still possible, but no longer necessary. Trichotomy makes it easier to count 4-squares, though the mathochist may still prefer to generate them via brutish force.

Proof of the Trichotomy Law We may assume that a_{11}, a_{12}, or a_{22} is equal to 1, where a_{ij} denotes the i, jth entry of the matrix. The remaining possibilities can be produced from these by rotation or reflection. Define the transformations ρ, σ as follows.

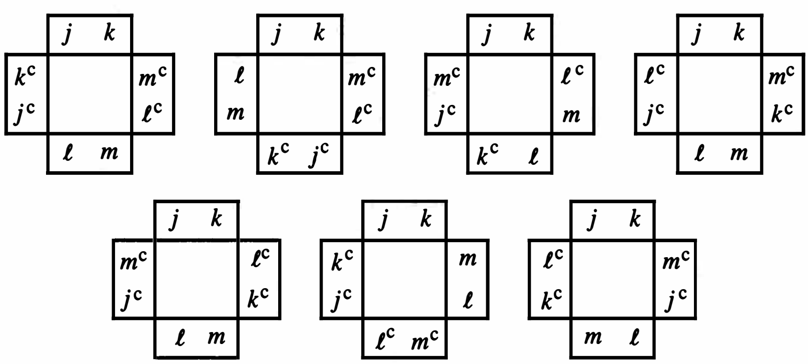
ρ = interchange first two columns, then interchange last two columns

σ = the commutator of ρ and τ, where τ is the ordinary transpose: $\sigma = \rho\tau\rho^{-1}\tau^{-1}$.

By Property B, we know that ρ, and therefore $\sigma = (\rho\tau)^2$, are bijections from the set of natural Franklin magic 4-squares onto itself, and that each map preserves type (I, II, or III). Then since ρ exchanges a_{11} with a_{12}, and σ exchanges a_{11} with a_{22}, we need only check that trichotomy holds when $a_{11} = 1$. Now, transposing if necessary, we can assume that $1^c = 16$ does not appear in the upper triangle of the matrix. If $16 = a_{22}$, then according to the bent row sums, $a_{11} + a_{22} = a_{23} + a_{14} = a_{33} + a_{44} = a_{32} + a_{41} = 17$, and so the matrix is also fully magic with each half-diagonal composed of a complementary pair.



That still leaves four complementary pairs to be placed somehow in the remaining eight spaces. However, no such pair can appear in the same row or column, else the remaining entries in that row or column would themselves be complementary (at odds with the placements already established). Under these restrictions, we claim that at least one of the remaining pairs must be diagonally adjacent (for example $a_{21}^c = a_{12}$). For, otherwise all four pairs would be placed in such a way as to avoid sharing a diagonal adjacency, a row, or a column, and while there are seven essentially different ways of doing so:



each of them leads to a contradiction by virtue of Property B. (For example, in the first diagram one would obtain

$$j + k^c + j^c + \ell = 34 = k + m^c + \ell^c + m,$$

and so $34 - k + \ell = 34 - \ell + k$, which results in a repeated entry, $\ell = k$.) This proves that at least one complementary pair must be diagonally adjacent; and therefore all four of them are, by Property B. Thus whenever $a_{22} = 16$, we have a magic square of Type III. If $16 = a_{41}$, on the other hand, then it is clear from columns, bent rows, and

Property A that the matrix is vertically symmetric, Type I. If $16 = a_{31}$, then Property C implies that

$$34 = (a_{31} + a_{34}) + (a_{32} + a_{33}) = (a_{31} + a_{34}) + (a_{11} + a_{14})$$
$$= 16 + a_{34} + 1 + a_{14} = 17 + (a_{14} + a_{34}),$$

so that $a_{14} + a_{34} = 17$; thus the first and last columns have shift-symmetry, and some tedious calculations give that the entire matrix is shift-symmetric, Type II. By similar machinations, one can verify that $16 \notin \{a_{21}, a_{32}, a_{33}, a_{42}, a_{43}, a_{44}\}$. ■

Thus, the question of enumeration comes down to the case where $a_{11} = 1$. With a simple computer program (or else pencil, paper, and a boring Saturday afternoon) you can use the trichotomy rule to find that there are 228 four-squares with $a_{11} = 1$, not counting reflections.

COROLLARY. *There are* 912 *natural Franklin magic squares of order* 4*, up to rotation and reflection.*

Here is one way to count them. Let A_{ij} be the set of natural Franklin magic 4-squares with $a_{ij} = 1$. We know that $|A_{11}| = 2(228)$, since each element of A_{11} has a transpose which is also in A_{11}. Likewise A_{22} is closed under transposition, while A_{12} is not. Note that ρ is a bijection from A_{12} onto A_{11}, and σ is a bijection from A_{22} onto A_{11}, so we have $|A_{11}| = |A_{12}| = |A_{22}|$. Thus if we do not count rotations and reflections, the number of natural Franklin magic squares of order 4 is

$$\tfrac{1}{2}|A_{11}| + |A_{12}| + \tfrac{1}{2}|A_{22}| = 2|A_{11}| = 4(228) = 912.$$

TABLE 1: Enumeration of natural squares (up to rotation and reflection)

Order	1	2	3	4	5	6
fully magic	1	0	1	880	275305224	unknown
Franklin magic	1	0	0	912	unknown	unknown
pandiagonal	1	0	0	48	3600	0
panfranklin	1	0	0	0	0	0

The first and third rows of this table are taken from *The On-Line Encyclopedia of Integer Sequences* [**13**].

Powerful new methods have been developed that provide an upper bound for an important subclass of the 8×8 panfranklin squares, but more precise results are still lacking.

The vertically symmetric squares, which appear naturally in the $n = 4$ case, were defined for arbitrary entry-sets by demanding that the sum of the (i, j) and $(n + 1 - i, j)$ cells be constant. Here is an algebraic result regarding Franklin squares of arbitrary order, generalizing an exercise that appeared in *Math Horizons* [**3**].

PROPOSITION. *Let* $n \geq 4$*. Consider the set of* $n \times n$ *Franklin magic squares with real entries, not necessarily distinct. The subset of vertically symmetric squares is a noncommutative ring under ordinary matrix addition and a product given by* $A * B := A(B^T)$*. Furthermore, the map* $A \longmapsto$ *(magic sum of* A*) is a surjective ring homomorphism.*

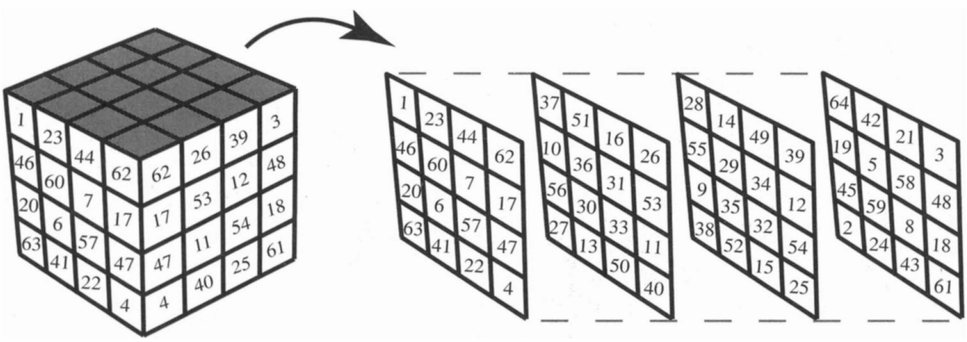
Thus the magic sum is both additive (like the trace) and multiplicative under $*$ (like the determinant). The proposition still holds if we replace $\mathbb{R}$ by any commutative ring.

A last word on the 4-case: While the trichotomy law does not hold for higher orders, it does illustrate the importance of placement of complementary pairs—and that is something Franklin exploited in every one of his squares. That's a little more evidence for the "4 as a model" hypothesis.

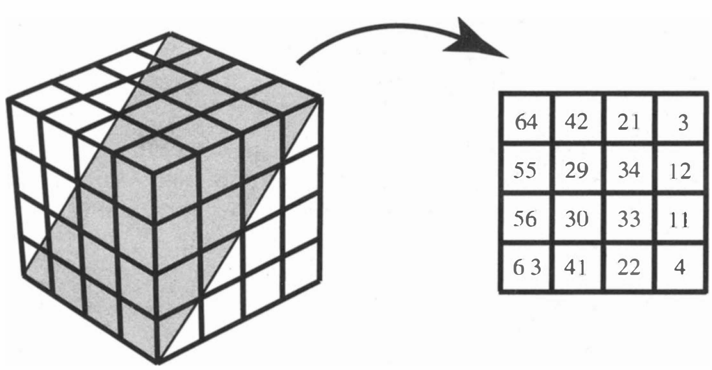
Journey to the third dimension

No article on the art of magic squares would be complete without a nod to the Cubist school.

When magicians tire of the plane, they turn to higher dimensions. How can we apply this approach to the current discussion? Shown below is a natural 4-cube in which each slice—from the 4 cross-sections shown to the 8 not pictured—is Franklin magic.



Indeed, there is a little extra magic, free with every such Franklin magic cube of order 4: One can even slice the cube diagonally (pictured below), and many of the resulting bent rows are guaranteed to sum magically. This is so despite the fact that this diagonal cross-section is not even nearly a magic square, for the rows work, but the columns do not.



The author believes this to be the first time anyone has bothered to construct a Franklin magic 4-cube. (Some partial successes for orders 4 and 6 have been published over the past century, but these were always deficient in various particulars, with some faces and other parallel slices lacking a bent row here or there.) In fact, many such examples are possible. For the interested novice: 4-cube constructions are made easier by good old Property C (diagram below).

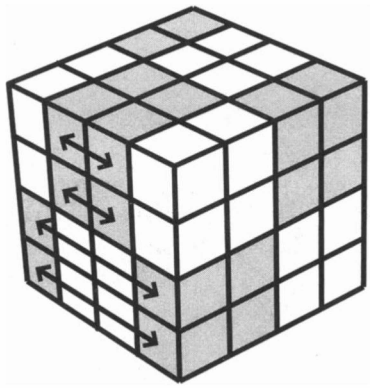


Figure 4 Some constant pair-sums that are automatic for any Franklin magic 4-cube

We have seen that there exist natural Franklin magic squares and cubes of order 4. What about tesseracts and other, higher-dimensional objects? Here is a conjecture, not for the faint of heart: If there exists a natural Franklin magic hypercube of order n and dimension k, then $n \geq 3 + \lfloor k/2 \rfloor$.

For investigators who seek to understand how Ben Franklin constructed his own magic squares we offer one final hint. Convert his 8- and 16-squares to octal and hexadecimal notation, respectively, then subtract 1 from each entry. Consider only the first (or second) digit of the resulting matrix. The results will surprise you!

REFERENCES

1. Elizabeth H. Agnew, Two problems on magic squares, this MAGAZINE **44** (1971), 13–15.
2. Paul C. Pasles, Benjamin Franklin's magic circle, `http://www.pasles.org/circle.html` (2001).
3. ———, Digging for squares, *Math Horizons* **9:4** (2002), 17–19.
4. ———, Every Magic Square for the Year, in preparation.
5. ———, Franklin's "Other" 8-Square, *J. Rec. Math.* **31** (2003), 161–166.
6. ———, Franklin Squares, `http://www.pasles.org/Franklin.html` (2000, 2001).
7. ———, The lost squares of Dr. Franklin: Ben Franklin's missing squares and the secret of the magic circle, *Amer. Math. Monthly* **108** (2001), 489–511.
8. ———, More magic squares, `http://www.pasles.org/magic.html` (2000, 2001).
9. ———, Benjamin Franklin, biographic entry in the *MacTutor History of Mathematics Archive*, `http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Franklin_Benjamin.html` (2001).
10. ———, Benjamin Franklin, Magician? *Franklin Gazette* **10:3** (Fall 2000), 4–8.
11. G. Pólya, *How to Solve It: A New Aspect of Mathematical Method*, 2nd edition, Princeton Univ. Pr., Princeton, 1971.
12. John P. Robertson, Magic squares of squares, this MAGAZINE **69** (1996), 289–293.
13. Neil J. A. Sloane, Editor-in-Chief, *The On-Line Encyclopedia of Integer Sequences*, `http://www.research.att.com/~njas/sequences/`. Sequences A006052, A027567.
14. Charles W. Trigg, Problem 882, this MAGAZINE **46** (1973), 285; solved **48** (1975), 54–55, 302–303.
15. ———, Problem 943, this MAGAZINE **48** (1975), 180; solved **49** (1975), 212–213.

Packing Boxes with Bricks

RICHARD J. BOWER
T. S. MICHAEL
United States Naval Academy
Annapolis, MD 21402
tsm@usna.edu

Brick-packing problems In a *packing problem* we must arrange a given collection of geometric objects in a nonoverlapping configuration to fill some larger object completely. Packing problems can be challenging even when only a few simple shapes are involved. For example, several innocuous-looking, but fiendish 3-dimensional packing problems were devised by J. H. Conway [**23**]. *Tiling problems* (2-dimensional packing problems) studied in the ancient world include tangrams in China and a conundrum of Archimedes, whose resolution merited a front page article in the *New York Times* in 2003 [**25**]. Jigsaw puzzles are familiar instances of more recent tiling problems. The packing problems most studied by mathematicians concern *polyominoes*—finite sets of rookwise-connected unit cells in an infinite chessboard—and their generalizations to higher dimensions.

In this article, we examine packings of rectangular boxes with rectangular bricks. Even in this basic case the problems that arise are interesting and difficult. We treat d-dimensional bricks and boxes, including those whose edge lengths are not integers, and answer the following questions as we introduce our packing theorems and constructions.

QUESTIONS. *Is it possible to*

(A) tile a 37×32 *rectangle with* 5×5, 3×3, *and* 2×2 *squares?*
(B) tile a 16×15 *rectangle with* 6×1 *rectangles?*
(C) tile a $\sqrt{4050} \times \sqrt{968}$ *rectangle with* $\sqrt{162} \times \sqrt{50}$ *rectangles?*
(D) pack a $12 \times 12 \times 11$ *box with* $4 \times 4 \times 4$ *and* $3 \times 3 \times 3$ *cubical bricks?*
(E) tile a 39×16 *rectangle with* 3×3 *and* 2×2 *squares?*
(F) pack a $720 \times 200 \times 100$ *box with* $15 \times 12 \times 10$ *bricks?*

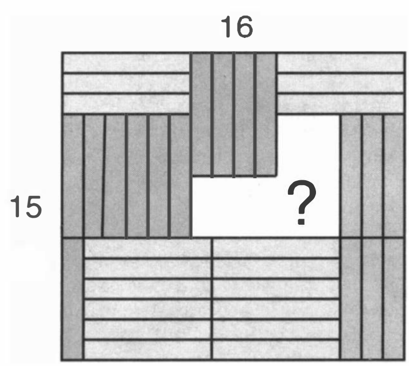


Figure 1 An attempt to answer Question (B)

To provide an affirmative answer to any of these questions we need only exhibit the requested packing or tiling. The reader has perhaps already found constructions for (A) and (D). Establishing a negative answer to a packing problem is often more difficult. FIGURE 1 shows an unsuccessful attempt to tile a 16×15 rectangle with 6×1 tiles. Further experimentation suggests that the tilings requested in (B), (C), and (E) do not exist, but it is not obvious how to confirm our suspicions without a brute-force, case-by-case search. Even a computer-search is daunting for (F), where 8000 bricks are to be placed. A central endeavor in the study of packing problems is the discovery of general nonexistence theorems that eliminate the need for brute-force attacks.

The following definitions are likely already clear from the preceding discussion. A d-dimensional *rectangular box* or *brick* of size $v_1 \times v_2 \times \cdots \times v_d$ is a set congruent to

$$\{(x_1, x_2, \dots, x_d) \in \mathbb{R}^d : 0 \leq x_i \leq v_i \text{ for } i = 1, 2, \dots, d\}.$$

An *integer brick* is one whose edge lengths v_1, v_2, ..., v_d are all integers. We use the term *pack* in the following sense: The union of the bricks must be the entire box, and the interiors of the bricks must be pairwise disjoint. In a packing, the edges of the bricks must be parallel to the edges of the box. In the 2-dimensional case we find it more natural to speak of rectangles that *tile* rectangles, and we sometimes refer to the bricks as *tiles.*

In Question (B), we tacitly assumed that we could rotate the 6×1 rectangles and place them in either orientation. However, it will be advantageous to forbid such rotations. We restrict our attention to packings with *translates* of a given set S of bricks. This is not a severe restriction as we may augment S to include the desired orientations of a particular brick. Here is the general problem we study.

BRICK-PACKING PROBLEM. *Which boxes can be packed by translates of a given set of d-dimensional bricks?*

We seek necessary and sufficient conditions for a packing, say, in terms of the edge lengths of the box and the given bricks. The Brick-Packing Problem is hopelessly difficult in full generality, and research has focused on the special cases involving a small number of integer bricks. In these situations the necessary and/or sufficient conditions for a packing usually hinge on divisibility conditions among the edge lengths of the bricks and box.

The partition-and-pack strategy Packings of a box with translates of one brick are readily characterized. We say that the $z_1 \times z_2 \times \cdots \times z_d$ box is a *multiple* of the $v_1 \times v_2 \times \cdots \times v_d$ brick provided z_i/v_i is an integer for $i = 1, 2, \dots, d$. The following observation is clear, as illustrated in FIGURE 2.

OBSERVATION. *The d-dimensional box R can be packed by translates of a given brick B if and only if R is a multiple of B. Moreover, any such packing is unique.*

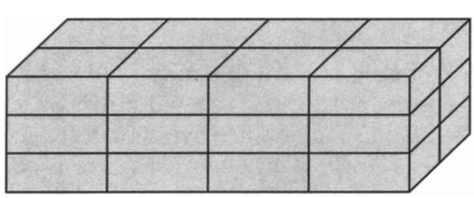


Figure 2 Packing a box with translates of one brick

Although the Brick-Packing Problem is not very interesting when we use translates of one brick, packings like the one in FIGURE 2 serve as the key building blocks when we solve more complex problems using a type of divide-and-conquer strategy:

PARTITION-AND-PACK STRATEGY. *To pack the box R with translates of some given bricks we partition R into sub-boxes, each of which is a multiple of a brick.*

We prefer to use a small number of sub-boxes in carrying out a partition-and-pack strategy.

Two themes and a preview Two themes run throughout this article. First, we show how elementary combinatorial arguments can be used to establish strong necessary conditions for prescribed packings. Second, we use the partition-and-pack strategy to obtain sufficient conditions for packings. In some fortuitous situations, the necessary and sufficient conditions coincide. We avoid the wide array of sophisticated algebraic techniques for solving packing problems **[1, 2, 9, 10, 30, 32, 33, 35]** in favor of more elementary combinatorial methods.

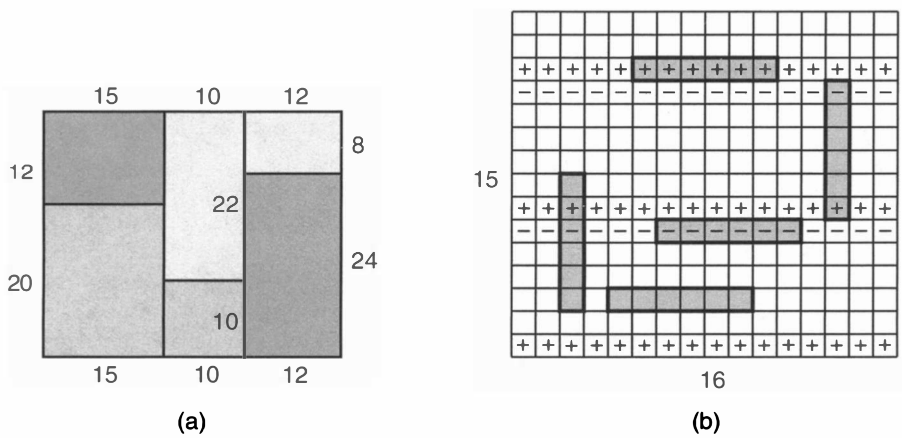


Figure 3 (a) Partition-and-pack strategy to tile a 37×32 rectangle. (b) Any 6×1 rectangle encloses a net charge of -6, 0, or $+6$

Let us illustrate our two themes by solving Questions (A) and (B). These solutions serve as a preview of much of the material to come. FIGURE 3(a) indicates a partition-and-pack scheme for the tiling requested in (A); each of the six sub-rectangles can be packed by a square of size 5×5, or 3×3, or 2×2. To prove that the tiling requested in (B) does not exist we consider the 16×15 rectangle in FIGURE 3(b) with unit positive and negative charges in some unit cells, as shown. Any 6×1 or 1×6 rectangle in a putative tiling encloses a net charge of -6, 0, or $+6$, and thus the net charge enclosed by all the rectangles would have to be a multiple of 6. However, the 16×15 rectangle encloses a net charge of $+16$, which is not a multiple of 6.

An outline In Part I we present some preliminary results and treat the 1-dimensional Brick-Packing Problem. In Part II we discuss and prove a recent theorem that completely solves the Brick-Packing Problem for two (not necessarily integer) bricks; this theorem answers Questions (B)-(E). Our exposition here is somewhat different from the one in **[5]**. Erdős and Straus considered two general problems about tiling large rectangles (with (A) and (E) as special cases)**[12]**. In Part III we discuss the context and history of these general problems and provide proofs for the theorems announced by Straus. In Part IV we consider packings of a d-dimensional box with just one brick (where all $d!$ orientations of the brick are allowed) and also with translates of cubes of different sizes. There is a type of duality between these two situations, which we illuminate with a fundamental theorem from combinatorial matrix theory.

PART I: Preliminaries and 1-dimensional packings

The 1-dimensional case of the Brick-Packing Problem deals with the packing of an interval by given intervals on the real line. The situation is familiar to schoolchildren who place colored rods end-to-end as a visual aid in learning how to add numbers. The conditions we obtain for 1-dimensional packings will serve as necessary conditions for results in higher dimensions.

Scaling We state an easy result without proof.

SCALING LEMMA. *Let r be a positive real number. Then there is a tiling of a $z_1 \times z_2$ rectangle with translates of $v_1 \times v_2$ and $w_1 \times w_2$ tiles if and only if there is a tiling of an $(rz_1) \times z_2$ rectangle with translates of $(rv_1) \times v_2$ and $(rw_1) \times w_2$ tiles.*

The number r represents a scaling factor applied to all horizontal lengths. There is a corresponding result for vertical scalings, as well as extensions to scalings in higher dimensions and packings involving more than two bricks. When we invoke any of these generalizations in our proofs, we simply refer to "a scaling argument."

We now answer Question (C).

EXAMPLE. *It is impossible to tile a $\sqrt{4050} \times \sqrt{968}$ rectangle with $\sqrt{162} \times \sqrt{50}$ rectangles (with both orientations allowed).*

Reason. We scale the edge lengths by a factor of $1/\sqrt{2}$ both horizontally and vertically and find that the requested tiling is equivalent to a tiling of a 45×22 rectangle with translates of 9×5 and 5×9 rectangular tiles. Assume there is such a tiling. Select an edge of length 22 and let c_1 and c_2 be the number of 9×5 and 5×9 tiles, respectively, along that edge. Then $22 = 5c_1 + 9c_2$. However, it is easy to check that this equation has no solution in nonnegative integers c_1 and c_2.

Counting We now generalize and formalize the counting argument in the example just given. Let the set of all nonnegative integer linear combinations of the real numbers $y_1, y_2, \ldots, y_t$ be denoted by

$$\langle y_1, y_2, \ldots, y_t \rangle = \{c_1 y_1 + c_2 y_2 + \cdots + c_t y_t : c_k = 0, 1, \ldots \text{ for } k = 1, \ldots, t\}.$$

Thus we may say that in the above example the requested tiling does not exist because 22 is not in the set $\langle 5, 9 \rangle$. More generally, we may count the bricks of each given type along one edge of a packed box and obtain the following necessary conditions for a packing.

COUNTING LEMMA. *Suppose that the box R is packed by translates of t bricks. Let z be the length of one edge of R, and let the corresponding edge lengths of the bricks be $y_1, y_2, \ldots, y_t$. Then z is in the set $\langle y_1, y_2, \ldots, y_t \rangle$.*

When we invoke this result, we simply cite "a counting condition." Counting conditions are not sufficient for a packing in general. For instance, 5 is in the set $\langle 2, 3 \rangle$, but a

5×5 square cannot be tiled by 2×2 and 3×3 squares. However, for 1-dimensional packings the counting conditions *are* sufficient, and we have the following answer to the problem of packing an interval with intervals.

THEOREM 1. *Let $v_1, v_2, \ldots, v_t$, and z be positive real numbers. An interval of length z can be packed by translates of intervals of length $v_1, v_2, \ldots, v_t$ if and only if z is in $\langle v_1, v_2, \ldots, v_t \rangle$.*

Proof. The counting lemma establishes necessity. Conversely, if $z = c_1 v_1 + c_2 v_2 + \cdots + c_t v_t$, then we may partition an interval of length z into sub-intervals of respective lengths $c_1 v_1, c_2 v_2, \ldots, c_t v_t$. Clearly, the kth sub-interval can be packed by c_k intervals of length v_k for $k = 1, 2, \ldots, t$, and thus the partition-and-pack strategy gives us the desired packing. ■

Intervals and integers: Frobenius and Sylvester Theorem 1 establishes the equivalence of the 1-dimensional Brick-Packing Problem and a problem in additive number theory, namely, determining whether the number z is in the set $\langle v_1, v_2, \ldots, v_t \rangle$. Throughout this section we assume that $v_1, v_2, \ldots, v_t$ are positive integers. In this case the set $\langle v_1, v_2, \ldots, v_t \rangle$ is the subject of a notoriously difficult problem in number theory that was popularized by Frobenius.

FROBENIUS STAMP PROBLEM. *Let $v_1, v_2, \ldots, v_t$ be positive integers with*

$$\gcd(v_1, v_2, \ldots, v_t) = 1.$$

Find the largest integer $n^* = n^*(v_1, v_2, \ldots, v_t)$ not in the set $\langle v_1, v_2, \ldots, v_t \rangle$.

If $v_1, v_2, \ldots, v_t$ are the available denominations of stamps, then n^* is the largest unachievable postage. It is not difficult to verify that $n^*(7, 3) = 11$ and $n^*(15, 10, 6) = 29$, for instance. The existence of the number n^* follows by induction on t and a consequence of the Euclidean algorithm: For any two positive integers v_1 and v_2 there are integers b_1 and b_2 such that $\gcd(v_1, v_2) = b_1 v_1 + b_2 v_2$. For $t \geq 3$ no general closed formula for n^* is known although several algorithmic schemes to compute n^* have been devised. An applet of Beukers computes n^* for $t \leq 4$ and moderate values of the arguments [**3**]. (We rely on this applet for some of our computations later.) The literature related to the Frobenius Stamp Problem and its variants is vast, and we refer the reader to the recent article by Owens as a starting point [**29**].

The Frobenius Stamp Problem may be recast in terms of 1-dimensional packings. Intervals with relatively prime lengths $v_1, v_2, \ldots, v_t$ pack all sufficiently large integer intervals, and we are requested to find the length n^* of the largest interval that cannot be packed. For $t = 2$ a result usually attributed to Sylvester [**34**] asserts that $n^* = v_1 v_2 - v_1 - v_2$ (proofs have appeared in the MAGAZINE [**28, 29**]) and gives us the following 1-dimensional packing theorem.

SYLVESTER'S THEOREM. *If v_1 and v_2 are relatively prime, then each interval of length greater than $v_1 v_2 - v_1 - v_2$ can be packed by intervals of length v_1 and v_2.*

Thus all sufficiently large integer intervals can be packed by two given intervals whose lengths are relatively prime. We often have to settle for just such an *asymptotic packing theorem*—a result that holds only when the edge lengths of the box to be packed are sufficiently large. We say that a set S of d-dimensional integer bricks *packs all sufficiently large boxes* provided there is an integer N such that every integer box whose edge lengths are all strictly greater than N can be packed by translates of bricks from S. We let $N^* = N^*(S)$ denote the smallest such N.

PART II: The two bricks theorem

We now characterize the boxes that can be packed by translates of two given bricks, solving the Brick-Packing Problem for two bricks. We do not require that the bricks and the box have integer edge lengths.

FIGURE 4 shows two packings of a box R with translates of two rectangular bricks B_1 and B_2. In one packing the box is partitioned by a plane into two sub-boxes R_1 and R_2, and the sub-box R_i is a multiple of the brick B_i for $i = 1, 2$. We refer to such a packing as a *bipartitioned packing* of R with B_1 and B_2. FIGURE 4 also shows a non-bipartitioned packing of R with the same bricks B_1 and B_2. Because the trivial box of size $0 \times 0 \times \cdots \times 0$ is a multiple of every nontrivial d-dimensional brick, either of the two sub-boxes may be trivial in a bipartitioned packing; this degenerate situation occurs precisely when the box is a multiple of one of the bricks.

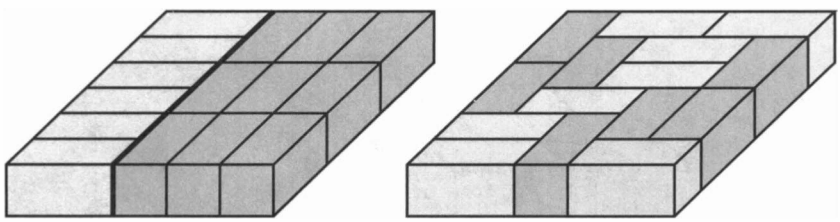


Figure 4 A bipartitioned packing and a non-bipartitioned packing of a box with two bricks

Clearly, the existence of a bipartitioned packing is sufficient for the existence of a packing of a box with translates of two given bricks. The thrust of the next theorem is that this obvious sufficient condition is also necessary.

TWO BRICKS THEOREM (Geometric). *In any dimension the box R can be packed by translates of two given bricks B_1 and B_2 if and only if R can be partitioned by a hyperplane into two sub-boxes R_1 and R_2 such that R_1 is a multiple of B_1, and R_2 is a multiple of B_2.*

We emphasize that the theorem does *not* say that every packing must be bipartitioned. (See FIGURE 4.) However, if there is a non-bipartitioned packing, then there must also be a bipartitioned packing.

We restate the Two Bricks Theorem in the next section. The proof extends over the following three sections. The 2-dimensional case is the crucial one, and we analyze it separately. Note that the 1-dimensional case follows from our proof of Theorem 1.

The two bricks theorem: Arithmetic version We have stated a satisfying and complete *geometric* characterization of the boxes that can be packed by translates of two given bricks. We now provide an equivalent *arithmetic* characterization in terms of the edge lengths of the box and the bricks. Recall that the set of all nonnegative integer linear combinations of v and w is denoted by

$$\langle v, w\rangle = \{bv + cw : b = 0, 1, 2, \ldots, c = 0, 1, 2, \ldots\}.$$

TWO BRICKS THEOREM (Arithmetic). *Let z_i, v_i, and w_i be positive real numbers for $i = 1, 2, \ldots, d$. There is a packing of a $z_1 \times z_2 \times \cdots \times z_d$ box with translates of $v_1 \times v_2 \times \cdots \times v_d$ and $w_1 \times w_2 \times \cdots \times w_d$ bricks if and only if*

(a) *z_i/v_i is an integer for $i = 1, 2, \ldots, d$, or*

(b) *z_i/w_i is an integer for $i = 1, 2, \ldots, d$, or*

(c) *there is an index k such that z_k is in $\langle v_k, w_k\rangle$, and the numbers z_i/v_i and z_i/w_i are integers for all $i \neq k$.*

This theorem helps us answer Question (D).

EXAMPLE. *It is possible to pack a* $12 \times 12 \times 11$ *box with* $4 \times 4 \times 4$ *and* $3 \times 3 \times 3$ *bricks.*

Reason: Condition (c) is satisfied (with $k = 3$) because $12/4$ and $12/3$ are both integers, and $11 = 2(4) + 1(3)$. Thus the desired packing exists. FIGURE 5 shows a bipartitioned packing, where the two sub-boxes have sizes $12 \times 12 \times 8$ and $12 \times 12 \times 3$.

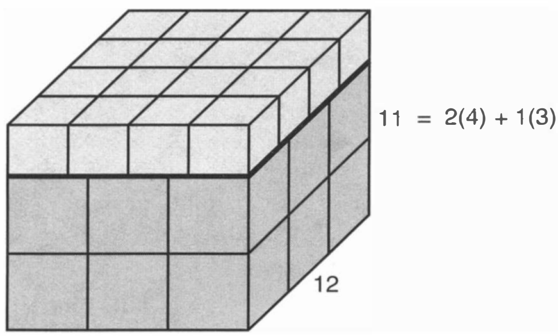


Figure 5 A packing of a $12 \times 12 \times 11$ box with $4 \times 4 \times 4$ and $3 \times 3 \times 3$ cubical bricks

The preceding example and FIGURE 5 help us see why the two versions of the Two Bricks Theorem are equivalent. The two sub-boxes R_1 and R_2 in the geometric version are nonempty and separated by a hyperplane perpendicular to the kth coordinate axis exactly when k is an index for which condition (c) holds in the arithmetic version. The integrality of z_i/v_i and z_i/w_i for $i \neq k$ guarantees that R_1 and R_2 are multiples of the two respective bricks.

Semi-integer rectangles Our proof of the Two Bricks Theorem relies on a result that is of considerable interest in its own right. A *semi-integer rectangle* is a rectangle with at least one edge with integer length. More than a dozen proofs of the following result appear in the prize-winning article by Wagon [**36**]. We include one of the simpler proofs, a generalization of our earlier charge-counting argument that solved Question (B). (Also see Golomb's book [**16**, pp 121–123].)

SEMI-INTEGER RECTANGLES THEOREM. *Suppose that a rectangle R is tiled by a finite number of semi-integer rectangles. Then R itself is a semi-integer rectangle.*

Proof. Place R in the first quadrant with the lower left corner at the origin. Partition R into cells of size $1/2$ by $1/2$ and color the cells black and white in a checkerboard manner with a black cell in the lower left corner, as in FIGURE 6. Each semi-integer rectangular tile encloses an equal amount of black and white area, and therefore R must enclose an equal amount of black and white area. Assume that R has size $z_1 \times z_2$

and is not a semi-integer rectangle. Then the lines $x = \lfloor z_1 \rfloor$ and $y = \lfloor z_2 \rfloor$ partition R into four sub-rectangles, three of which enclose an equal amount of black and white area because they have an edge with integer length. The fourth rectangle R^* has size $z_1^* \times z_2^*$, where $z_k^* = z_k - \lfloor z_k \rfloor$. We assert that R^* encloses more black area than white, and this will give us a contradiction. Now the assertion is clear from a picture unless $z_k^* > 1/2$ for both $k = 1$ and $k = 2$. But in this case a brief calculation shows that the black area of R^* exceeds the white area by $(1 - z_1^*)(1 - z_2^*)$, which is positive. ■

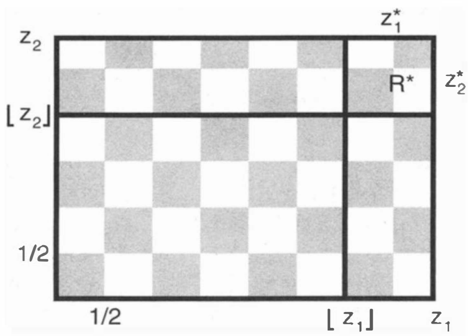


Figure 6 A checkerboard proof of the Semi-Integer Rectangles Theorem

The two tiles theorem We now use the Semi-Integer Rectangles Theorem to give necessary and sufficient arithmetic conditions for a rectangle to be tiled by translates of two (not necessarily integer) rectangles. In other words, we prove the 2-dimensional case of the Two Bricks Theorem.

TWO TILES THEOREM. *A $z_1 \times z_2$ rectangle can be tiled by translates of $v_1 \times v_2$ and $w_1 \times w_2$ rectangular tiles if and only if*

(a) *z_1/v_1 and z_2/v_2 are integers, or*

(b) *z_1/w_1 and z_2/w_2 are integers, or*

(c) *z_1/v_1 and z_1/w_1 are integers and z_2 is in $\langle v_2, w_2 \rangle$ or*

(d) *z_2/v_2 and z_2/w_2 are integers and z_1 is in $\langle v_1, w_1 \rangle$.*

Proof. Suppose that the $z_1 \times z_2$ rectangle R is tiled by translates of $v_1 \times v_2$ and $w_1 \times w_2$ rectangular tiles. Then certainly z_i is in $\langle v_i, w_i \rangle$ for $i = 1, 2$ by the counting conditions. Now scale the tiling of R horizontally by a factor of $1/v_1$ and vertically by a factor of $1/w_2$ to obtain a tiling of a $(z_1/v_1) \times (z_2/w_2)$ rectangle R' by semi-integer rectangles of sizes $1 \times (v_2/w_2)$ and $(w_1/v_1) \times 1$. By the Semi-Integer Rectangles Theorem R' must be a semi-integer rectangle. Hence z_1/v_1 or z_2/w_2 is an integer. Similarly, if we scale R by a factor of $1/w_1$ horizontally and $1/v_2$ vertically, we find that z_1/w_1 or z_2/v_2 is an integer. It now follows that one of the conditions (a)–(d) holds.

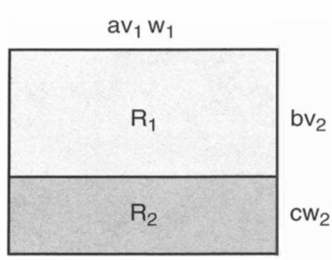


Figure 7 The proof of the Two Tiles Theorem

For the converse, first note that if (a) or (b) holds, then R can be tiled by translates of one of the given tiles. Suppose that (c) holds with $z_1 = av_1w_1$ and $z_2 = bv_2 + cw_2$, where av_1, aw_1, b, and c are nonnegative integers. Then a horizontal line partitions R into two sub-rectangles R_1 and R_2 of respective sizes $av_1w_1 \times bv_2$ and $av_1w_1 \times cw_2$, as in FIGURE 7. Now R_1 can be packed by $(aw_1)(b)$ translates of $v_1 \times v_2$ rectangles, and R_2 can be packed by $(av_1)(c)$ translates of $w_1 \times w_2$ rectangles to give a packing of R. Condition (d) is treated similarly, but uses a vertical line to partition R. ■

Proof of the Two Bricks Theorem We are ready to prove the Two Bricks Theorem in its arithmetic formulation. The theorem holds in dimensions $d = 1$ and $d = 2$ by the proof of Theorem 1 and our work above. Now suppose that $d \geq 3$. If (a) or (b) is true, then the box can be packed by translates of one brick, while if (c) is true, then there is a bipartitioned packing.

Conversely, suppose that a $z_1 \times z_2 \times \cdots \times z_d$ box R is packed by translates of bricks of size $v_1 \times v_2 \times \cdots \times v_d$ and $w_1 \times w_2 \times \cdots \times w_d$. Also, suppose that neither (a) nor (b) holds. Then there are indices j and k such that neither z_j/v_j nor z_k/w_k is an integer. If $j \neq k$, then we inspect a suitable 2-dimensional face of our packing of R to see that there exists a tiling of a $z_j \times z_k$ rectangle with translates of $v_j \times v_k$ and $w_j \times w_k$ tiles. However, none of the conditions (a)–(d) in the Two Tiles Theorem holds, and thus no such tiling exists. Therefore $j = k$. Of course, a counting condition implies that z_k is in the set $\langle v_k, w_k \rangle$. It follows that (c) holds, which completes the proof. ■

We remark that Kolountzakis has recently used Fourier transforms to give a short proof of the Two Bricks Theorem [**26**].

Tiling with integer rectangles When specialized to integer rectangles, the Two Tiles Theorem takes the following useful form.

COROLLARY 2. *Let v_1, v_2, w_1, and w_2 be positive integers with* $\gcd(v_1, w_1) = \gcd(v_2, w_2) = 1$. *Then an integer rectangle of size $n_1 \times n_2$ can be tiled by translates of $v_1 \times v_2$ and $w_1 \times w_2$ rectangles if and only if*

(a) *v_1 divides n_1, and v_2 divides n_2, or*

(b) *w_1 divides n_1, and w_2 divides n_2, or*

(c) *v_1w_1 divides n_1, and n_2 is in $\langle v_2, w_2 \rangle$, or*

(d) *v_2w_2 divides n_2, and n_1 is in $\langle v_1, w_1 \rangle$.*

A mild objection to our route to Corollary 2 may be raised: Our reliance on the Semi-Integer Rectangles Theorem in the proof of the Two Tiles Theorem means that we have gone outside the realm of integers to prove Corollary 2, which deals entirely with integer tiles. In [**5**] Corollary 2 and related results are established using only counting arguments and integers; the Semi-Integer Rectangles Theorem is sidestepped by means of an extension of the charge-counting proof used in FIGURE 3(b) to answer Question (B).

The hypothesis that $\gcd(v_1, w_1) = \gcd(v_2, w_2) = 1$ is not a significant impediment in applying Corollary 2; a scaling argument allows us to produce this case, as in our proof of the following important tiling theorem [**22**].

KLARNER'S THEOREM. *An integer rectangle of size $n_1 \times n_2$ can be tiled by $v \times w$ rectangles (with both orientations allowed) if and only if*

(a) *v divides n_1 or n_2, and*

(b) *w divides n_1 or n_2, and*

(c) n_1 *is in* $\langle v, w\rangle$, *and*

(d) n_2 *is in* $\langle v, w\rangle$.

Proof. If v and w are relatively prime, then the result follows from Corollary 2 with $v_1 = w_2 = v$ and $w_1 = v_2 = w$. If v and w are not relatively prime, then we divide n_1, n_2, v, and w by $\gcd(v, w)$ and apply a scaling argument. ■

Klarner's Theorem implies that there is a tiling of the integer rectangle R with $v \times 1$ and $1 \times v$ rectangles if and only if there is a tiling of R using rectangles in just one orientation. (This fact is also a consequence of a famous theorem of de Bruijn discussed in Part IV.) One implication is immediate. Thus Klarner's Theorem gives a negative answer to Question (A) since neither 16 nor 15 is divisible by 6.

Part III: Two squares and three rectangles

At a combinatorics conference in 1977, Erdős [**12**] posed questions about asymptotic tilings of rectangles. Straus (also in [**12**]) responded by announcing two theorems that extend Sylvester's Theorem to two dimensions. Roughly, *two square tiles with relatively prime edge lengths do not suffice to tile all sufficiently large rectangles, but three square tiles do.* We have been unable to locate proofs in the published works of Straus, and several notable mathematicians at the conference do not recall any arguments that Straus may have communicated [**18**]. In any case, theorems that vindicate Straus have subsequently appeared in the literature. We discuss these results here.

The two squares theorem: Fricke In 1995 Fricke proved the following result about tiling integer rectangles with squares [**13**].

TWO SQUARES THEOREM. *Let* v *and* w *be relatively prime positive integers. An* $n_1 \times n_2$ *rectangle can be tiled by* $v \times v$ *and* $w \times w$ *squares if and only if*

(a) v *divides* n_1 *and* n_2, *or*

(b) w *divides* n_1 *and* n_2, *or*

(c) vw *divides* n_1, *and* n_2 *is in* $\langle v, w\rangle$, *or*

(d) vw *divides* n_2, *and* n_1 *is in* $\langle v, w\rangle$.

Proof. In Corollary 2 let $v_1 = v_2 = v$ and $w_1 = w_2 = w$. ■

We now answer Question (E).

EXAMPLE. *It is impossible to tile a* 39×16 *rectangle with* 3×3 *and* 2×2 *squares.*

Reason: Select $v = 3$, $w = 2$, $n_1 = 39$, and $n_2 = 16$. None of conditions (a)-(d) is satisfied in the Two Squares Theorem.

If v and w are relatively prime and greater than 1, then conditions (a)-(d) imply that no square with edge length $(vw + 1)^s$ can be tiled by $v \times v$ and $w \times w$ square tiles for $s = 1, 2, \ldots$. Thus Fricke's result implies that *two square tiles do not suffice to tile all sufficiently large integer rectangles* (in agreement with Straus). Fricke was apparently unaware of the earlier interest by Erdős and Straus in tiling rectangles with squares.

We have worded the Two Squares Theorem and Klarner's Theorem in a manner that highlights a curious "dual" relationship between them. If we interchange the words "and" and "or" throughout conditions (a)-(d) in either of these theorems, then we nearly obtain the statement of the other theorem. In Part IV we use combinatorial properties of a certain matrix to discover and explain a similar phenomenon in any dimension $d \geq 2$.

Three rectangles suffice, asymptotically The partition-and-pack scheme in FIGURE 8 supplies us with a "proof without words" of an asymptotic tiling theorem for three rectangles. The verbal proof makes the details clear.

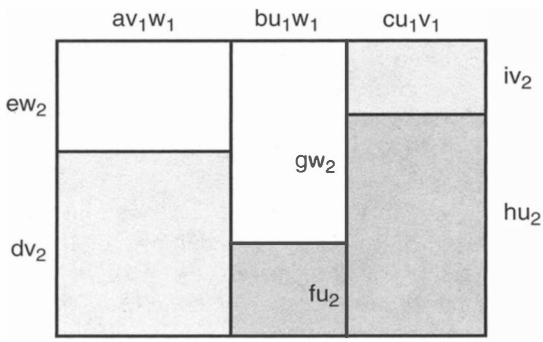


Figure 8 Partition-and-pack strategy for the Three Rectangles Theorem

THREE RECTANGLES THEOREM. *Let T_1, T_2, and T_3 be integer rectangles of respective sizes $u_1 \times u_2$, $v_1 \times v_2$, and $w_1 \times w_2$. Suppose that the three numbers u_j, v_j, and w_j are pairwise relatively prime for $j = 1$ and $j = 2$. Then all sufficiently large rectangles can be tiled by translates of T_1, T_2, and T_3. In particular, any $n_1 \times n_2$ rectangle can be tiled if $n_1 > n^*(v_1w_1, u_1w_1, u_1v_1)$ and $n_2 > \max\{n^*(v_2, w_2), n^*(u_2, w_2), n^*(u_2, v_2)\}$.*

Proof. Let R be an $n_1 \times n_2$ rectangle with n_1 and n_2 satisfying the stated lower bounds. Then there are nonnegative integers $a, b, \ldots, i$ so that

$$n_1 = av_1w_1 + bu_1w_1 + cu_1v_1,$$
$$n_2 = dv_2 + ew_2 = fu_2 + gw_2 = hu_2 + iv_2. \tag{1}$$

We may now partition R into six sub-rectangles, as in FIGURE 8, where each subrectangle can be tiled by one of T_1, T_2, and T_3. For instance, the two darkest rectangles in FIGURE 8 can be tiled by T_1. ■

Consequences of the three rectangles theorem We mention three applications of the theorem in the previous section. First, as an immediate consequence, we see that Straus (in [**12**]) was also correct concerning three square tiles.

THREE SQUARES THEOREM. *All sufficiently large integer rectangles can be tiled by three given squares with pairwise relatively prime edge lengths.*

The following counterexample indicates a lapse in [**12**]; the edge lengths must be *pairwise* relatively prime.

COUNTEREXAMPLE. *Square tiles with edge lengths* 6, 10, *and* 15 *do not tile all sufficiently large integer rectangles, even though* gcd(6, 10, 15) = 1.

Reason: If a rectangle can be tiled by squares with edge lengths 6, 10, and 15, then it certainly can be tiled by squares with edge lengths 2 and 3. The Two Squares Theorem shows that no square with edge length 7^s can be tiled by these two squares for $s = 1, 2, \ldots$.

For our second application, we solve Problem B-3 from the 1991 William Lowell Putnam Mathematical Competition, which asks whether 7×5 and 6×4 rectangles (with both orientations allowed) tile all sufficiently large rectangles. The answer is "yes," and the solutions in **[20, 24]** imply that $N^* \leq 2213$. The detailed analysis by Narayan and Schwenk gives the exact value $N^* = 33$ **[28]**. The Three Rectangles Theorem implies that asymptotic tilings exist even if we fix the orientation of the 6×4 rectangle; in this case we have $N^* \leq n^*(35, 28, 20) = 197$.

Our final application of the Three Rectangles Theorem arises from the proof itself under the additional hypothesis that $n_2 > n^*(u_2, v_2w_2)$. Now there are nonnegative integers h and j so that $n_2 = hu_2 + jv_2w_2$, and thus we may take $f = h$, $g = jv_2$, and $i = jw_2$ in (1). In FIGURE 8 the two darkest sub-rectangles along the lower edge may be merged to form a single sub-rectangle of size $(bu_1w_1 + cu_1v_1) \times hu_2$, which can be tiled by T_1. We have proved the following result, which is in the spirit of the geometric version of the Two Bricks Theorem.

COROLLARY. *Let T_1, T_2, and T_3 be integer rectangles with pairwise relatively prime horizontal edge lengths and pairwise relatively prime vertical edge lengths. Then all sufficiently large integer rectangles can be partitioned into five sub-rectangles, each of which can be tiled by one of T_1, T_2, and T_3.*

Part IV: More bricks, higher dimensions

We now study packings of a d-dimensional box for $d \geq 3$. In some problems our bricks will be d-dimensional cubes, and in other problems we allow all $d!$ orientations of one given brick. Our results are special cases of deeper theorems of Barnes **[1, 2]** and of Katona and Szász **[19]**, who provide necessary and asymptotically sufficient conditions for packings of boxes by sets of given integer bricks. Barnes associates polynomials with bricks and boxes and then applies ideas from algebraic geometry. Katona and Szász employ the theory of matchings from combinatorics. Our proofs in the special cases we examine are simpler than the proofs for the general results in the papers just cited. We also get some better bounds for the asymptotic existence of packings. Finally, we see how a theorem from combinatorial matrix theory can be applied to packing problems.

We restrict our attention to integer boxes and bricks. However, we rely on the following result, which also holds when the edge lengths are not integers. The 2-dimensional case was a key ingredient in our proof of the Two Bricks Theorem in Part II. Several proofs based on combinatorial arguments are featured in Wagon **[36]**.

SEMI-INTEGER BRICKS THEOREM. *Suppose that a d-dimensional box R is packed by a finite number of d-dimensional bricks. If each brick has at least m integer edge lengths, then R itself has at least m integer edge lengths.*

We remark that for packings with cubes our restriction to integer edge lengths is not really a restriction at all, due to a beautiful result of Dehn **[11]**.

DEHN'S THEOREM. *If a box is packed by a finite number of cubes in d dimensions ($d \geq 2$), then all edge lengths of the cubes and the box become integers after scaling the entire configuration by a suitable positive number.*

One brick, all orientations: de Bruijn and Kelly The following famous theorem of de Bruijn gives necessary conditions for a d-dimensional box to be packed by a single brick with all orientations allowed [**9**].

DE BRUIJN'S THEOREM. *If the d-dimensional integer box of size* $n_1 \times n_2 \times \cdots \times n_d$ *is packed by an integer brick of size* $v_1 \times v_2 \times \cdots \times v_d$ *(with all d! orientations allowed), then for each index i in* $\{1, 2, \ldots, d\}$ *there exists an index j such that* v_i *divides* n_j.

A brick of size $v_1 \times v_2 \times \cdots \times v_d$ is *harmonic* provided either v_i divides v_j, or v_j divides v_i for all pairs $i, j = 1, 2, \ldots, d$. De Bruijn also showed that the divisibility conditions in the above theorem are sufficient for packings with harmonic bricks.

The seminal paper by de Bruijn [**9**] (which promulgated the results in the earlier Hungarian publications [**7, 8**]) introduced polynomials and complex exponentials to the study of packings. The elegance and power of de Bruijn's method inspired a host of variations and generalizations [**1, 2, 4, 19, 21, 36**]. We now give a different proof of one of these generalizations [**21**].

KELLY'S THEOREM. *If the d-dimensional integer box of size* $n_1 \times n_2 \times \cdots \times n_d$ *is packed by integer bricks of size* $v_1 \times v_2 \times \cdots \times v_d$ *(with all d! orientations allowed), then the following statement holds for* $m = 1, 2, \ldots, d$*: For each index set* $\{i_1, i_2, \ldots, i_m\}$ *there exists an index set* $\{j_1, j_2, \ldots, j_m\}$ *such that* $\gcd(v_{i_1}, v_{i_2}, \ldots, v_{i_m})$ *divides* $\gcd(n_{j_1}, n_{j_2}, \ldots, n_{j_m})$.

Proof. Let R be a box of size $n_1 \times n_2 \times \cdots \times n_d$ and let $g = \gcd(v_{i_1}, v_{i_2}, \ldots, v_{i_m})$. Scale the packing of R by a factor of $1/g$ in all d directions to produce a box R' of size $(n_1/g) \times (n_2/g) \times \cdots \times (n_d/g)$ that is packed by bricks, each of which has (at least) m integer edge lengths. By the Semi-Integer Bricks Theorem at least m of the edge lengths of R' are integers, that is, g divides at least m of the numbers $n_1, n_2, \ldots, n_d$. ∎

Note that de Bruijn's packing conditions correspond to the case $m = 1$ in Kelly's Theorem. We now answer Question (F).

EXAMPLE. *It is impossible to pack a* $720 \times 200 \times 100$ *box with* $15 \times 12 \times 10$ *bricks (with all six orientations allowed).*

Reason. Although de Bruijn's conditions are satisfied (as are the counting conditions), Kelly's conditions fail for $m = 2$ because $\gcd(15, 12) = 3$, and 3 divides just one edge length of the box.

Packing boxes with cubes in d dimensions As suggested in [**12**], the results about tilings with squares in Part III are elements of a grander picture: *With mild conditions on the edge lengths no set of d cubes in d dimensions packs all sufficiently large boxes, but every set of* $d + 1$ *cubes does.* We now discuss theorems that make this assertion precise.

THEOREM. *Let* $v_1, v_2, \ldots, v_d$ *be integers greater than* 1. *Then the set of d-dimensional cubes with edge lengths* $v_1, v_2, \ldots, v_d$ *does not pack all sufficiently large integer boxes.*

Proof. The case $d = 1$ is easy. Suppose that $d \geq 2$.

CLAIM. *If a box R of size $n_1 \times n_2 \times \cdots \times n_d$ is packed by translates of d cubes with edge lengths $v_1, v_2, \ldots, v_d$, then v_j divides n_j for some index j.*

Scale the packing of R by a factor of $1/v_i$ in the direction of the ith axis for $i = 1, 2, \ldots, d$ to produce a packing of a box R' of size $(n_1/v_1) \times (n_2/v_2) \times \cdots \times (n_d/v_d)$ with translates of d bricks, each of which has an edge of length 1. By the Semi-Integer Bricks Theorem n_j/v_j is an integer for some index j, which establishes the claim. The divisibility restrictions imposed by the claim yield the conclusion of the theorem. ■

Because the proof of the claim remains true for any permutation of $v_1, v_2, \ldots, v_d$, the preceding argument gives us strong necessary conditions for a d-dimensional box to be packed by d cubes. These conditions turn out to be asymptotically sufficient when the edge lengths of the cubes are pairwise relatively prime [**19**].

THEOREM 3. (a) *If an $n_1 \times n_2 \times \cdots \times n_d$ box is packed by cubes with edge lengths $v_1, v_2, \ldots, v_d$, then for each permutation π of $\{1, 2, \ldots, d\}$ there exists an index i such that v_i divides $n_{\pi(i)}$.*

(b) *If $v_1, v_2, \ldots, v_d$ are pairwise relatively prime and $n_1, n_2, \ldots, n_d$ are sufficiently large, then the conditions in* (a) *are sufficient for a packing.*

Assertion (a) follows from the argument given earlier. We postpone the proof of (b) to the final section of this article. We now turn to packings with $d + 1$ cubes.

THEOREM 4. *In d dimensions all sufficiently large boxes can be packed by any given set of $d + 1$ cubes with pairwise relatively prime edge lengths.*

Proof. Let $v_1, v_2, \ldots, v_{d+1}$ be the pairwise relatively prime edge lengths of the given cubes. We induct on d. Sylvester's Theorem is the case $d = 1$. Suppose that $d \geq 2$ and let R be an $n_1 \times n_2 \times \cdots \times n_d$ box with

$$n_1 \geq n^*(\widehat{v}_1, \widehat{v}_2, \ldots, \widehat{v}_{d+1}),$$

where

$$\widehat{v}_k = (v_1 v_2 \cdots v_{d+1}) / v_k$$

for $k = 1, 2, \ldots, d + 1$. There are nonnegative integers $c_1, c_2, \ldots, c_{d+1}$ such that

$$n_1 = c_1\widehat{v}_1 + c_2\widehat{v}_2 + \cdots + c_{d+1}\widehat{v}_{d+1}.$$

Thus we may partition R into sub-boxes $R_1, R_2, \ldots, R_{d+1}$, where R_k is of size $(c_k\widehat{v}_k) \times n_2 \times \cdots \times n_d$. By the induction hypothesis, if $n_2, n_3, \ldots, n_d$ are sufficiently large, then for each $k = 1, 2, \ldots, d + 1$ we can pack the $(d - 1)$-dimensional box R'_k of size $n_2 \times n_3 \times \cdots \times n_d$ with $(d - 1)$-dimensional cubes with edge lengths in $\{v_1, v_2, \ldots, v_{d+1}\} \setminus \{v_k\}$. Observe that R'_k is a $(d - 1)$-dimensional face of the sub-box R_k for $k = 1, 2, \ldots, d + 1$. Because $c_k\widehat{v}_k$ is divisible by each length in $\{v_1, v_2, \ldots, v_{d+1}\} \setminus \{v_k\}$, the packing of R'_k with $(d - 1)$-dimensional cubes can be "lifted" to a packing of R_k with d-dimensional cubes with the same edge lengths for $k = 1, 2, \ldots, d + 1$. The packings of $R_1, R_2, \ldots, R_{d+1}$ yield the desired packing of R. ■

The proof shows that

$$N^* \leq n^*(\widehat{v}_1, \widehat{v}_2, \ldots, \widehat{v}_{d+1}).$$

(It is not hard to show that the successive bounds that arise for $n_2, n_3, \ldots, n_d$ in the induction cannot be larger than this expression.) As an example, for four cubes with edge lengths 7, 5, 3, and 2 we have $N^* \leq n^*(105, 70, 42, 30) = 383$, which is vastly superior to the gigantic bound $N^* \leq 3^{49152} \cdot 7^{4098}$ from the general constructions in [**19**]. An explicit bound for N^* does not appear in [**2**].

A construction similar to the one in the proof of Theorem 4 establishes a more general result.

THEOREM 3. *In d dimensions all sufficiently large boxes can be packed by any given set of $d+1$ bricks whose corresponding edge lengths are pairwise relatively prime.*

Divisibility matrices We now introduce a matrix to record the divisibility relations between the brick and box edge lengths. Let $n_1, n_2, \ldots, n_d$ and $v_1, v_2, \ldots, v_d$ be two sequences of positive integers. The *divisibility matrix* $A_{n/v} = [a_{ij}]$ is a d by d matrix with (i, j)-entry

$$a_{ij} = \begin{cases} 0 & \text{if } v_i \text{ divides } n_j \\ 1 & \text{if } v_i \text{ does not divide } n_j \end{cases}$$

Our definition differs slightly from the one in Katona and Szász, but carries the same information [**19**].

EXAMPLE. *If* $(n_1, n_2, n_3) = (720, 200, 100)$ *and* $(v_1, v_2, v_3) = (15, 12, 10)$, *then the divisibility matrix is*

$$A_{n/v} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

A *transversal* in a matrix A of order d is a set of d positions $(i, \pi(i))$ for $i = 1, 2, \ldots, d$, where π is a permutation of $\{1, 2, \ldots, d\}$. A transversal includes exactly one position in each row and one position in each column of A. The $d!$ transversals in A correspond to the terms in an expansion of the determinant of A. Let us use the divisibility matrix to restate two of our earlier results. We continue to assume all lengths are positive integers.

DE BRUIJN'S THEOREM (Restated). *If an $n_1 \times n_2 \times \cdots \times n_d$ box is packed by $v_1 \times v_2 \times \cdots \times v_d$ bricks (with all $d!$ orientations allowed), then each of the d rows of the divisibility matrix $A_{n/v}$ contains a* 0.

THEOREM 3 (Restated). (a) *If an $n_1 \times n_2 \times \cdots \times n_d$ box is packed by d cubes with edge lengths $v_1, v_2, \ldots, v_d$, then each of the $d!$ transversals of the divisibility matrix $A_{n/v}$ contains a* 0.

(b) *The transversal condition in* (a) *is sufficient for a packing if $v_1, v_2, \ldots, v_d$ are pairwise relatively prime and $n_1, n_2, \ldots, n_d$ are sufficiently large.*

The divisibility matrix has revealed an unusual relationship between de Bruijn's Theorem and Theorem 3 for fixed edge lengths $v_1, v_2, \ldots, v_d$ and $n_1, n_2, \ldots, n_d$. In de Bruijn's Theorem we have translates of $d!$ bricks, and there are d conditions imposed on the divisibility matrix. In Theorem 3 there are d bricks and $d!$ conditions on the divisibility matrix. When $d = 2$, these two theorems become Klarner's Theorem and the Two Squares Theorem, respectively, and the curious relationship we noted earlier is now placed in a broader context.

The transversal condition in Theorem 3(a) has combinatorial ramifications. Let A be a matrix of order d. If A has a row or a column of 0's, then every transversal of A clearly contains a 0. More generally, one may show that if A has a p by $d-p+1$ submatrix of 0's for some p in $\{1, 2, \ldots, d\}$, then every transversal contains a 0. Frobenius established the converse, and proved the following cornerstone of combinatorial matrix theory **[14, 15]**.

FROBENIUS'S THEOREM. *Let A be a matrix of order d. Then every transversal of A contains a 0 if and only if A has a p by $d-p+1$ submatrix of 0's for some p in $\{1, 2, \ldots, d\}$.*

Frobenius's Theorem is a member of a celebrated family of equivalent combinatorial results (including the Marriage Theorem, Dilworth's Theorem for partially ordered sets, and König's Theorem for bipartite graphs). Proofs and discussions of the relationships among these theorems can be found in combinatorics books, e.g., **[6, 17, 31]**.

Proof of an Asymptotic Packing Theorem Katona and Szász applied the Marriage Theorem to establish necessary and asymptotically sufficient conditions for a box to be packed by translates of a set of given bricks **[19]**. As a glimpse of their more advanced techniques we use the divisibility matrix to give a short proof of Theorem 3(b). Our proof combines several ideas in this article.

Proof of Theorem 3(b) Our strategy is to use Frobenius's Theorem to reduce the problem to the packing of a $(p-1)$-dimensional box with p cubes, apply Theorem 4, and then "lift" to a d-dimensional packing. By the transversal hypothesis and Frobenius's Theorem we know that the divisibility matrix $A_{n/v} = [a_{ij}]$ has a p by $d-p+1$ submatrix of 0's for some p in $\{1, 2, \ldots, d\}$. Without loss of generality we may place the submatrix of 0's in the upper right corner of A so that $a_{ij} = 0$ for $i = 1, 2, \ldots, p$ and $j = p, p+1, \ldots, d$. We will show that the box R of size $n_1 \times n_2 \times \cdots \times n_d$ can be packed by cubes with edge lengths $v_1, v_2, \ldots, v_p$. If $p = 1$, then each edge length of R is divisible by v_1, and R can certainly be packed by cubes with edge length v_1. Now suppose that $p > 1$ and consider the $(p-1)$-dimensional box R' of size $n_1 \times n_2 \times \cdots \times n_{p-1}$. It is helpful to observe that R' is a $(p-1)$-dimensional face of R. By Theorem 4 the $(p-1)$-dimensional cubes with edge lengths $v_1, v_2, \ldots, v_p$ can be used to pack the box R' if $n_1, n_2, \ldots, n_{p-1}$ are sufficiently large. The known positions of 0's in A tell us that v_i divides n_j for $i = 1, 2, \ldots, p$ and $j = p, p+1, \ldots, d$. From these divisibility relations it follows that we may "lift" our packing of the $(p-1)$-dimensional face R' to obtain a packing of the d-dimensional box R with d-dimensional cubes having edge lengths $v_1, v_2, \ldots, v_p$. ∎

Acknowledgment. We thank Courtney Moen for his help and ideas and the referees for some improvements in clarity. TSM was partially supported by the Naval Academy Research Council.

REFERENCES

1. F. W. Barnes, Algebraic theory of brick packing I, *Disc. Math.* **42** (1982), 7–26.
2. F. W. Barnes, Algebraic theory of brick packing II, *Disc. Math.* **42** (1982), 129–144.
3. F. Beukers, Frobenius applet, `http://www.math.uu.nl/people/beukers/frobenius/`, posted January 2001.
4. P. Boisen, Polynomials and packings: a new proof of de Bruijn's theorem, *Disc. Math.* **146** (1995), 285–287.
5. R. J. Bower and T. S. Michael, Tiling a box with translates of two rectangular bricks, *Elec. J. of Combin.* **11** (2004), #N7, 9 pages, (electronic).

6. R. A. Brualdi and H. J. Ryser, *Combinatorial Matrix Theory*. Cambridge University Press, 1991.
7. N. G. de Bruijn, Problem 119, *Matematikai Lapok.* **12** (1961), 103.
8. N. G. de Bruijn (with G. Hajos, G. Katona, and D. Szász), Solution to Problem 119, *Matematikai Lapok.* **13** (1962), 314–317.
9. N. G. de Bruijn, Filling boxes with bricks, *Amer. Math. Monthly* **76** (1969), 37–40.
10. J. H. Conway and J.C. Lagarias, Tiling with polyominoes and combinatorial group theory, *J. Combin. Th. A* **53** (1990), 183–208.
11. M. Dehn, Über Zerlegung von Rechtecken in Rechtecke, *Math. Annalen* **57** (1903), 314–332.
12. P. Erdős, A tiling problem, Problem session, Proceedings Combinatorial Mathematics, Canberra 1977, *Lecture Notes in Mathematics* 686, A. Dold and B. Eckham, eds. Springer, 1977.
13. J. Fricke, Quadratzerlegung eines Rechtecks, *Math. Semesterber.* **42** (1995), 53–62.
14. G. Frobenius, Über Matrizen aus nicht negativen Elementen, *S.-B K. Preuss. Akad. Wiss. Berlin* (1912), 456–477.
15. G. Frobenius, Über zerlegbare Determinanten, *S.-B K. Preuss. Akad. Wiss. Berlin* (1917), 274–277.
16. S. W. Golomb, *Polyominoes,* 2nd ed. Princeton Univ. Press, 1994.
17. M. Hall, *Combinatorial Theory,* 2nd ed. Wiley, New York, 1986.
18. D. A. Holton, B. D. McKay, R. G. Stanton, and W. D. Wallis, personal communications, March, 1999.
19. G. Katona and D. Szász, Matching problems, *J. Combin. Th. B* **10** (1971), 60–92.
20. K. S. Kedlaya, B. Poonen, and R. Vakil, *The William Lowell Putnam Mathematical Competition 1985–2000: Problems, Solutions, and Commentary*, Math. Association of America, Washington, DC, 2002.
21. J. B. Kelly, Polynomials and polyominoes, *Amer. Math. Monthly* **73** (1966), 464–471.
22. D. A. Klarner, Packing a rectangle with congruent N-ominoes, *J. Combin. Th. A* **7** (1969), 107–115.
23. D. A. Klarner, Brick-packing puzzles, *J. Rec. Math.* **6** (1973), 112–117.
24. L. F. Klosinski, G. L. Alexanderson, and L. C. Larson, The fifty-second William Lowell Putnam Mathematical Competition, *Amer. Math. Monthly* **99** (1992), 715–724.
25. G. Kolata, In Archimedes' puzzle, a new eureka moment, *New York Times,* (14 December 2003) pp 1, 46 (National Edition pp 1, 32); `http://www.nytimes.com/2003/12/14/science/14MATH.html`.
26. M. N. Kolountzakis, Filling a box with translates of two bricks, *Elec. J. of Combin.* **11** (2004), #N16, 4 pages, (electronic).
27. G. E. Martin, *Polyominoes, A Guide to Puzzles and Problems in Tiling.* Math. Association of America, Washington, DC, 1991.
28. D. A. Narayan and A. J. Schwenk, Tiling large rectangles, this MAGAZINE **75** (2002), 372–380.
29. R. W. Owens, An algorithm to solve the Frobenius problem, this MAGAZINE **76** (2003), 264–275.
30. J. Propp, A pedestrian approach to a method of Conway, or, a tale of two cities, this MAGAZINE **70** (1997), 327–340.
31. P. Reichmeider, *The Equivalence of Some Combinatorial Matching Theorems.* Polygonal Publ. House, Washington, NJ, 1984.
32. M. Reid, Tile homotopy groups, *Enseign. Math.* **49** (2003), 123–155.
33. S. K. Stein and S. Szabó, *Algebra and Tiling: Homomorphisms in the Service of Geometry* Math. Association of America, Washington, DC, 1994.
34. J. J. Sylvester, Mathematical questions with their solutions, *Educational Times* **41** (1884), 21.
35. W. Thurston, Conway's tiling groups, *Amer. Math. Monthly* **97** (1990), 757–773.
36. S. Wagon, Fourteen proofs of a result about tiling a rectangle, *Amer. Math. Monthly* **94** (1987), 610–617.

Don't Let Them Know It's Math

The instructions for the Philadelphia Metro's (and many other newspapers') Sudoku puzzle read:

> "Fill in the grid so that every row, every column, and every 3×3 box contains the digits 1–9. There is no math involved. You solve the puzzle with reasoning and logic."!!!

Leslie Hayes
Saint Joseph's University
Philadelphia, PA 19131

A New Model for Ribbons in $\mathbb{R}^3$

TOM FARMER
Miami University
Oxford, OH 45056
farmerta@muohio.edu

Let's start with a hands-on demonstration. We take a long, rectangular strip of adhesive tape (or postage stamps) and wind it diagonally around a cylinder in such a way that the sticky surface of the tape is entirely in contact with the cylinder. It follows that the two long edges of the tape must lie along parallel helices. This is true by definition— if we start with a line in a plane and then wrap the plane around a cylinder we get a helix. What about the short edges of the tape? Well, by the same definition, they must lie along parallel helices that are orthogonal to the first two! We call the resulting, familiar object a *helical band* (FIGURE 1). It plays a central role in this discussion.

Figure 1 A helical band

Here is a more general demonstration. Starting with a long, rectangular strip of paper, we fix in our minds a three-dimensional curve C having the same length as the strip. Now we force one long edge of the strip to follow the curve, and call the result a *ribbon* from now on. Surely the whole ribbon must bend and twist if one edge is to stick to the curve. Can we describe the shape of the ribbon? More to the point, can we present a mathematical model that would help us visualize such a ribbon and might help us illustrate or understand the curve?

At this point it is helpful to introduce some terminology. In the plane of the rectangular strip of paper, let's choose an orientation so that the cross-sections parallel to the long edges are horizontal and the cross-sections parallel to the short edges are vertical. When the ribbon is bent along a curve in space, these horizontal and vertical cross-sections are positioned along curves that we call *threads* and *transversals*, respectively. We force one thread (one long edge of the strip) to follow the given curve C. But what can we say about the other threads and about the transversals? Should the other threads be pictured as congruent copies of C? Should the transversals be pictured as line segments orthogonal to C? Before trying to answer these questions, it is very important to state that our limited goal is to construct the threads and transversals of a realistic-looking ribbon—we hope only for verisimilitude, and make no assertions about the physics of the situation.

As we have already seen, if the curve C is a helix then a helical band can be formed as one physical realization of a ribbon following this curve. All the threads of the ribbon are helices parallel to C, and all the transversals are helices orthogonal to C. But this realization is not unique. For example, without changing the position of one edge along the given helix, an opposite corner can be curled up so that we no longer have a helical band. Thus, we cannot claim that forcing one edge of a ribbon to follow a given curve uniquely determines the shape of the ribbon.

How do we construct a ribbon model? Let's focus on the transversals. The main requirements are that each transversal must be a curve of the proper length and be orthogonal to C, since these properties appear to be true of real-life ribbons. The simplest choice (perhaps too simple) is to use line segments. If each transversal is rendered as a line segment orthogonal to C then we call the model a (generalized) Linear Transversal (LT) model. An LT model may be a reasonable way to thicken a curve, and this can be useful in computer graphics. If a curve has been thickened into some sort of ribbon or tube, then hidden-surface and lighting techniques can be effective in illustrating the curve. However, the LT model often won't provide a realistic shape. In particular, it fails to yield a helical band when C is a helix (FIGURE 2).

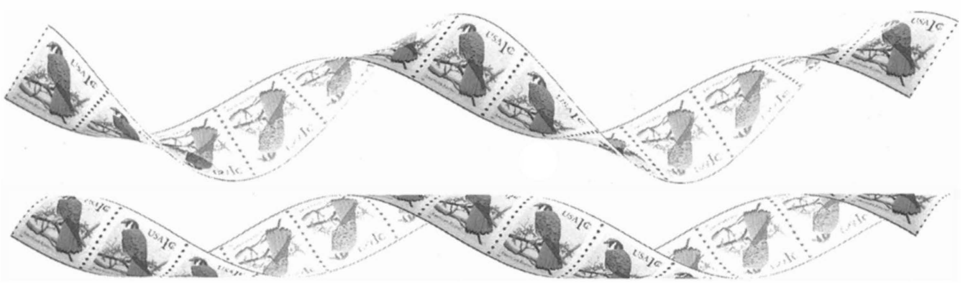


Figure 2 LT model (top) and HT model (bottom)

On the other hand, in our new model, we propose to emulate the helical band and render each transversal as an arc of a helix. This is the Helical Transversal (HT) model. The transversals at the top of the ribbons in FIGURE 3 show the subtle, but significant, difference between the two models.

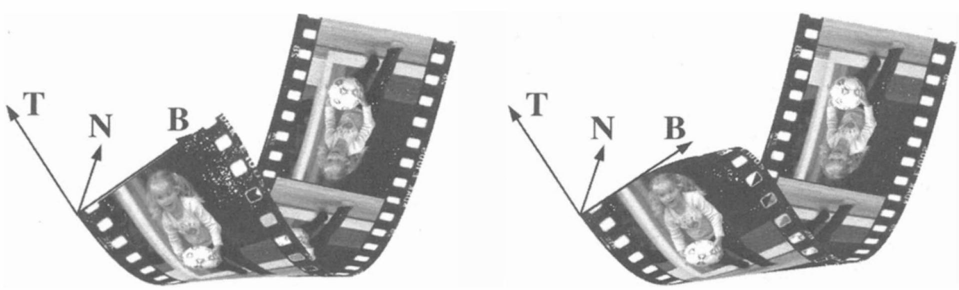


Figure 3 LT model (left) and HT model (right)

Of course, both the LT and HT models are just models and they do not yield the exact shapes of real-world ribbons, except in special cases. Even so, models can provide plausible and attractive pictures helping to illustrate interesting curves. Indeed, the illustrations in this article have been produced using MATLAB to implement the formulas for both models. By the way, computer-drawn ribbons are used in science to visualize certain large molecules. For example, Carson [**1**] has employed a construction of threads (rather than transversals) to form ribbons that aid in understanding and illustrating the molecular structure of proteins. Our focus on transversals is apparently a different approach from Carson's.

For simplicity, throughout this discussion we assume that the curve C can be given in parametric form $s \mapsto \langle x(s), y(s), z(s)\rangle$, with arc length $s \in [0, l]$ as parameter and in terms of functions that are sufficiently differentiable. To be more precise, although we won't dwell on this matter, we assume that the fifth derivatives of the coordinate functions exist and are continuous on $[0, l]$. This suffices to force the curvature and torsion functions (discussed in the next section) to have at least two continuous derivatives. We further assume that C never intersects itself and that it has positive curvature at every point. Such assumptions are commonly used in the study of space curves, and we want to focus on the successful modeling of ribbons rather than on the potential pitfalls of troublesome curves.

Ribbon maps and ribbon models

Let us clarify the distinction between actual ribbons and our mathematical models of ribbons. For the purposes of this article, we think of a real-world ribbon (such as a flexible ruler, a roll of postage stamps, a strip of film, etc.) as being made of a material that cannot be distorted or stretched or shrunk in any direction. A mathematical way to describe such a ribbon is by a ribbon map, defined on a rectangular domain D. Let $D = [0, l] \times [0, w]$ (typically with $l \gg w$), where l is the length and w is the width of the ribbon. Now let's define a *ribbon map* to be a function $\mathbf{f} : D \to \mathbb{R}^3$, having continuous first partial derivatives on D, and satisfying the following three properties:

(R1) The transversal $v \mapsto \mathbf{f}(s, v)$ has $v \in [0, w]$ as its arc length parameter, for every fixed $s \in [0, l]$. That is, $\mathbf{f}_v \cdot \mathbf{f}_v \equiv 1$.

(R2) The thread $s \mapsto \mathbf{f}(s, v)$ has $s \in [0, l]$ as its arc length parameter, for every fixed $v \in [0, w]$. That is, $\mathbf{f}_s \cdot \mathbf{f}_s \equiv 1$.

(R3) Transversals and threads intersect at right angles. That is, $\mathbf{f}_s \cdot \mathbf{f}_v \equiv 0$.

Some readers may recognize that these properties amount to $\mathbf{f}$ being a local isometry. Indeed, in terms of the classical notation in the study of surfaces, the properties can be compactly expressed as $G = 1$, $E = 1$, and $F = 0$.

Here is a more precise version of our problem of forcing one edge of a ribbon to follow a given curve. Let C be a curve of finite length l in $\mathbb{R}^3$ and let $D = [0, l] \times [0, w]$ be given. We would like to produce a ribbon map $\mathbf{f} : D \to \mathbb{R}^3$ such that $s \mapsto \mathbf{f}(s, 0)$ is the curve C.

The author is not aware of any exact solutions to this problem of finding ribbon maps, except in special cases, so let us speak of approximate solutions—models—instead. A *ribbon model* for D and C is a map $\mathbf{f} : D \to \mathbb{R}^3$ (with continuous first partials on D) such that $s \mapsto \mathbf{f}(s, 0)$ is the curve C, and such that (R1) holds. Whether the map $\mathbf{f}$ satisfies (R2) and (R3) is what separates a ribbon model from an exact ribbon map.

By the way, our definition of a ribbon map does not require the map to be one-to-one. Likewise, a ribbon model may not be one-to-one. This flexibility might be desirable, as in the case of modeling a helical band that partially overlaps itself. Alternatively, the domain of a ribbon map that is not one-to-one can be restricted to a subset $D_1 = [0, l] \times [0, w_1]$ of D in order to achieve a one-to-one result. In effect, we would just be choosing a narrower ribbon that is easier to bend along a curve without bumping into itself.

Again, let C be represented by the arc length parametrization

$$s \longmapsto \langle x(s), y(s), z(s) \rangle.$$

Ribbon models are expressed in terms of the *Frenet frame*, a "moving" coordinate system that we construct at each point of the curve. At the point $P = P(s)$ on the curve, we define unit vectors $\mathbf{T}$, $\mathbf{N}$, and $\mathbf{B}$. The *unit tangent vector* $\mathbf{T}$ is a vector of length 1 pointing in the direction of the derivative $\langle x'(s), y'(s), z'(s)\rangle$. The *unit normal vector* $\mathbf{N}$ is the vector of length 1 in the direction of $d\mathbf{T}/ds$. We can understand $\mathbf{T}$ and $\mathbf{N}$ by thinking of an object moving at unit speed along the curve. In this situation, $\mathbf{T}(s)$ is the velocity vector at time s and $\mathbf{N}(s)$ is in the direction of the acceleration vector. These vectors are necessarily orthogonal to each other because

$$0 = \frac{d}{ds}(1) = \frac{d(\mathbf{T}\cdot\mathbf{T})}{ds} = 2\mathbf{T}\cdot\frac{d\mathbf{T}}{ds}.$$

We complete the coordinate system (the Frenet frame) by defining the *unit binormal vector* $\mathbf{B}$ to be $\mathbf{T}\times\mathbf{N}$. The properties of cross products guarantee that this is a unit vector orthogonal to both $\mathbf{T}$ and $\mathbf{N}$.

Our discussion of ribbon models also requires the notions of curvature and torsion along a curve. These are scalar quantities that characterize the geometry of the curve. Using the Frenet frame, we have a straightforward way to define and describe the *curvature* κ and *torsion* τ of a space curve. In fact, the *Frenet-Serret* formulas (described in detail by Stewart [**4**] or Oprea [**3**]) tell the story:

$$\frac{d\mathbf{T}}{ds} = \kappa\mathbf{N}, \qquad \frac{d\mathbf{N}}{ds} = -\kappa\mathbf{T} + \tau\mathbf{B}, \qquad \frac{d\mathbf{B}}{ds} = -\tau\mathbf{N}. \tag{1}$$

By the first of these formulas, the curvature κ is a measure of how fast the curve is turning within the *osculating plane* spanned by $\mathbf{T}$ and $\mathbf{N}$. And by the third formula, since $\mathbf{B}$ is normal to the osculating plane, the torsion τ is a measure of how quickly the normal to the osculating plane is turning—or how the curve is twisting.

As one example, the torsion of a circle, or of any other plane curve, is zero since the osculating plane is fixed. As a second example, think of a helix. We derive in the next section the fact that both the curvature and the torsion of a helix are constants. A nonzero, constant torsion reflects the understanding that, along a helix, the vector $\mathbf{B}$ turns at a constant rate.

In the Linear Transversal model, each transversal is the line segment of length w starting at a point P and heading in the direction of $\mathbf{B}$. Why $\mathbf{B}$? The easiest way to explain this choice is to think of the special case in which the given curve is planar. In this case, the ribbon can be naturally realized as a (right, noncircular) cylinder. For example, imagine a coil of postage stamps standing on a desktop and following a spiral curve in the plane of the desktop (FIGURE 4). The vectors $\mathbf{T}$ and $\mathbf{N}$ are in the plane of the curve and the transversals are line segments in the direction of $\mathbf{B}$ or $-\mathbf{B}$.

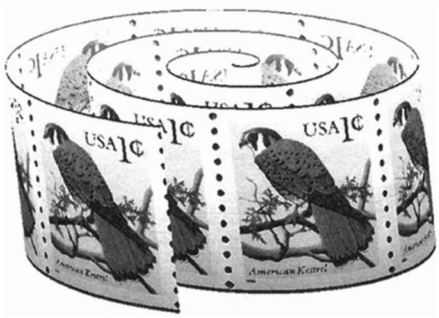


Figure 4 Ribbon for a planar curve

For simplicity, we focus first on one choice of sign. The formula for the LT model is simply

$$\mathbf{f}(s, v) = \langle x(s), y(s), z(s)\rangle + v\mathbf{B}(s), \tag{2}$$

where $(s, v) \in D$. The differentiability assumptions for the curve guarantee that the first partials of $\mathbf{f}$ exist and are continuous on D. Also, it is clear by our construction of $\mathbf{f}$ that (R1) holds. In fact, $\mathbf{f}_v \cdot \mathbf{f}_v = \mathbf{B} \cdot \mathbf{B} \equiv 1$. Thus, the LT model is a ribbon model.

Similar models are obtained by replacing $\mathbf{B}$ in this formula with another unit vector in the normal plane to the curve. In this way, we could regard the *generalized* LT model to be

$$\mathbf{f}(s, v) = \langle x(s), y(s), z(s)\rangle + v((\cos q)\mathbf{N} + (\sin q)\mathbf{B}), \tag{3}$$

where $s \mapsto q(s)$ is a continuously differentiable function. We investigate these models in the next section.

The Helical Transversal model is strongly motivated by the helical band. We have already noted that when a rectangle is wrapped around a (right, circular) cylinder, the boundaries of the rectangle become helical arcs on the cylinder. The wrapping operation we speak of preserves distances and angles, so it sends perpendicular lines to orthogonal helices. Thus, in fact, for any helix H through a point P on a given cylinder, there is a unique helix H^{orth} lying on the same cylinder and orthogonal to H at P.

There are two fundamental ideas behind the HT model. One is that we may view a curve as being approximated at each point by a certain helix—in general, a different helix at each point. In fact, we show in Theorem 1 that at any point P of a given curve C there is a unique, approximating helix that has the same curvature, torsion and Frenet frame at P. This is the main result in a MAGAZINE article by McHugh [2]. The other fundamental idea is to choose the transversal at P to be an arc of the helix that is orthogonal to the approximating helix, just as if the ribbon were a helical band. In order to work out the formula for this HT model, we need to do some preparation in the next section.

By the way, a quick comparison of the two illustrations in FIGURE 5 shows that the LT and HT models can appear remarkably similar. In this case, however, the HT model seems to do a better job of portraying a ribbon whose bottom edge follows a 3-dimensional, spiral curve C. (Look at the "tightly wound" end of the spiral.)

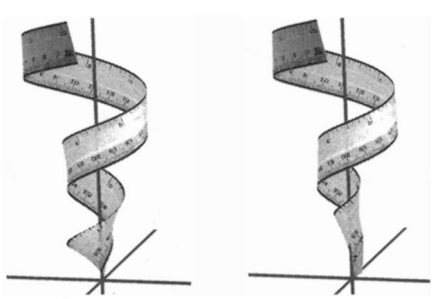


Figure 5 LT model (left) and HT model (right)

Helix redux

The helix plays a major role in the study of space curves. Just as the shape of a plane curve at a point is compared to a circle by calculating the curvature, the shape of a space curve at a point can be compared to a helix by calculating the curvature and the torsion.

Tangent vectors to a helix make a constant angle $\theta \in [0, \pi]$ with the chosen direction vector of the axis of the cylinder. Note that $\theta = \pi/2$ corresponds to a circle and $\theta = 0$ or π corresponds to a line parallel to the axis of the cylinder. These are degenerate helices, and we will avoid them in the rest of this section.

A helix on a given cylinder can be right-handed, like the threads of a standard screw, or left-handed. One way to characterize the difference is in terms of the tangent vector **T** and the binormal vector **B** at each point. Once the positive direction of the axis of the cylinder is selected, a right-handed helix has the property that the angles that **T** and **B** make with the axis are both acute or both obtuse. By contrast, for a left-handed helix one of these angles is acute and the other is obtuse.

We need formulas describing the geometry of a helix. By choosing convenient coordinates, it is enough to work with a helix on the cylinder of radius r having the z-axis as its directed axis and passing through the point $(r, 0, 0)$. Such a helix can be parameterized by

$$\begin{cases} x = r\cos t \\ y = r\sin t \quad t \in \mathbb{R} \\ z = ct \end{cases} \tag{4}$$

The helix is right-handed if $c > 0$ and left-handed if $c < 0$.

Let s denote an arc length parameter for the helix (4), so

$$\frac{ds}{dt} = \left\| \left\langle \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} \right\rangle \right\| = \sqrt{r^2 + c^2}.$$

It is routine to calculate the Frenet frame for this helix. The vectors are

$$\mathbf{T} = \langle -r\sin t, r\cos t, c\rangle / \sqrt{r^2 + c^2},$$

$$\mathbf{N} = \frac{d\mathbf{T}}{dt} \bigg/ \left| \frac{d\mathbf{T}}{dt} \right| = \langle -\cos t, -\sin t, 0\rangle$$

and

$$\mathbf{B} = \mathbf{T} \times \mathbf{N} = \langle c\sin t, -c\cos t, r\rangle / \sqrt{r^2 + c^2}.$$

Next we calculate the curvature and torsion of the helix (4). The curvature is

$$\kappa = \left| \frac{d\mathbf{T}}{ds} \right| = \frac{r}{r^2 + c^2}. \tag{5}$$

And, since

$$\frac{d\mathbf{B}}{ds} = \langle c\cos t, c\sin t, 0\rangle / (r^2 + c^2) = -\frac{c}{(r^2 + c^2)}\mathbf{N} = -\tau\mathbf{N},$$

it is apparent that the torsion is

$$\tau = \frac{c}{r^2 + c^2}. \tag{6}$$

It is worth noting that both the curvature and the torsion are constant. This is a property that characterizes helices, in the sense that a space curve is a (circular) helix if and only if the curvature and torsion are constant.

Observe that the angle θ between the tangent vector $\mathbf{T}$ of the helix and the positive z-axis satisfies

$$\cos\theta = \mathbf{T}\cdot\mathbf{k} = \frac{c}{\sqrt{r^2+c^2}}. \tag{7}$$

It follows that

$$\sin\theta = \frac{r}{\sqrt{r^2+c^2}} = \mathbf{B}\cdot\mathbf{k}, \tag{8}$$

since $\sin\theta$ is positive on the interval $(0, \pi)$. Finally, note that the arc length along the helix between the two points $(r, 0, 0)$ and $(r\cos t, r\sin t, ct)$ is given by

$$\int_0^t \frac{ds}{dt}dt = \sqrt{r^2+c^2}|t|. \tag{9}$$

We actually apply this arc length formula to the orthogonal helix. For the helix (4), the orthogonal helix drawn on the same cylinder and passing through the point $(r, 0, 0)$ is given by

$$\begin{cases} x = r\cos t \\ y = r\sin t \quad t\in\mathbb{R} \\ z = \frac{-r^2}{c}t \end{cases} \tag{10}$$

To see that these two helices ((4) and (10)) are orthogonal at the point $(r, 0, 0)$, just note that the dot product of their tangent vectors is equal to zero.

Developing the HT model Now that we have the necessary background, we can explain the HT model.

THEOREM 1. *Assume we are given the following data: a point $P \in \mathbb{R}^3$, orthogonal unit vectors* $\mathbf{T}$, $\mathbf{N}$, *and* $\mathbf{B} = \mathbf{T}\times\mathbf{N}$, *and real numbers $\kappa > 0$ and τ. Then there is exactly one helix passing through P and having κ as curvature, τ as torsion, and* $\mathbf{T}$, $\mathbf{N}$, *and* $\mathbf{B}$ *as Frenet frame at P.*

Proof. The formulas (5) and (6) for the curvature and torsion of a helix allow us to uniquely choose the radius r of the needed cylinder, and to calculate the needed parameter c. In fact,

$$r = \frac{\kappa}{\kappa^2+\tau^2} \quad \text{and} \quad c = \frac{\tau}{\kappa^2+\tau^2}.$$

Also, from (7) and (8) we find that the angle θ between the tangent vector $\mathbf{T}$ of our intended helix at P and the unknown axis of the cylinder (with a suitable direction vector $\mathbf{A}$) satisfies

$$\cos\theta = \frac{\tau}{\sqrt{\kappa^2+\tau^2}} \quad \text{and} \quad \sin\theta = \frac{\kappa}{\sqrt{\kappa^2+\tau^2}}.$$

To locate the axis of the cylinder, note that the normal vector at any point of a helix is directed towards the axis and is orthogonal to $\mathbf{A}$. Thus, the axis we seek passes through a point that is a distance r from the point P in the direction of $\mathbf{N}$ and satisfies $\mathbf{A}\cdot\mathbf{N} = 0$. Furthermore, (7) and (8) require $\mathbf{T}\cdot\mathbf{A} = \cos\theta$ and $\mathbf{B}\cdot\mathbf{A} = \sin\theta$. It follows that $\mathbf{A} = (\cos\theta)\mathbf{T} + (\sin\theta)\mathbf{B}$, which is uniquely determined by the data. The axis, then, passes through the terminal point Q of the position vector $\overrightarrow{OP} + r\mathbf{N}$, and has direction vector $\mathbf{A}$.

To present a formula for the desired helix, we use the coordinate system with Q as origin and with direction vectors $-\mathbf{N}$, $\mathbf{A} \times (-\mathbf{N})$, and $\mathbf{A}$. These are the appropriate vectors because $\mathbf{A}$ is the direction of the axis of the cylinder (playing the role of the z-axis), and $-\mathbf{N}$ points in the direction of $\overrightarrow{QP}$ (playing the role of the x-axis). The other vector, $\mathbf{A} \times (-\mathbf{N})$, is exactly the one that completes the right-handed system. Now a point $X = (x(t), y(t), z(t))$ of the helix is determined, using the vector form of (4), by

$$\overrightarrow{QX} = (r\cos t)(-\mathbf{N}) + (r\sin t)(\mathbf{A} \times (-\mathbf{N})) + (ct)(\mathbf{A}).$$

It is not difficult to verify that this helix has the stated properties. ■

The construction used in the proof of the theorem is the first step in deriving a formula for the HT ribbon model. Assume that a curve C is given in parametric form by $P(s) = (x(s), y(s), z(s))$, where $0 \leq s \leq l$ and s is arc length. Also, let $D = [0, l] \times [0, w]$ be given. According to Theorem 1, at each point $P(s)$ the curve has the same curvature, torsion, and Frenet frame as a certain, unique helix (which varies with s). As in the proof of the theorem, this helix is on a cylinder whose axis passes through the terminal point Q of the position vector $\overrightarrow{OP} + r\mathbf{N}$, and has direction vector $\mathbf{A} = (\cos\theta)\mathbf{T} + (\sin\theta)\mathbf{B}$. Of course, all the quantities involved in this discussion are functions of s, since they vary along the curve.

Now we need to construct the orthogonal helix at P on the same cylinder. As in the proof of the theorem, we locate the origin at Q and use the vectors $-\mathbf{N}$, $\mathbf{A} \times (-\mathbf{N})$, and $\mathbf{A}$ as the coordinate directions. In these terms and using (10), a point X of the orthogonal helix satisfies

$$\overrightarrow{QX} = (r\cos t)(-\mathbf{N}) + (r\sin t)(\mathbf{A} \times (-\mathbf{N})) + \left(\frac{-r^2}{c}t\right)(\mathbf{A}), \tag{11}$$

where

$$r = \frac{\kappa}{\kappa^2 + \tau^2}, \qquad c = \frac{\tau}{\kappa^2 + \tau^2}, \tag{12}$$

and t is a parameter. Of course, r and c, like κ and τ, are functions of s. We mention this to emphasize that the specific orthogonal helix emanating from the point $P(s)$ varies with s. The shape of the orthogonal helix is governed by r and c, and these parameters, in turn, are determined by the curvature and torsion of the curve at $P(s)$.

The HT model $\mathbf{f} : D \to \mathbb{R}^3$ is characterized by requiring each transversal $v \mapsto \mathbf{f}(s, v)$ to be an arc of the orthogonal helix starting at $P(s)$ and having v as arc length parameter. Thus, each point $(s, v) \in D$ is sent to the point X as in (11), chosen so that the arc length from $P(s)$ to X is equal to v. According to (9), this says that the parameter t in (11) should satisfy

$$|t| = \frac{v}{\sqrt{r^2 + (-r^2/c)^2}} = \frac{v|c|}{r\sqrt{r^2 + c^2}}.$$

We pick

$$t = \frac{-vc}{r\sqrt{r^2 + c^2}} \tag{13}$$

from among the two options. It turns out that with this choice the resulting formula for the HT model agrees with the LT model, when the curve happens to be planar.

Finally (see FIGURE 6), since

$$\overrightarrow{OX} = \overrightarrow{OP} - \overrightarrow{QP} + \overrightarrow{QX} = \overrightarrow{OP} + r\mathbf{N} + \overrightarrow{QX},$$

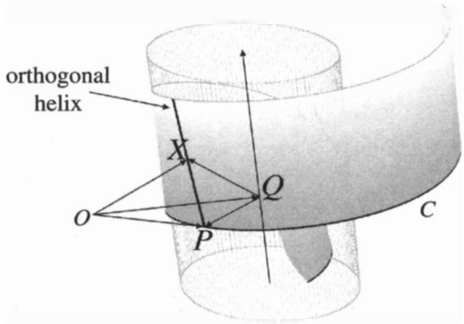


Figure 6 Constructing transversals in the HT model

we can revise (11) and write an explicit formula for the HT ribbon model:

$$\mathbf{f}(s, v) = \langle x(s), y(s), z(s)\rangle + r\mathbf{N} + (r\cos t)(-\mathbf{N}) + (r\sin t)(\mathbf{A}\times(-\mathbf{N})) + \left(\frac{-r^2}{c}t\right)(\mathbf{A}).$$

If we use $\mathbf{A} = (\cos\theta)\mathbf{T} + (\sin\theta)\mathbf{B}$ and simplify, we obtain

$$\mathbf{f}(s, v) = \langle x(s), y(s), z(s)\rangle + \left[r\sin t\sin\theta - \frac{r^2}{c}t\cos\theta\right]\mathbf{T} + [r - r\cos t]\mathbf{N} + \left[-r\sin t\cos\theta - \frac{r^2}{c}t\sin\theta\right]\mathbf{B}.$$

Next, we replace r, c, $\cos\theta$, $\sin\theta$, and t with their equivalent formulas ((12), (7), (8), (13)) involving κ and τ. The result is

$$\mathbf{f}(s, v) = \langle x(s), y(s), z(s)\rangle + \frac{\kappa}{\kappa^2+\tau^2}\left\{\left[\frac{\kappa}{\sqrt{\kappa^2+\tau^2}}\sin\left(\frac{-v\tau\sqrt{\kappa^2+\tau^2}}{\kappa}\right) + v\tau\right]\mathbf{T} + \left(1 - \cos\left(\frac{-v\tau\sqrt{\kappa^2+\tau^2}}{\kappa}\right)\right)\mathbf{N} + \left[\frac{-\tau}{\sqrt{\kappa^2+\tau^2}}\sin\left(\frac{-v\tau\sqrt{\kappa^2+\tau^2}}{\kappa}\right) + v\kappa\right]\mathbf{B}\right\}.$$

Although this expression prominently shows the dependence on κ and τ, we can simplify the appearance of the formula by letting $\rho = \sqrt{\kappa^2+\tau^2}$ and $t = -v\tau\rho/\kappa$ to obtain what we will refer to as the HT model:

$$\mathbf{f}(s, v) = \langle x(s), y(s), z(s)\rangle + \frac{\kappa}{\rho^2}\left\{\left[\frac{\kappa}{\rho}\sin t + v\tau\right]\mathbf{T} + (1-\cos t)\mathbf{N} + \left[\frac{-\tau}{\rho}\sin t + v\kappa\right]\mathbf{B}\right\}. \tag{14}$$

It is not hard to see from this formula that the first partials of **f** exist and are continuous on D. And the construction guarantees that (R1) holds. Thus, the HT model is, indeed, a ribbon model. We explore this matter further in the next section. For now, it is revealing to note that there is a component in the tangent direction—the transversal is not just drawn in the normal plane at each point of the curve as is the case with all of the LT models. We can also see in this formula, that if $\tau \to 0$ then $t \to 0$ and, in the limit, we obtain the LT model (2).

When do the LT and HT models become ribbon maps?

We need to test each model to determine if the ribbon map properties (R2) and (R3) hold. Let's look first at (R2), the property saying that distance is preserved along threads. This is equivalent to requiring that, for each fixed v, the magnitude of $\mathbf{f}_s(s, v)$ is identically equal to 1.

In the case of the LT model, we have

$$\mathbf{f}_s(s, v) = \frac{d}{ds}\{\langle x(s), y(s), z(s)\rangle + v\mathbf{B}(s)\} = \mathbf{T}(s) - v\tau\mathbf{N}(s),$$

and it follows that

$$\mathbf{f}_s \cdot \mathbf{f}_s = 1 + v^2\tau^2. \tag{15}$$

Thus, (R2) fails to hold in the case of the LT model, since $\mathbf{f}_s \cdot \mathbf{f}_s > 1$ (if $v \neq 0$) unless the torsion τ of the curve is zero at every point. This is the same as saying that the curve C must be planar.

On the other hand, the LT model satisfies (R3) for every curve. That is, the transversals are orthogonal to the threads at every point. To see this, we calculate the dot product

$$\mathbf{f}_s \cdot \mathbf{f}_v = (\mathbf{T}(s) - v\tau\mathbf{N}(s)) \cdot \mathbf{B}(s) = 0.$$

The generalized LT models (3) behave in a similar way. In fact, (R2) fails unless $\tau = 0$ but (R3) always holds. For if

$$\mathbf{f}(s, v) = \langle x(s), y(s), z(s)\rangle + v((\cos q)\mathbf{N} + (\sin q)\mathbf{B})$$

then, by the Frenet-Serret formulas (1),

$$\begin{aligned}\mathbf{f}_s(s, v) &= \mathbf{T} + v((-\sin q)q'\mathbf{N} + \cos q(-\kappa\mathbf{T} + \tau\mathbf{B}) + (\cos q)q'\mathbf{B} + (\sin q)(-\tau\mathbf{N}))\\ &= (1 - v\kappa\cos q)\mathbf{T} - v(q' + \tau)(\sin q)\mathbf{N} + v(q' + \tau)(\cos q)\mathbf{B}\end{aligned}$$

and

$$\mathbf{f}_v(s, v) = (\cos q)\mathbf{N} + (\sin q)\mathbf{B}.$$

Thus,

$$\mathbf{f}_s \cdot \mathbf{f}_v = -v(q' + \tau)\sin q\cos q + v(q' + \tau)\cos q\sin q = 0,$$

so (R3) holds. Also, after a little simplification,

$$\mathbf{f}_s \cdot \mathbf{f}_s = (1 - v\kappa\cos q)^2 + v^2(q' + \tau)^2.$$

The only way to make this expression be identically equal to 1 is to have $\cos q = 0$ and $q' + \tau = 0$. But $\cos q = 0$ means that the model is reduced to one of the special LT models in which the transversals are drawn in the direction of $\pm\mathbf{B}$. It follows that q is constant, $q' = 0$, and $\tau = 0$. That is, the curve must once again be planar for (R2) to hold. The conclusion is that the LT model and all the generalized LT models are ribbon maps if and only if the given curve is planar.

Before we begin the discussion of whether the HT function (14) satisfies the requirements to be a ribbon map, it may be useful to make an additional remark for the purposes of comparison. It seems reasonable to say that the LT model is designed for planar curves and can be expected to provide a good approximation to a ribbon map if τ is close to zero (so the curve is close to being planar). Indeed, this is born out by our computation of $\mathbf{f}_s \cdot \mathbf{f}_s$ as in (15). By contrast, the HT model is designed for helices (κ and τ constant) and should be expected to provide a good approximation to a ribbon map if κ' and τ' are close to zero. The next theorem verifies that this is the case.

We need to analyze the partial derivatives of the HT ribbon model $\mathbf{f}$ given by (14), and straightforward (though somewhat arduous) computations yield the expressions below.

One partial derivative is

$$\mathbf{f}_v(s, v) = \delta_1 \mathbf{T} + \delta_2 \mathbf{N} + \delta_3 \mathbf{B},$$

where

$$\delta_1 = \frac{\kappa\tau(1-\cos t)}{\rho^2}, \qquad \delta_2 = \frac{-\tau \sin t}{\rho}, \qquad \text{and} \qquad \delta_3 = \frac{\tau^2 \cos \mathrm{t} + \kappa^2}{\rho^2}. \tag{16}$$

Just as in (14), $\rho = \sqrt{\kappa^2 + \tau^2}$ and $t = -v\tau\rho/\kappa$.

The other partial derivative is

$$\mathbf{f}_s(s, v) = (\alpha_1\kappa' + \beta_1\tau' + \gamma_1)\mathbf{T} + (\alpha_2\kappa' + \beta_2\tau' + \gamma_2)\mathbf{N} + (\alpha_3\kappa' + \beta_3\tau' + \gamma_3)\mathbf{B},$$

where

$$\gamma_1 = 1 + \frac{\kappa^2(\cos t - 1)}{\rho^2}, \qquad \gamma_2 = \frac{\kappa \sin t}{\rho}, \qquad \gamma_3 = \frac{\kappa\tau(1-\cos t)}{\rho^2}, \tag{17}$$

and the alphas and betas are somewhat more complicated functions of the same sorts. They are sums of fractions whose denominators are powers of ρ and whose numerators are polynomials in κ, τ, v, $\sin t$, and $\cos t$. For example,

$$\alpha_1 = \frac{(2\kappa\tau^2 - \kappa^3)\sin t}{\rho^5} + \frac{v(\tau^3 - \kappa^2\tau + \tau^3\cos t)}{\rho^4}.$$

It is clear from the form of all these *coefficient functions* (the alphas, betas, gammas, and deltas), that they are continuous functions on D. This is because κ, and hence ρ, is nonzero for $(s, v) \in D$. It follows that the coefficient functions are bounded on the compact set D. To shorten the statement of the following theorem, we can take advantage of the vector notation $\boldsymbol{\alpha} = \langle \alpha_1, \alpha_2, \alpha_3 \rangle$ and similarly for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$.

THEOREM 2. *Let C be a curve of length l in $\mathbb{R}^3$ and let $D = [0, l] \times [0, w]$. The HT function $\mathbf{f} : D \to \mathbb{R}^3$ given by (14) satisfies the following properties.*

(i) *Distance is preserved along transversals (that is,* (R1) *holds):*

$$\mathbf{f}_v \cdot \mathbf{f}_v = \delta_1^2 + \delta_2^2 + \delta_3^2 = \|\boldsymbol{\delta}\|^2 = 1.$$

(ii) *The ribbon map condition* (R2) *holds if $\kappa' = 0 = \tau'$ and it holds approximately if κ' and τ' are sufficiently small:*

$$\mathbf{f}_s \cdot \mathbf{f}_s = 1 + \|\boldsymbol{\alpha}\|^2(\kappa')^2 + (2\boldsymbol{\alpha}\cdot\boldsymbol{\beta})\kappa'\tau' + \|\boldsymbol{\beta}\|^2(\tau')^2 + (2\boldsymbol{\gamma}\cdot\boldsymbol{\alpha})\kappa' + (2\boldsymbol{\gamma}\cdot\boldsymbol{\beta})\tau',$$

since $\|\boldsymbol{\gamma}\| = 1$.

(iii) *The ribbon map condition* (R3) *holds if $\kappa' = 0 = \tau'$, and it holds approximately if κ' and τ' are sufficiently small:*

$$\mathbf{f}_s \cdot \mathbf{f}_v = (\boldsymbol{\alpha} \cdot \boldsymbol{\delta})\kappa' + (\boldsymbol{\beta} \cdot \boldsymbol{\delta})\tau',$$

since $\boldsymbol{\gamma} \cdot \boldsymbol{\delta} = 0$.

Proof. The proof amounts to using the formulas for the partial derivatives, $\mathbf{f}_v$ and $\mathbf{f}_s$. It is not difficult to verify, using (16) and (17), that

$$\|\boldsymbol{\delta}\|^2 = 1, \quad \|\boldsymbol{\gamma}\| = 1, \quad \text{and} \quad \boldsymbol{\gamma} \cdot \boldsymbol{\delta} = 0. \qquad \blacksquare$$

It should be recognized that we are not apt to apply Theorem 2 by letting κ' and τ' tend to zero. Since the curve C would typically be given as part of the data, the functions κ and τ, along with their derivatives, are already determined and fixed. Qualitatively, the theorem just confirms that the HT model is apt to perform best when the curvature and torsion of the curve are changing gradually.

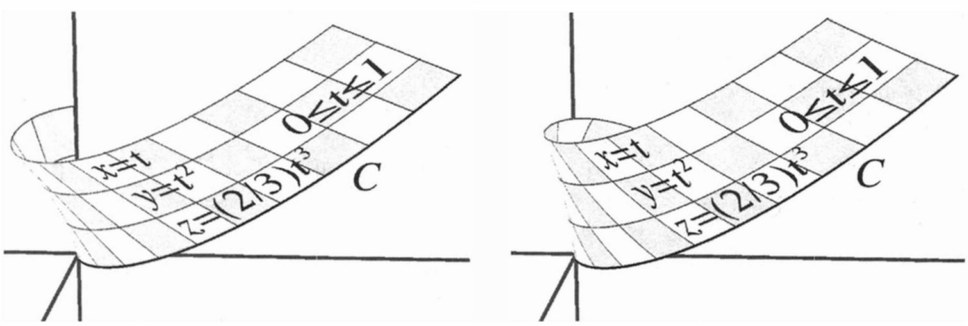


Figure 7 LT model (left) and HT model (right)

In closing, we return to the empirical aspect of this topic—after all, the HT and LT models actually do provide pictures. What may be most striking about the side-by-side pictures obtained from the two models is how similar they are. Given the relatively complicated derivation of the HT model, we may have expected more dramatic differences between the two. The fact is, however, that the transversals in both models start out in the direction of the same binormal vector **B**. In cases where the models are noticeably different, it is an indication of the effect of torsion in the given curve. As a final example, FIGURE 7 is offered as a suggestion of how a curve can be illustrated using a ribbon to provide some added depth and interest. The curve in this case is given by

$$\begin{cases} x = t \\ y = t^2 \\ z = (2/3)t^3 \end{cases} \quad t \in [0, 1], \tag{18}$$

and has the nice property that the torsion and curvature are the same. In fact,

$$\tau = \frac{2}{(1 + 2t^2)^2} = \kappa.$$

Apparently, the torsion decreases by a factor of 9, from $\tau = 2$ at one endpoint to $\tau = 2/9$ at the other. And perhaps the HT model does a better job of signaling this change. Recall that in the HT model each transversal is plotted along a different helix. And each helix lies on a cylinder of varying radius, given in this example by $r = \kappa/(\kappa^2 + \tau^2) = 1/(2\tau)$. The relatively small radius when $t = 0$, one ninth of the radius when $t = 1$, causes the distinctive curl in the HT ribbon. This may not be a dramatic improvement over the result of the LT model, but the curled edge is eye-catching and directs our attention to the part of the curve where torsion is high.

Acknowledgment. I wish to thank the referees for suggesting many improvements.

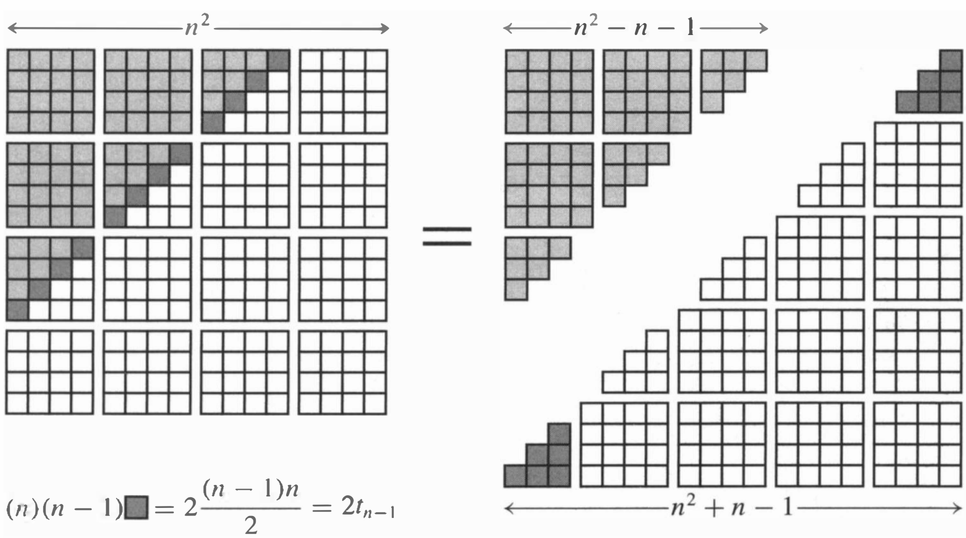
REFERENCES

1. Mike Carson, Ribbons 2.0, *J. Appl. Cryst.* **24** (1994), 958–961.
2. Joseph McHugh, The osculating spiral, this MAGAZINE, **61** (1988), 106–113.
3. John Oprea, *Differential Geometry and Its Applications*, 2nd Ed., Pearson Prentice Hall, 2004.
4. James Stewart, *Multivariable Calculus Early Transcendentals*, 5th Ed., Brooks/Cole, 2003.

NOTES

Proof Without Words: Every Fourth Power Greater than One Is the Sum of Two Nonconsecutive Triangular Numbers

$$t_k = 1 + 2 + \cdots + k \Rightarrow 2^4 = 15 + 1 = t_5 + t_1,$$
$$3^4 = 66 + 15 = t_{11} + t_5,$$
$$4^4 = 190 + 66 = t_{19} + t_{11},$$
$$\vdots$$
$$n^4 = t_{n^2+n-1} + t_{n^2-n-1}.$$



Note: Since $k^2 = t_{k-1} + t_k$, we also have $n^4 = t_{n^2-1} + t_{n^2}$.

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

Another Step Further... On a Problem of the 1988 IMO

ISTVÁN G. LAUKÓ
GABRIELLA A. PINTÉR
University of Wisconsin, Milwaukee
Milwaukee, WI 53201-0413
iglauko@uwm.edu
gapinter@uwm.edu

LAJOS PINTÉR
Bolyai Institute
University of Szeged
Szeged, Hungary 6720
pinter@math.u-szeged.hu

Unexpected depth and complexity can arise from problems that are stated surprisingly simply. One valuable tool is simplifying even more and analyzing special cases. On the other hand, generalizing a problem whose solution is well understood can be just as useful. The problem we analyze in this paper provides a case study for both activities. We hope that our discussion, with its open questions and proposed connections, allows the reader to take yet another step forward.

Our starting point is a problem on the 1988 International Mathematical Olympiad (IMO). It is stated in the following way:

P1 *If* a, b, *and* $q = \dfrac{a^2+b^2}{1+ab}$ *are nonnegative integers, then* q *is a perfect square.*

We note that there is an account of the problem's interesting IMO history in A. Engel's book titled *Problem-Solving Strategies* [**1**]:

> The ... problem was submitted by the FRG [Federal Republic of Germany]. Nobody of the six members of the Australian problem committee could solve it.... Since it was a number theoretic problem it was sent to four most renowned Australian number theorists. They were asked to work on it for six hours. None of them could solve it in this time. The problem committee submitted it to the jury of the XXIX. IMO marked with a double asterisk, which meant a superhard problem, possibly too hard to pose. After a long discussion, the jury finally had the courage to choose it as the last problem of the competition. Eleven students gave perfect solution.

Engel [**1**] gives two solutions to our problem. The first, in Chapter 6 (*Number Theory*, E15, p. 127), uses the idea that there is a lattice point on the hyperbola $x^2+y^2-qxy-q=0$ if and only if q is a perfect square. The second solution can be found in Chapter 8, (*The Induction Principle*, Problem 6, p. 211). It follows from this solution that if there exist integers a, b, and q satisfying $q=(a^2+b^2)/(1+ab)$, then $q=(\gcd(a,b))^2$. In this paper we investigate the problem further, with a bit of help from a computer, and account for all integers a and b for which q is an integer.

Using a simple computer code to generate solution pairs (a, b) satisfying the requirement, we discover recursive relationships within the solution set. We investigate these relationships further, and effectively find the structure of the set of all solutions and so obtain a new way to solve the original problem. The recursive relationships are reminiscent of that of the Fibonacci sequence; indeed, several Fibonacci-like properties of these solutions can be established. Similar properties are shared by a whole family of sequences defined by recurrence relations, namely, the generalized Fibonacci sequences studied by Kalman and Mena [**2**].

Throughout the article, there are a number of exercises posed for the interested reader. These are typically easy to complete based on the demonstrated techniques. The last section contains suggestions for generalization, as well as questions of what interesting structural properties can be established for the extended solution set.

Discovering solutions and structures It is certainly not difficult to come up with integer pairs (a, b) for which q in **P1** is an integer. We will refer to such pairs in the future as solutions to the problem **P1**. Since the problem is symmetric in a and b we would rather consider nonordered pairs, and for simplicity we use the notation (a, b) with the understanding that $a \leq b$. We can easily point out the trivial solution $(a, b) = (0, 0)$ with $q = 0$. Moreover, if $a = 0$, then $q = (a^2 + b^2)/(1 + ab) = b^2$, and we generate an infinite list of solutions

$$(0, 0), (0, 1), (0, 2), (0, 3) \ldots$$

with q-values $0^2, 1^2, 2^2, 3^2, \ldots$. Furthermore, if $a > 0$ is an integer and $b = a^3$, then

$$q = \frac{a^2 + b^2}{1 + ab} = \frac{a^2 + a^6}{1 + aa^3} = a^2 \frac{1 + a^4}{1 + a^4} = a^2,$$

which means that the pairs

$$(1, 1^3), (2, 2^3), (3, 3^3), (4, 4^3), \ldots$$

are also solutions with corresponding q-values $1^2, 2^2, 3^2, 4^2, \ldots$. Let us consider another simple case, and assume that $a = b \neq 0$. In this case

$$q = \frac{a^2 + a^2}{1 + a^2} = \frac{a^2 + 1 - 1 + a^2}{1 + a^2} = 1 + \frac{a^2 - 1}{a^2 + 1},$$

and therefore $(a, b) = (1, 1)$ is a solution with $q = 1$; we also see that there are no solutions with $a = b > 1$.

A natural question can be posed here: Are there any other solutions with $q = 1$? Since $q = 1$ implies $1 = (a^2 + b^2)/(1 + ab)$, or $a^2 + b^2 = ab + 1 \leq b^2 + 1$, we conclude that $a = 0$ or $a = 1$. If $a = 0$, then $b^2 = 1$, which yields $b = 1$. If $a = 1$, then $1 = (1 + b^2)/(1 + b)$, which implies $b^2 = b$ and in turn $b = 1$ (since $1 = a \leq b$). So, not unexpectedly, the answer to our question is no, the only solutions corresponding to $q = 1$ are $(0, 1)$ and $(1, 1)$.

We easily found a good number of solutions (a, b) to **P1**, but are there more? Many more? Experimentation with a doubly nested loop, presented in pseudo-code below, helped us to get additional pairs and, ultimately, insight into the structure of all solutions.

```
ABQs=[]                      %to store the a,b,q solution triples
 for a=0 to 1000             %scan through integers for a
   for b=a to 1000           %scan through integers for b
      q=(a^2+b^2)/(a*b+1)    %compute q
      if integer(q)          %if q is integer
        store([a,b,q],ABQs)  %store solution triples in ABQs
      end
    end
 end
```

If we use this code to search for solutions with either component between 1 and 1000, we find the following list:

$$(1,1),(2,8),(3,27),(4,64),(5,125),(6,216),(7,343),(8,30),$$
$$(8,512),(9,729),(10,1000),(27,240),(30,112),(112,418).$$

The values assumed by q for these pairs are $1, 2^2, 3^2, \ldots, 10^2$. Some of these solutions are new, not having the form (a, a^3). These are (8, 30), (27, 240), (30, 112), and (112, 418), with corresponding qs, 4, 9, 4,and 4. Let us list the solution pairs we know for $q = 4$: (0, 2), (2, 8), (8, 30), (30, 112), (112, 418). Here is a pattern, a chain! Does it continue?

With a bit of effort we can notice another pattern, a recursion:

$$8 = 4\cdot 2 - 0; \quad 30 = 4\cdot 8 - 2; \quad 112 = 4\cdot 30 - 8; \quad 418 = 4\cdot 112 - 30.$$

Based on this observation we suspect that the next solution pair is (418, 1560), since $1560 = 4\cdot 418 - 112$, and indeed, $(418^2 + 1560^2)/(1 + 418\cdot 1560) = 4$. Our conjecture in general terms is the following:

If $a_0 = 0$, $a_1 = 2$, and $a_{n+1} = 4a_n - a_{n-1}$ for $n = 1, 2, 3\ldots$, then a_{n+1} is a positive integer, $a_n < a_{n+1}$ and $(a_n^2 + a_{n+1}^2)/(1 + a_n a_{n+1}) = 4$ for $n = 0, 1, 2, 3\ldots$.

The first two parts follow quickly from the recursion equation. We prove the essential part of our statement by induction, using a simple property of proportions: if A, B, C, D are positive real numbers and $B \neq D$, then $A/B = (A - C)/(B - D)$ exactly when $A/B = C/D$. To begin the induction, observe that the statement holds for $n = 0$:

$$\frac{a_n^2 + a_{n+1}^2}{1 + a_n a_{n+1}} = \frac{0^2 + 2^2}{1 + 0\cdot 2} = 4.$$

Let us assume that $(a_{n-1}^2 + a_n^2)/(1 + a_{n-1}a_n) = 4$ (with $n \geq 1$) and try to show that $(a_n^2 + a_{n+1}^2)/(1 + a_n a_{n+1}) = 4$. Since

$$4 = \frac{4a_n}{a_n} = \frac{a_{n+1} + a_{n-1}}{a_n} = \frac{a_{n+1}^2 - a_{n-1}^2}{a_n(a_{n+1} - a_{n-1})}$$
$$= \frac{a_n^2 + a_{n+1}^2 - (a_{n-1}^2 + a_n^2)}{1 + a_n a_{n+1} - (1 + a_{n-1}a_n)} = \frac{a_{n-1}^2 + a_n^2 - (a_n^2 + a_{n+1}^2)}{1 + a_{n-1}a_n - (1 + a_n a_{n+1})},$$

with $C = a_n^2 + a_{n+1}^2$, $D = 1 + a_n a_{n+1}$, $A = a_{n-1}^2 + a_n^2$ and $B = 1 + a_{n-1}a_n$ we have

$$\frac{a_n^2 + a_{n+1}^2}{1 + a_n a_{n+1}} = \frac{a_{n-1}^2 + a_n^2}{1 + a_{n-1}a_n} = 4.$$

Thus the sequence of solutions (0, 2), (2, 8), (8, 30), (30, 112), (112, 418) can be continued, (a_{n-1}, a_n) is followed by (a_n, a_{n+1}), where $a_{n+1} = 4a_n - a_{n-1}$. Let us call this sequence of pairs the *chain corresponding to* $q = 4$. Alternatively, in the following we will also consider the corresponding sequence 0, 2, 8, 30, 112, ... $(a_{n+1} = 4a_n - a_{n-1})$.

EXERCISE 1. *Generalize the case $q = 2^2$ by showing that for all $q = k^2$, with $k > 2$, there corresponds an infinite chain: $(b_0, b_1), (b_1, b_2), \ldots (b_{n-1}, b_n), (b_n, b_{n+1}), \ldots$ where $b_0 = 0$, $b_1 = k$, and $b_{n+1} = k^2 b_n - b_{n-1}$ for all $n \geq 1$ and, furthermore, $(b_n^2 + b_{n+1}^2)/(1 + b_n b_{n+1}) = k^2$ for $n = 0, 1, \ldots$.*

We list a few of these solutions chains:

$$
\begin{aligned}
&(0,0),\ (0,0),\ (0,0),\ \ldots\\
&(0,1),\ (1,1),\ (1,1),\ \ldots\\
&(0,2),\ (2,8),\ (8,30),\ (30,112),\ \ldots \qquad (1)\\
&(0,3),\ (3,27),\ (27,240),\ (240,2133),\ \ldots,\ \text{and, in general}\\
&(0,k),\ (k,k^3),\ (k^3,k^5-k),\ \ldots.
\end{aligned}
$$

Let us summarize what we have obtained so far. The known solutions are contained in chains $(a_0, a_1), (a_1, a_2), \ldots, (a_{n-1}, a_n), \ldots$, given by the recursive formulas $a_0 = 0$, $a_1 = k$, and $a_{n+1} = k^2 a_n - a_{n-1}$ for $k = 0, 1, 2, \ldots$; in each such chain the corresponding q is $q = k^2$. The first two chains, corresponding to $q = 0^2$ and $q = 1^2$, are periodic, producing finitely many solutions, while the rest provide infinitely many.

All solutions found It is not clear at this point whether these chains contain *all* solutions, but they certainly include all pairs that we were able to detect computationally, no matter how far we tried. Now we show that they indeed include all solutions.

Assume that a is a positive integer for which there exists a corresponding nonnegative integer x so that $(a^2 + x^2)/(1 + ax) = m$, where $m > 1$ is an integer (we have already seen the cases $m = 1$ and $m = 0$). Note that we do not know at this point whether m is a perfect square. Now, $x^2 - amx + a^2 - m = 0$ has two real solutions $x_{1,2} = (am \pm \sqrt{a^2m^2 - 4(a^2 - m)})/2$, since $a^2(m^2 - 4) + 4m \geq 4m > 0$. The integer x must coincide with either x_1 or x_2, and since $x_1 + x_2 = am$, it follows that both x_1 and x_2 are integers. We now show that $x_1 > a$ and $a > x_2 \geq 0$. It is clear that

$$x_1 = \frac{am + \sqrt{a^2m^2 - 4(a^2 - m)}}{2} > \frac{am}{2} \geq a.$$

On the other hand, $(x_2^2 + a^2)/(1 + x_2 a) = m$ implies $ax_2 + 1 > 0$, which in turn implies $x_2 > -1/a$. Thus $x_2 \geq 0$, since a is a positive integer and x_2 is also known to be an integer. Also, we assert that

$$x_2 = \frac{am - \sqrt{a^2m^2 - 4(a^2 - m)}}{2} < a.$$

This holds since it follows from $m \geq 2$ that $8a^2 < 4a^2m + 4m$, and adding $a^2m^2 - 4a^2m - 4a^2$ to each side gives $a^2(m-2)^2 = a^2m^2 - 4a^2m + 4a^2 < a^2m^2 - 4(a^2 - m)$, which implies

$$a(m-2) < \sqrt{a^2m^2 - 4(a^2 - m)}.$$

Now we start going backwards to discover a known chain behind us: For the (x_2, a) pair we have that $(x_2^2 + a^2)/(1 + x_2 a) = m$ and $0 \leq x_2 < a$. If $x_2 = 0$ then

$$\frac{x_2^2 + a^2}{1 + x_2 a} = a^2 = m.$$

If $x_2 > 0$ then x_2 will take over the role of a. Using the ideas above we obtain that there exists a nonnegative integer x_3 such that $0 \leq x_3 < x_2$ and

$$\frac{x_3^2 + x_2^2}{1 + x_3 x_2} = m.$$

Repeating this procedure finitely many times we must get to zero, that is, there exists n so that $a > x_2 > x_3 > \cdots > x_{n-1} > x_n = 0$ and

$$\frac{x_i^2 + x_{i-1}^2}{1 + x_i x_{i-1}} = m$$

for $i = 2, \ldots, n$. In particular,

$$\frac{0^2 + x_{n-1}^2}{1 + 0 \cdot x_{n-1}} = x_{n-1}^2 = m.$$

We have gained two results at once: On one hand m is proved to be a perfect square, on the other hand, the original solution pair (a, x) must be in the chain corresponding to the perfect square m, since any value in a solution pair corresponding to m uniquely determines the *forward* and *backward* chain links (because of the quadratic equation). Thus we proved that all pairs of nonnegative integers (a, b) such that $(a^2 + b^2)/(1 + ab)$ is an integer are members of chains in (1).

Moreover, we note that, for $m > 0$, the recursive formula implies that $\gcd(a_{n-1}, a_n) = \gcd(a_n, a_{n+1})$, since it follows from $a_{n+1} = m^2 a_n - a_{n-1}$ that $\gcd(a_{n-1}, a_n)$ divides a_{n+1} and $\gcd(a_n, a_{n+1})$ divides a_{n-1} for $n = 2, 3, \ldots$. Thus when $(a_1, a_2) = (k, k^3)$, $q = k^2 = (\gcd(a_n, a_{n+1}))^2$, as we remarked earlier.

Fibonacci-like properties We can see from Exercise 1 that the sequences in (1) arise from second-order linear recurrence relations similar to that of the famous Fibonacci sequence, $f_{n+1} = f_n + f_{n-1}$ for $n \geq 1$, with $f_0 = 0$ and $f_1 = 1$. A number of properties of the Fibonacci sequence also hold for the sequences $\{b_n\}$ as defined in Exercise 1.

Indeed, in the June 2003 issue of the MAGAZINE [**2**], Kalman and Mena "expose" the Fibonacci recurrence as just one example of a general second-order linear recurrence relation; the popular properties of the Fibonacci numbers are shared by every sequence generated by such a recurrence. Rather than prove all these properties again, we list some of them here as exercises, referring the reader to the Kalman and Mena article for more details.

EXERCISE 2. **General Binet formula** *Show that the terms of the sequence $\{b_n\}$ formed from the chain corresponding to k^2, (that is, $b_0 = 0$, $b_1 = k$, and $b_{n+1} = k^2 b_n - b_{n-1}$ for $n \geq 1$), satisfy*

$$b_n = \frac{k}{\sqrt{k^4 - 4}} \left(\left(\frac{k^2 + \sqrt{k^4 - 4}}{2} \right)^n - \left(\frac{k^2 - \sqrt{k^4 - 4}}{2} \right)^n \right), \qquad n = 0, 1, 2, \ldots.$$

A formula for the sum of the first n Fibonacci numbers is easy to find. Let us compute the partial sums $s_n = a_0 + a_1 + \cdots + a_n$ for the sequence defined by $a_0 = 0$, $a_1 = 2$, and $a_{n+1} = 4a_n - a_{n-1}$. We sum the equations

$$a_2 = 4a_1 - a_0, \quad a_3 = 4a_2 - a_1, \quad \ldots \quad a_{n+1} = 4a_n - a_{n-1},$$

to obtain

$$s_n + a_{n+1} - a_1 - a_0 = 4s_n - 4a_0 - s_n + a_n,$$

which yields $s_n = (a_{n+1} - a_n)/2 - 1$.

EXERCISE 3. *Show that for the sequence $b_0 = 0$ and $b_1 = k$, $b_{n+1} = k^2 b_n - b_{n-1}$ for $n \geq 1$, we have*

$$s_n = b_0 + b_1 + \cdots + b_n = \frac{1}{k^2 - 2}(b_{n+1} - b_n - k).$$

EXERCISE 4. *Show that for the sequence $b_0 = 0$ and $b_1 = k$, we have $b_{n+1} = k^2 b_n - b_{n-1}$ for $n \geq 1$*

$$\begin{pmatrix} k^2 & 1 \\ -1 & 0 \end{pmatrix}^n = \frac{1}{k} \begin{pmatrix} b_{n+1} & b_n \\ -b_n & -b_{n-1} \end{pmatrix}.$$

EXERCISE 5. *Show that for the sequence $\{b_n\}$*

$$b_{n+m} = \frac{1}{k}(b_{n+1} b_m - b_n b_{m-1}) \qquad \text{for } n, m \geq 1.$$

EXERCISE 6. *Is it true that $b_n \mid b_{pn}$ for $n, p \geq 1$?*

The property $\det(AB) = \det A \cdot \det B$ implies that $\det(A^n) = (\det A)^n$ for $n \in \mathbb{N}^+$, where $\det A$ denotes the determinant of the matrix A. Thus we see that, for

$$R = \frac{1}{2} \begin{pmatrix} a_2 & a_1 \\ -a_1 & -a_0 \end{pmatrix},$$

we have

$$\begin{aligned} 1 &= (\det R)^n = \det(R^n) \\ &= \det\left(\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} a_{n+1} & a_n \\ -a_n & -a_{n-1} \end{pmatrix} \right) = \frac{1}{4} \det \begin{pmatrix} a_{n+1} & a_n \\ -a_n & -a_{n-1} \end{pmatrix}, \end{aligned}$$

that is,

$$a_{n+1} a_{n-1} = a_n^2 - 4, \qquad n \in \mathbb{N}^+.$$

Hence it follows that $a_{n+1} \mid a_n^2 - 4$ and $a_{n-1} \mid a_n^2 - 4$.

EXERCISE 7. *Show that for the sequence $\{b_n\}$ with $n \geq 1$*

$$b_{n+1} b_{n-1} = b_n^2 - k^2 \qquad n \in \mathbb{N}^+.$$

EXERCISE 8. *Show that the last digits of the terms in the sequence corresponding to any chain form a periodic sequence. We can also form a periodic sequence from the last digits of the sequence of partial sums $\{s_n\}$.*

One step toward generality Suppose we drop the nonnegativity condition for a, b in the original problem and consider a more general one:

P2 *Find integers a, b, so that $\dfrac{a^2 + b^2}{1 + ab} = q$ is an integer.*

It is quite clear that the linear recurrence relations $a_{n+1} = q a_n - a_{n-1}$ (or $a_{n-1} = q a_n - a_{n+1}$) would generate solution pairs in the negative direction as well, and that the chains of solution pairs, formed from adjacent terms in the doubly infinite sequences below, would continue in a symmetric fashion about 0.

$$\ldots, 0, 0, 0, 0, 0, \ldots$$
$$\ldots, -1, -1, -1, 0, 1, 1, 1, \ldots$$
$$\ldots, -112, -30, -8, -2, 0, 2, 8, 30, 112, \ldots \tag{2}$$
$$\ldots, -240, -27, -3, 0, 3, 27, 240, \ldots, \text{ and, in general,}$$
$$\ldots, -(k^5 - k), -k^3, -k, 0, k, k^3, k^5 - k, \ldots.$$

But would other kinds of solutions appear? The only possibility for this is that a and b are opposite in sign and $q < 0$. By changing the ranges of the loops for a and b into negative integers in our pseudo-code, our findings show that there seems to be only one negative q corresponding to new solutions, $q = -5$ that is. Also those solutions can be pieced together into a pair of chains (negatives of one another) associated with the doubly infinite sequences emanating from the solutions $(a, b) = (-3, 1)$ or $(1, -3)$. The two chains,

$$\ldots, 321, -67, 14, -3, 1, -2, 9, -43, 206, \ldots$$
$$\ldots, -206, 43, -9, 2, -1, 3, -14, 67, -321 \ldots \tag{3}$$

correspond to $q = -5$ and are generated by the recurrence $a_{n+1} = -5a_n - a_{n-1}$. Clearly, the second can be recovered from the first by pivoting around 1 and changing the signs (and vice versa). We set out to prove a new conjecture:

If a, b are integers, $ab < 0$, and $(a^2 + b^2)/(1 + ab) = q$ is a negative integer, then $q = -5$, and all solutions are formed from adjacent pairs in the two sequences in (3).

We can observe the symmetries in (3) and reformulate this problem by setting $m = -q$ and switching the sign of the negative member in each (a, b) pair (or every other term in the sequences in (3)). The equivalent formulation of the conjecture obtained this way is the following:

If a, b, and $(a^2 + b^2)/(ab - 1) = m$ are positive integers then $m = 5$. Furthermore, all solutions pairs are adjacent members of two sequences generated from $(a_1, a_2) = (1, 2)$ and $(a_1, a_2) = (1, 3)$ by the recursion $a_{n+1} = 5a_n - a_{n-1}$ $(n = 2, 3, \ldots)$.

If one member of a solution pair (a, b), say a, is 1, then $(b^2 + 1)/(b - 1) = m$, which implies that the two values for b are

$$b_{1,2} = \frac{m \pm \sqrt{m^2 - 4m - 4}}{2}.$$

Thus $b_1 + b_2 = m$ and both roots must be integers. This in turn yields that $m^2 - 4m - 4 = p^2$, where p is a positive integer. A short calculation shows that the only integer solution to this equation is $p = 1$, which implies that $m = 5$, and $b = 2$ or $b = 3$ are the two solutions when $a = 1$.

Next assume that x is a positive integer solution to the equation

$$\frac{a^2 + x^2}{ax - 1} = m,$$

where a and m are positive integers. Necessarily, $m \geq 3$, since $a^2 + x^2 \geq 2ax > max - m$ if $m = 1$ or $m = 2$. We further assume $a \geq 2$ (the case $a = 1$ has been treated above). Then $(a^2+x^2)/(ax-1)=m$ implies $x^2-max+a^2+m=0$, so the possible values of x are

$$x_{1,2}=\frac{ma\pm\sqrt{m^2a^2-4a^2-4m}}{2}.$$

Since $x_1+x_2=ma$ is an integer and one of the solutions x_1 or x_2 is an integer, so is the other. Let $x_1\geq x_2$, then $x_1\geq ma/2>a$ (since $m\geq 3$). On the other hand, $x_1x_2=a^2+m>0$ implies $x_2>0$. We now show that $x_2<a$, or $ma-\sqrt{m^2a^2-4a^2-4m}<2a$, which is equivalent to

$$a(m-2)<\sqrt{m^2a^2-4a^2-4m},\ \text{equivalently}\ 8a^2<4m(a^2-1).$$

This is true, because $a\geq 2$ implies $8a^2<12(a^2-1)$, and, if also $m\geq 3$, we conclude

$$8a^2<12(a^2-1)\leq 4m(a^2-1).$$

We have shown that there is an integer $0<x_2<a$ so that

$$\frac{a^2+x_2^2}{ax_2-1}=m.$$

If $x_2=1$, then our previous argument applies: $a=2$ or $a=3$ and $m=5$. If $x_2>1$, then x_2 takes the role of a and we can repeat the procedure above to obtain a solution x_3 satisfying $0<x_3<x_2$ and

$$\frac{x_1^2+x_2^2}{x_1x_2-1}=\frac{x_2^2+x_3^2}{x_2x_3-1}=m.$$

In a similar fashion we obtain the finite sequence $x_1>x_2>\cdots>x_{k-1}>x_k$ corresponding to the same m, and in finitely many steps we reach $x_k=1$. This in turn shows that $x_{k-1}=2$ or $x_{k-1}=3$ and $m=5$ is the only possibility for solutions to exist. We can see now that all solution pairs to the reformulated version of problem **P2** are adjacent pairs from two different infinite sequences starting with $(a_1,a_2)=(1,2)$ or $(a_1,a_2)=(1,3)$, respectively.

It can be easily seen that these two sequences are in fact $a_{n+1}=5a_n-a_{n-1}$ with $(a_1,a_2)=(1,2)$ and $(a_1,a_2)=(1,3)$. Indeed, take two adjacent solution pairs (a_{n-1},a_n), (a_n,a_{n+1}) with $n>1$ from one of the two sequences generated by the recursion above. We subtract the two equalities

$$a_n^2+a_{n+1}^2=5(a_na_{n+1}-1)\ \text{and}\ a_{n-1}^2+a_n^2=5(a_{n-1}a_n-1)$$

to obtain $a_{n+1}^2-a_{n-1}^2=5a_n(a_{n+1}-a_{n-1})$, which implies $a_{n+1}+a_{n-1}=5a_n$, since $a_{n+1}>a_{n-1}$. This proves the reformulated conjecture and establishes the description of all solutions of **P2** given in (2) and (3).

More challenges needed? We now are confident that we have a very good understanding of the solutions and their relationships for problem **P1** and its generalization **P2**. Of course, if the reader is hungry for more challenges we can always widen the scope of our investigation. Problem **P2** is a special case of the following problem:

P3 *Find integers a, b, and c so that* $\dfrac{a^2+b^2+c^2}{1+ab+bc+ac}=q$ *is an integer.*

Clearly, there are a number of questions to ask about the solutions to **P3**:

What can we discover about the solutions to **P3** numerically?
Are there new kinds of recursive relations among the solutions?
Do we have closed forms for the solutions?
How do the new solutions (if any) enrich the structure that we have unveiled?
Do they form new linearly ordered subsets or would a more complicated ordering appear?
Would there still be only finitely many negative values for q?
Can we prove our numerical discoveries?

We invite the interested reader to join the investigation.

Acknowledgment. The authors are grateful to J. Kincses and L. Megyesi, as well as the anonymous reviewers for their interest and valuable suggestions.

REFERENCES

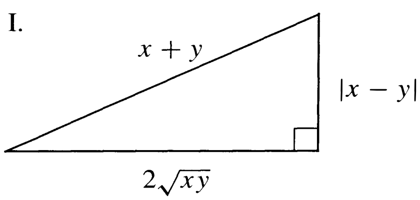
1. A. Engel, *Problem Solving Strategies*, Springer-Verlag, New York, 1997.
2. D. Kalman and R. Mena, The Fibonacci Numbers—Exposed, this MAGAZINE **76** (2003), 167–181.

Proof Without Words: Padoa's Inequality (Alessandro Padoa, 1868–1937)

If a, b, c are the sides of a triangle, then

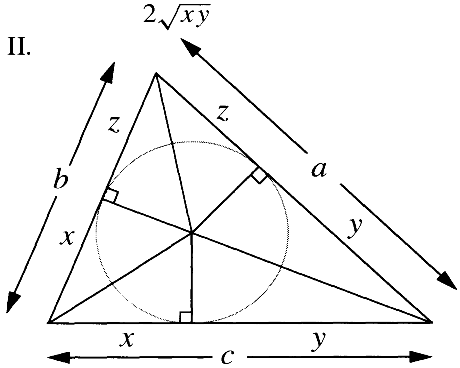
$$abc \geq (a+b-c)(b+c-a)(c+a-b).$$

I.



$$x + y \geq 2\sqrt{xy}$$

II.



$$\begin{aligned} abc &= (y+z)(z+x)(x+y) \\ &\geq 2\sqrt{yz}\cdot 2\sqrt{zx}\cdot 2\sqrt{xy} \\ &= (2z)(2x)(2y) \\ &= (a+b-c)(b+c-a)(c+a-b) \end{aligned}$$

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

The Cross Ratio *Is* the Ratio of Cross Products!

LEAH WRENN BERMAN
GORDON IAN WILLIAMS
Ursinus College
Collegeville, PA 19426
lberman@ursinus.edu
gwilliams@ursinus.edu

BRADLEY JAMES MOLNAR
11653 Gas Light Court #F
Reston, VA 20190
brad@llanos.us

On a recent geometry exam at Ursinus College, students were asked to find the cross ratio of four collinear points in the real projective plane. One of the students forgot how to do the problem and thought to himself, "I know that the cross ratio must involve a ratio...it must be a ratio of...hmm...cross products!" and proceeded to calculate a "ratio" of cross products. Imagine the surprise of the authors when all the numbers worked out.

Cross products are a familiar computation from calculus: The cross product of vectors **a** and **b** gives a vector normal to the plane spanned by **a** and **b** whose magnitude is determined by their lengths and the angle between them. However, the cross ratio, while quite useful both in classical Euclidean geometry and in projective geometry, is probably less familiar. In Euclidean geometry, the cross ratio of four collinear points is a "ratio of ratios," computed by looking at signed distances: if A, B, C, and D are all collinear, then the cross ratio is defined to be

$$(ABCD) = \frac{|AC|}{|CB|} \Big/ \frac{|AD|}{|DB|},$$

where distances are assigned with respect to some fixed choice of orientation on the line, so that for example, $|AD| = -|DA|$ (see Figure 1). One interesting property of the cross ratio is that it is independent of the orientation assigned to the line.

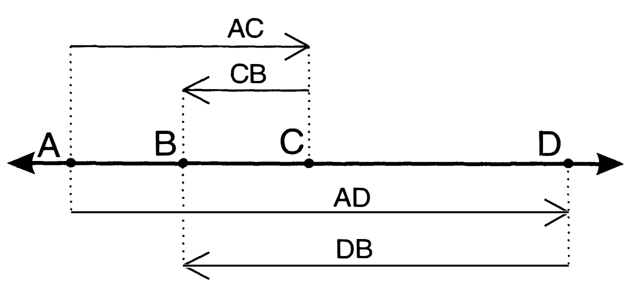


Figure 1 Signed distances for four collinear points; leftward pointing arrows correspond to negative distances

The cross ratio can be quite useful. For example, one can use the cross ratio to determine actual distances from aerial photographs. Suppose an aerial photograph is taken of a road with three objects that are at known distances apart from each other, and a car whose relative position is uncertain. Using measurements taken from the photograph of the images of the objects and the car, the known distances between the objects, and the fact that the cross ratio of the four points (the objects and the car) in the photograph is equal to the cross ratio of the analogous four points on the ground, it is possible to determine how far the car is from the objects. The cross ratio also may be used to provide alternate, slick proofs for classical results in geometry such as Pappus' theorem.

The real projective plane The real projective plane, $\mathbb{RP}^2$, has a somewhat counterintuitive but useful geometry. In $\mathbb{RP}^2$, the "points" of the geometry are lines through the origin in $\mathbb{R}^3$, and the "lines" of the geometry are planes through the origin in $\mathbb{R}^3$. A projective point A in $\mathbb{RP}^2$ may be represented using homogeneous coordinates $[a_1, a_2, a_3]$, where (a_1, a_2, a_3) is any nonzero vector lying along the line in $\mathbb{R}^3$ that defines A; two homogeneous coordinates $[a_1, a_2, a_3]$ and $[\alpha_1, \alpha_2, \alpha_3]$ represent the same projective point if and only if one is a nonzero scalar multiple of the other. While this formulation is computationally efficient, it may be hard to visualize.

One way to visualize $\mathbb{RP}^2$ is to imagine positioning your eye at the origin in $\mathbb{R}^3$; then two points lying on a ray through the origin are indistinguishable. You may choose as representative projective points those points on the rays through the origin that are at unit distance from the origin, that is, points that lie on the unit sphere centered at the origin. Projective lines, which correspond to planes through the origin in $\mathbb{R}^3$, become great circles on the unit sphere. In this model, projective geometry may be thought of as the geometry of what one can distinguish from a viewpoint at the origin. The only complication with this presentation of the spherical model of $\mathbb{RP}^2$ is that there is one more identification of points in the projective plane than has occurred by collapsing all the points we see from the origin to points on the unit sphere, namely that antipodal points of the sphere must also be identified. Thus, we have the rather comical image of a viewer who can't distinguish her front from her back—but this is a consequence of the fact that the real projective plane is a nonorientable surface.

The projective plane shares many properties with the ordinary Euclidean plane, but it also has fundamentally different properties (Brannan, et. al. [**1**], Eves [**5, 6**] or Seidenberg [**13**] provide thorough introductions). For example, in the projective plane, every two projective lines intersect, so there is no notion of parallel projective lines. (This is intuitively clear if we consider the spherical model of the projective plane, since projective lines correspond to great circles and every two great circles intersect.) The fundamental transformations of the projective plane, projective transformations, are maps from $\mathbb{RP}^2$ to $\mathbb{RP}^2$ that correspond (in homogeneous coordinates) to invertible linear transformations from $\mathbb{R}^3$ to itself.

Projective transformations fail to preserve many of the quantities that one might expect to be preserved. Collinearity of projective points and coincidence of projective lines are preserved by projective transformations, but distances between projective points, angles between projective lines, ratios of lengths between projective points on a projective line, and even "betweenness" are *not* preserved! However, there *is* one important quantity that is preserved by projective transformations, the projective version of the cross ratio, which will be defined shortly (Brannan, et al. [**1**] and Eves [**5**] give detailed proofs that the cross ratio is preserved). The cross ratio is such an important invariant in projective geometry that Felix Klein mentions "the widespread tendency in projective geometry to resolve all magnitudes which exhibit invariant character back to cross ratios" [**12**], although he thought this was an unfortunate tendency.

For the remainder of the paper, let A, B, C, and D be four collinear projective points in $\mathbb{RP}^2$, and let homogeneous coordinates be chosen so that $A = [\mathbf{a}]$, $B = [\mathbf{b}]$, $C = [\mathbf{c}]$, and $D = [\mathbf{d}]$, where $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ are vectors in $\mathbb{R}^3$. Since A, B, C, and D are collinear as projective points in $\mathbb{RP}^2$, **a**, **b**, **c**, and **d** must lie in the same plane in $\mathbb{R}^3$. Thus, **c** and **d** may be represented as linear combinations of **a** and **b**. In other words, there exist scalars α, β, γ, and δ so that $\mathbf{c} = \alpha\mathbf{a} + \beta\mathbf{b}$ and $\mathbf{d} = \gamma\mathbf{a} + \delta\mathbf{b}$. Using this notation [1], the cross ratio $(ABCD)$ of collinear projective points with given homogeneous coordinates

$$(ABCD) = \frac{\beta}{\alpha}\Big/\frac{\delta}{\gamma} = \frac{\beta\gamma}{\alpha\delta}. \tag{1}$$

Note that this definition collapses to the Euclidean definition if the vectors **a**, **b**, **c**, and **d** are themselves collinear in $\mathbb{R}^3$. To see this, observe that if **a**, **b**, **c**, and **d** are collinear, then $\mathbf{c} = (1-t)\mathbf{a} + t\mathbf{b}$ and $\mathbf{d} = (1-s)\mathbf{a} + s\mathbf{b}$ for some scalars s and t, so that $\alpha = 1-t$, $\beta = t$, $\gamma = 1-s$, and $\delta = s$. If $||\mathbf{v}||$ is the usual Euclidean norm of **v** and $\langle\mathbf{uv}\rangle$ gives the sign (± 1) assigned to the oriented line segment **uv** according to the chosen orientation of the line, then the ratio of signed distances is

$$\begin{aligned}\frac{|\mathbf{ac}|}{|\mathbf{cb}|}\Big/\frac{|\mathbf{ad}|}{|\mathbf{db}|} &= \frac{\langle\mathbf{ac}\rangle||\mathbf{c}-\mathbf{a}||}{\langle\mathbf{cb}\rangle||\mathbf{b}-\mathbf{c}||}\Big/\frac{\langle\mathbf{ad}\rangle||\mathbf{d}-\mathbf{a}||}{\langle\mathbf{db}\rangle||\mathbf{b}-\mathbf{d}||}\\ &= \frac{t|||\mathbf{b}-\mathbf{a}||}{(1-t)|||\mathbf{b}-\mathbf{a}||}\Big/\frac{s|||\mathbf{b}-\mathbf{a}||}{(1-s)|||\mathbf{b}-\mathbf{a}||}\\ &= \frac{\beta}{\alpha}\Big/\frac{\delta}{\gamma}.\end{aligned}$$

Using cross products to calculate cross ratios Recall that the cross ratio of four collinear points in ordinary Euclidean space can be calculated using a ratio of ratios of signed distances:

$$(ABCD) = \frac{|AC|}{|CB|}\Big/\frac{|AD|}{|DB|} = \frac{|AC|}{|CB|}\cdot\frac{|DB|}{|AD|}.$$

Naively, one could attempt to calculate a cross ratio of projective points, using cross products, by mimicking the Euclidean formulation and attempting to calculate

$$\frac{\mathbf{a}\times\mathbf{c}}{\mathbf{c}\times\mathbf{b}}\Big/\frac{\mathbf{a}\times\mathbf{d}}{\mathbf{d}\times\mathbf{b}}.$$

Of course, vectors cannot be divided, but since **a**, **b**, **c**, and **d** are coplanar (and rooted at the origin), the cross products of any pair of these vectors must lie on the same line. Thus, the four vectors are scalar multiples of each other.

THEOREM 1. *Let A, B, C, and D be four collinear points in* $\mathbb{RP}^2$, *and let homogeneous coordinates be chosen so that* $A = [\mathbf{a}]$, $B = [\mathbf{b}]$, $C = [\mathbf{c}]$, *and* $D = [\mathbf{d}]$, *where* **a**, **b**, **c**, *and* **d** *are vectors in* $\mathbb{R}^3$. *Furthermore, let j and k be chosen so that* $(\mathbf{a}\times\mathbf{c}) = j(\mathbf{c}\times\mathbf{b})$ *and* $(\mathbf{a}\times\mathbf{d}) = k(\mathbf{d}\times\mathbf{b})$. *Then the ratio* j/k *is the cross ratio* $(ABCD)$.

Proof. If $\mathbf{z} = \mathbf{a}\times\mathbf{b}$ and $\mathbf{c} = \alpha\mathbf{a} + \beta\mathbf{b}$, then

$$\mathbf{a}\times\mathbf{c} = \mathbf{a}\times(\alpha\mathbf{a}+\beta\mathbf{b}) = \alpha(\mathbf{a}\times\mathbf{a}) + \beta(\mathbf{a}\times\mathbf{b}) = \beta\mathbf{z}$$

and

$$\mathbf{c}\times\mathbf{b} = (\alpha\mathbf{a}+\beta\mathbf{b})\times\mathbf{b} = \alpha(\mathbf{a}\times\mathbf{b}) + \beta(\mathbf{b}\times\mathbf{b}) = \alpha\mathbf{z},$$

so that

$$(\mathbf{a}\times\mathbf{c}) = \beta\mathbf{z} = \frac{\beta}{\alpha}(\alpha\mathbf{z}) = \frac{\beta}{\alpha}(\mathbf{c}\times\mathbf{b}),$$

and therefore $j = \beta/\alpha$. Similarly, $k = \delta/\gamma$, so

$$j/k = \frac{\beta}{\alpha}\bigg/\frac{\delta}{\gamma} = (ABCD). \tag{2}$$

A related statement with a similar proof that reflects the Euclidean formulation more clearly is the following:

THEOREM 2. *Let A, B, C, and D be four collinear points in $\mathbb{RP}^2$, and let homogeneous coordinates be chosen so that $A = [\mathbf{a}]$, $B = [\mathbf{b}]$, $C = [\mathbf{c}]$, and $D = [\mathbf{d}]$, where $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ are vectors in $\mathbb{R}^3$. Then*

$$(ABCD) = \frac{(\mathbf{a}\times\mathbf{c})\cdot(\mathbf{d}\times\mathbf{b})}{(\mathbf{c}\times\mathbf{b})\cdot(\mathbf{a}\times\mathbf{d})}.$$

Proof. As above, let $\mathbf{c} = \alpha\mathbf{a} + \beta\mathbf{b}$ and $\mathbf{d} = \gamma\mathbf{a} + \delta\mathbf{b}$. Consider the ratio

$$\frac{(\mathbf{a}\times\mathbf{c})\cdot(\mathbf{d}\times\mathbf{b})}{(\mathbf{c}\times\mathbf{b})\cdot(\mathbf{a}\times\mathbf{d})} = \frac{(\mathbf{a}\times[\alpha\mathbf{a}+\beta\mathbf{b}])\cdot([\gamma\mathbf{a}+\delta\mathbf{b}]\times\mathbf{b})}{([\alpha\mathbf{a}+\beta\mathbf{b}]\times\mathbf{b})\cdot(\mathbf{a}\times[\gamma\mathbf{a}+\delta\mathbf{b}])}.$$

Evaluating cross products, we obtain

$$\frac{\beta(\mathbf{a}\times\mathbf{b})\cdot\gamma(\mathbf{a}\times\mathbf{b})}{\alpha(\mathbf{a}\times\mathbf{b})\cdot\delta(\mathbf{a}\times\mathbf{b})} = \frac{\beta\gamma||\mathbf{a}\times\mathbf{b}||^2}{\alpha\delta||\mathbf{a}\times\mathbf{b}||^2} = \frac{\beta\gamma}{\alpha\delta} = (ABCD).$$

Cross ratio of a pencil of four distinct lines If we have four coplanar lines in $\mathbb{R}^3$ that intersect in a single point, we can assign to them a cross ratio that has many properties in common with the cross ratio of collinear points. The term *pencil of lines* refers to a set of lines that meet at a common point, so we call this the *cross ratio of a pencil of four distinct lines* (for a more detailed discussion see Eves [**5**] or Seidenberg [**13**]). Given four coplanar lines l, m, n, and p that intersect at a point $\mathcal{O}$ and (Euclidean) points $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ on lines l, m, n, and p, respectively, the cross ratio assigned to the set of four lines through the point $\mathcal{O}$ is denoted $\mathcal{O}(\mathbf{abcd})$ and defined as

$$\mathcal{O}(\mathbf{abcd}) = \frac{\sin(\mathbf{a}\mathcal{O}\mathbf{c})}{\sin(\mathbf{c}\mathcal{O}\mathbf{b})}\bigg/\frac{\sin(\mathbf{a}\mathcal{O}\mathbf{d})}{\sin(\mathbf{d}\mathcal{O}\mathbf{b})}.$$

It is a well-known result that the cross ratio of a pencil of four distinct lines is equal to the cross ratio of the four collinear points formed by the intersection of the four lines with any transversal [**6, 13**]. However, one can prove directly that the cross ratio of the pencil of four distinct lines is also a ratio of cross products, as shown in Theorem 3 below. The method of proof presented here, except for the explicit use of the notation of cross products in the definitions of j and k, is similar to that used in more classical proofs about the role of the transversal. A key geometrical insight is that cross products may be used to calculate areas of triangles, and that when a pencil of four distinct lines is cut by a transversal, a collection of triangles is formed that all have the same height and so have areas related to the lengths of their bases.

THEOREM 3. *Suppose four coplanar lines l, m, n, and p intersect at the origin $\mathcal{O} = (0, 0, 0)$ in $\mathbb{R}^3$, and suppose that vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ lie on lines l, m, n, and p, respectively. If j and k are chosen so that $(\mathbf{a}\times\mathbf{c}) = j(\mathbf{c}\times\mathbf{b})$ and $(\mathbf{a}\times\mathbf{d}) = k(\mathbf{d}\times\mathbf{b})$, then the ratio j/k is equal to $\mathcal{O}(\mathbf{abcd})$.*

Proof. The signed length of $\mathbf{a} \times \mathbf{c}$ is $||\mathbf{a}||\,||\mathbf{c}|| \sin(\mathbf{a}\mathcal{O}\mathbf{c})$ and the signed length of $\mathbf{c} \times \mathbf{b}$ is $||\mathbf{c}||\,||\mathbf{b}|| \sin(\mathbf{c}\mathcal{O}\mathbf{b})$. Note that here we are using the interpretation of cross product in which the resulting vector has a length equal to the area of the parallelogram constructed from the two vectors and whose direction is determined by the order of the operation. If j and k are defined so that $(\mathbf{a} \times \mathbf{c}) = j(\mathbf{c} \times \mathbf{b})$ and $(\mathbf{a} \times \mathbf{d}) = k(\mathbf{d} \times \mathbf{b})$, then

$$j = \frac{||\mathbf{a}||\,||\mathbf{c}|| \sin(\mathbf{a}\mathcal{O}\mathbf{c})}{||\mathbf{c}||\,||\mathbf{b}|| \sin(\mathbf{c}\mathcal{O}\mathbf{b})} = \frac{||\mathbf{a}|| \sin(\mathbf{a}\mathcal{O}\mathbf{c})}{||\mathbf{b}|| \sin(\mathbf{c}\mathcal{O}\mathbf{b})}. \tag{3}$$

Similarly, $k = \dfrac{||\mathbf{a}|| \sin(\mathbf{a}\mathcal{O}\mathbf{d})}{||\mathbf{b}|| \sin(\mathbf{d}\mathcal{O}\mathbf{b})}$, so

$$j/k = \frac{\sin(\mathbf{a}\mathcal{O}\mathbf{c})}{\sin(\mathbf{c}\mathcal{O}\mathbf{b})} \Bigg/ \frac{\sin(\mathbf{a}\mathcal{O}\mathbf{d})}{\sin(\mathbf{d}\mathcal{O}\mathbf{b})} = \mathcal{O}(\mathbf{abcd}). \tag{4}$$

Historical background We started with a student, desperately and creatively attempting to solve a problem on a final exam. We end with the delightful discovery that the cross ratio of projective geometry is related to the cross product of Euclidean geometry. The authors are particularly enamored of the third theorem, because the proof provides a geometric indication for why the "ratio" of cross products has anything to do with the cross ratio, namely that the cross ratio is a ratio of the sines of angles formed by the lines (or vectors) in question, while the cross product is a compact tool for isolating the sines of the angles between vectors or lines. We wondered whether there was a more than coincidental historical reason that these two definitions share such similar names, but sadly, this does not appear to be the case.

The best source we have been able to find about the origins of the term "cross ratio" is a brief mention by Eves [**5**, p. 87]:

> Essentially the notation (AB, CD)

Here the notation (AB, CD) is our $(ABCD)$. Unfortunately, the only publication we could find of Clifford's in 1878 containing the term *cross ratio* was his *Elements of Dynamic* [**3**], where he uses it without comment, as though referring to a well-known usage. Other papers from 1878 by Clifford appear in his *Mathematical Papers* [**4**], but we could not find an instance of his use of the term in those works.

The origin of the term cross product, on the other hand, seems to be related to the introduction by J. W. Gibbs of the notation $\mathbf{a} \times \mathbf{b}$ to indicate the vector product of vectors $\mathbf{a}$ and $\mathbf{b}$. The notation was introduced in an unpublished pamphlet on vector analysis [**7**, p. 20] during the period 1881–84 (see also Heaviside's *Electromagnetic Theory* [**10, 11**]), and in an 1891 paper he said that the notation he used was called a "cross" [**7**, p. 159]. The notation was popularized by Gibbs' student E. B. Wilson [**15**] in a book he wrote based on Gibbs' pamphlet and lectures; Wilson's text also carried the instruction that $\mathbf{a} \times \mathbf{b}$ should be read aloud as "a cross b."

Although our formulation of cross ratios in terms of cross products does not seem to appear explicitly anywhere in the literature, there are several references in the literature that seem to relate cross ratios to cross products or cross product-like objects.

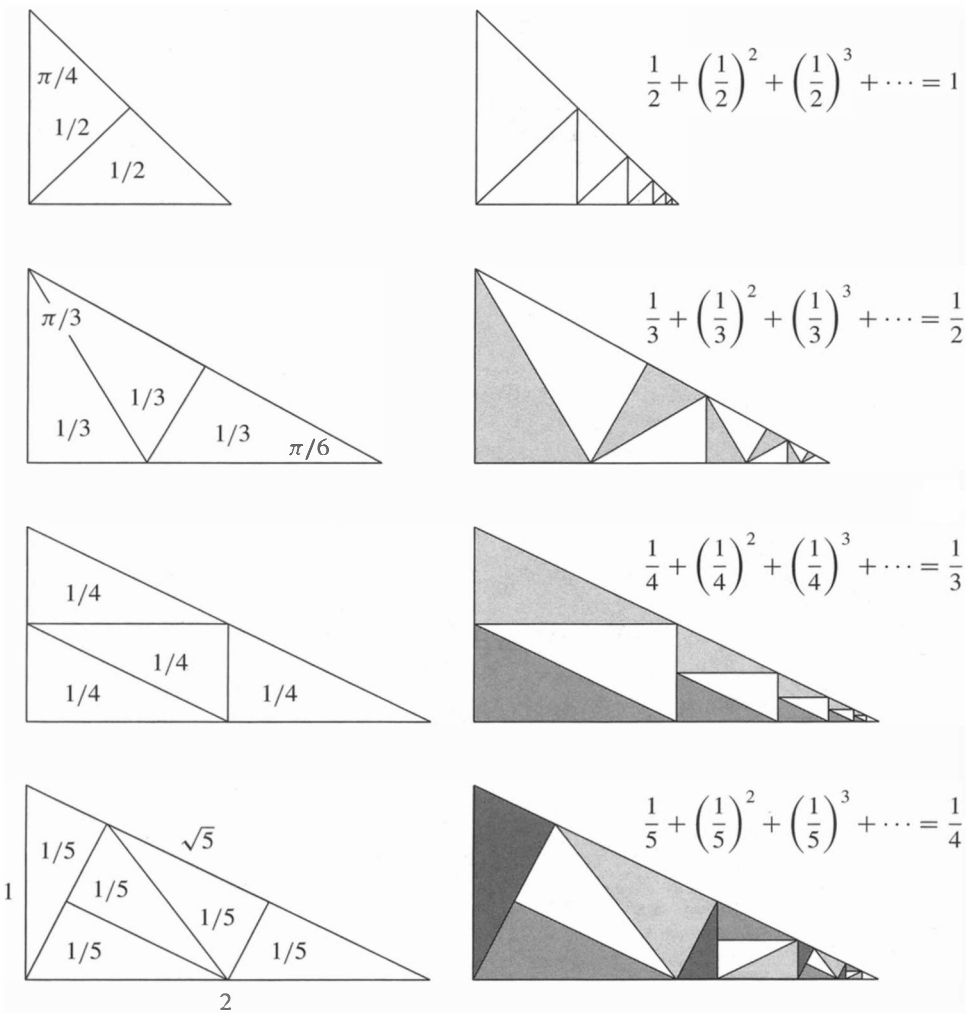
We are grateful to the anonymous referees for suggesting many of these references. For example, (3) appears (with a change of notation) in A. Seidenberg's *Lectures in Projective Geometry* [**13**] in his proof relating the cross ratio of a pencil of four distinct lines to the cross ratio of four collinear points. In addition, there are hints of results similar to Theorem 1 in several references. Cross products have some interesting properties, including being anticommutative, nilpotent, and distributing over addition (as can be seen in any standard calculus text such as Stewart [**14**]). Klein [**12**] defines the cross ratio as a ratio of products of 2×2 determinants, which share many of the properties of cross products (they "act like" cross products). Hermann Ernst Grassmann (sometimes called Grassmann the Younger, to differentiate him from his more famous father, Hermann Günther Grassmann), defined cross ratio, which he called double ratio, using a ratio of "outer products" (*äußere Multiplikation*) [**8**]. These outer products seem to be closely related to the "combinatorial product" developed in his father's book on Extension Theory [**9**] and were defined to have to have precisely the same properties as those listed above for cross products, so his definition of cross ratio is very similar to our Theorem 1. Finally, Busemann and Kelly [**2**] use cross products explicitly when discussing the fact that the cross ratio of a pencil of four lines is equal to the cross ratio of four points formed by the intersection of the pencil of four distinct lines with a transversal.

Acknowledgment. The authors would like to thank Joan Rhodes and Judith Fryer of Myrin Library, Ursinus College, and Debbi Gaspar and Nancy Strobel of Reeves Library, Moravian College, for help with historical research, and the anonymous referees, for suggesting some very interesting historical references.

REFERENCES

1. David Brannan, Matthew Esplin, and Jeremy Gray, *Geometry*, Cambridge University Press, 1999, pp. 136–141.
2. Herbert Busemann and Paul J. Kelly, *Projective Geometry and Projective Metrics*, Academic Press Inc., 1953, p. 28–31.
3. William Kingdon Clifford, *Elements of Dynamic; An Introduction to the Study of Motion and Rest in Solid and Fluid Bodies (Parts 1–3)*, McMillan and Co., 1878, p. 42. http://historical.library.cornell.edu:80/Dienst/UI/1.0/Display/cul.math/04370002
4. William Kingdon Clifford, *Mathematical Papers*, Chelsea Publishing Co., 1968.
5. Howard Eves, *A Survey of Geometry, Volume 1*, Allyn and Bacon, Inc., 1965, pp. 86–89.
6. Howard Eves, *College Geometry*, Jones and Bartlett Publishers International, 1995, p. 73.
7. J. Willard Gibbs, *The Scientific Papers of J. Willard Gibbs, Vol. 2*, Dover Publications Inc, 1961.
8. Hermann Ernst Grassmann, *Projektive Geometrie der Ebene, unter Benutzung der Punktrechnung (Volume 1)*, Druck und Verlag von B.G. Teubner, 1909, pp. 1–49. http://historical.library.cornell.edu:80/cgi-bin/cul.math/docviewer?did=00900002&seq=5
9. Hermann Günther Grassmann, *Extension Theory*, (Die Ausdehnungslehre von 1862) Lloyd C. Kannenberg, trans. American Mathematical Society, 2000, pp. 19–49.
10. Oliver Heaviside, *Electromagnetic Theory, Vol. I*, "The Electrician" Printing and Publishing Company, Ltd., 1893, pp. 138, 304.
11. Oliver Heaviside, *Electromagnetic Theory, Vol. III*, "The Electrician" Printing and Publishing Company, Ltd., 1912, pp. 135–143.
12. Felix Klein, *Geometry*, E.R. Hedrick and Charles Noble, trans. Dover Publications, Inc., 1939, pp. 144–147.
13. A. Seidenberg, *Lectures in Projective Geometry, Vol. 1*, D. Van Nostrand Company, Inc. 1962, pp. 22–23.
14. James Stewart, *Calculus, Early Transcendentals*, 5th edition, Brooks/Cole, 2003, pp. 814–820.
15. E. B. Wilson, *Vector Analysis; A Text-book for the Use of Students of Mathematics and Physics, Founded upon the Lectures of J. Willard Gibbs*, C. Scribner's Sons, 1901, p. 61.

Proof Without Words: Right Triangles and Geometric Series



CHALLENGE: Can you create the next two rows?

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

The Euler-Maclaurin Formula and Sums of Powers

MICHAEL Z. SPIVEY
University of Puget Sound
Tacoma, WA 98416
mspivey@ups.edu

Mathematicians have long been intrigued by the sum $1^m + 2^m + \cdots + n^m$ of the first n integers, where m is a nonnegative integer. The study of this sum of powers led Jakob Bernoulli to the discovery of Bernoulli numbers and Bernoulli polynomials. There are expressions for sums of powers in terms of Eulerian numbers and Stirling numbers [**5**, p. 199]. In addition, past articles in this MAGAZINE contain algorithms for producing a formula for the sum involving powers of $m+1$ from that involving powers of m [**1, 4**]. (The algorithm in Bloom [**1**] is actually Bernoulli's method.)

This note involves a curious property concerning sums of integer powers, namely,

$$1^m + 2^m + \cdots + (m-1)^m < m^m, \quad \text{for } m \geq 1. \tag{1}$$

In other words, the sum of the $m-1$ terms from 1^m to $(m-1)^m$ is always less than the single term m^m, regardless of how large m is. This inequality is not true for an arbitrary number of terms; $1^m + 2^m + \cdots + (n-1)^m$ is not necessarily less than n^m for all n, but the inequality is true when $n = m$.

Proving (1) is not too difficult. In fact, one proof is a nice first-semester calculus problem using left-hand Riemann sums to underestimate the integral $\int_0^m x^m\,dx$. Another establishes $(x+1)^m - x^m > x^m$ for $x < m$ via the binomial theorem; replacing x successively with $0, 1, 2, \ldots, m-1$ and summing yields (1).

There is a deeper question here, though. Dividing (1) by m^m produces the inequality

$$\left(\frac{1}{m}\right)^m + \left(\frac{2}{m}\right)^m + \cdots + \left(\frac{m-1}{m}\right)^m < 1. \tag{2}$$

Since this relation holds regardless of the value of m, a natural question to ask is this: What is the limiting value of the expression on the left of (2) as m approaches infinity? Our investigation of this value involves a useful tool in any mathematician's bag of tricks—one that is, unfortunately, not often taught in undergraduate courses—the Euler-Maclaurin formula for approximating a finite sum by an integral. Along the way we also prove (1) using Euler-Maclaurin, thus illustrating the use of the Euler-Maclaurin formula with remainder.

Rota calls Euler-Maclaurin "one of the most remarkable formulas of mathematics" [**6**, p. 11]. After all, it shows us how to trade a finite sum for an integral. It works much like Taylor's formula: The equation involves an infinite series that may be truncated at any point, leaving an error term that can be bounded.

The formula uses the very numbers discovered by Bernoulli during his investigations into the power sum, and the error term uses Bernoulli's polynomials. For example, the second-order formula with error term is given in *Concrete Mathematics* [**2**, p. 469]:

$$\sum_{j=0}^{n-1} f(j) = \int_0^n f(x)dx + \frac{B_1}{1!}\Big(f(n) - f(0)\Big) + \frac{B_2}{2!}\Big(f'(n) - f'(0)\Big)$$
$$+ (-1)^3 \frac{1}{2!} \int_0^n B_2(\{x\}) f''(x)dx, \tag{3}$$

where

- B_i is the ith Bernoulli number ($B_1 = -1/2$, $B_2 = 1/6$),
- $B_2(x)$ is the second Bernoulli polynomial: $x^2 - x + 1/6$,
- $\{x\} = x - \lfloor x \rfloor$, and
- f is twice-differentiable.

Since $\{x\}$ is the fractional part of x, the function $B_2(\{x\})$ in (3) is just the periodic extension of the parabola $B_2(x) = x^2 - x + 1/6$ from [0, 1] to the entire real number line. In other words, $B_2(\{x\})$ agrees with $B_2(x)$ on [0, 1] and is periodic with period 1.

Proving (3) involves nothing more complicated than integration by parts. A brief outline is as follows: Start with $(1/2)\int_0^1 (y^2 - y + 1/6)\, g''(y)\, dy$. Use integration by parts twice and solve for $g(0)$. Let $g(y) = f(y + j)$, and then substitute x for $y + j$ to find an expression for $f(j)$. Sum this expression as j varies from 0 to $n - 1$, noting that the terms involving $f'(j)$ and $f'(j + 1)$ telescope, while those involving $f(j + 1)$ are absorbed into the sum. This yields (3), since $B_2(\{y\}) = B_2(\{x\})$. The interested reader is invited to fill in the details.

The full Euler-Maclaurin formula with no remainder term (for infinitely differentiable f) is given in *Concrete Mathematics* [**2**, p. 471]:

$$\sum_{j=0}^{m-1} f(j) = \int_0^m f(x)dx + \sum_{k=1}^{\infty} \frac{B_k}{k!}\Big(f^{(k-1)}(m) - f^{(k-1)}(0)\Big). \tag{4}$$

Unfortunately, the infinite sum on the right-hand side often diverges. This formula can also be proved using integration by parts; Lampret, in fact, shows how to use parts to prove Euler-Maclaurin for arbitrary orders [**3**].

On to the proof of (1): We can easily verify the inequality for small values of m. In particular, for $m = 1$, we have $0 < 1 = 1^1$, and for $m = 2$, we have $1^2 = 1 < 4 = 2^2$. For $m \geq 3$, we turn to Euler-Maclaurin. Plugging $f(x) = x^m$ and $n = m$ into (3) yields

$$\sum_{j=1}^{m-1} j^m = \int_0^m x^m dx - \frac{1}{2}m^m + \frac{1}{12} m\, m^{m-1} - \frac{1}{2!}\int_0^m B_2(\{x\}) m(m-1)x^{m-2}dx$$
$$= \frac{m^{m+1}}{m+1} - \frac{5}{12}m^m - \frac{1}{2}\int_0^m B_2(\{x\}) m(m-1)x^{m-2}dx. \tag{5}$$

Now, let's deal with the error term. Completing the square on the parabola $B_2(x)$ gives us $B_2(x) = (x - 1/2)^2 - 1/12$. This tells us that the minimum value of $B_2(x)$ on [0, 1] is $-1/12$, occurring at $x = 1/2$, and the maximum value on [0, 1] is $1/6$, occurring at the two endpoints $x = 0$ and $x = 1$. Since $B_2(\{x\})$ is the periodic extension of $B_2(x)$ from [0, 1] to the real number line, the minimum and maximum values of $B_2(\{x\})$ over the real numbers are $-1/12$ and $1/6$, respectively (which, incidentally, occur infinitely often). This tells us that $-1/2 B_2(\{x\}) \leq (-1/2)(-1/12) = 1/24$.

Therefore,

$$\frac{-1}{2}\int_0^m B_2(\{x\})m(m-1)x^{m-2}dx \leq \frac{1}{24}\int_0^m m(m-1)x^{m-2}dx$$
$$= \frac{m}{24}m^{m-1} = \frac{1}{24}m^m.$$

Plugging back into (5) produces

$$\sum_{j=1}^{m-1} j^m \leq \frac{m^{m+1}}{m+1} - \frac{5}{12}m^m + \frac{1}{24}m^m$$
$$< m^m - \frac{3}{8}m^m = \frac{5}{8}m^m.$$

This establishes the inequality (1), namely $1^m + 2^m + \cdots + (m-1)^m < m^m$, for all positive integers m, via the second-order Euler-Maclaurin formula with remainder.

We now move on to our main question—determining the limiting expression for

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

From our proof of (1), we know that the limit must be less than $5/8$. To find the exact value we use the full Euler-Maclaurin formula (4). For fixed m and $f(x) = x^m$, we have

$$\sum_{j=1}^{m-1}\left(\frac{j}{m}\right)^m = \frac{1}{m^m}\sum_{j=0}^{m-1} j^m$$
$$= \frac{1}{m^m}\int_0^m x^m dx + \frac{1}{m^m}\sum_{k=1}^{\infty}\frac{B_k}{k!}\left(f^{(k-1)}(m) - f^{(k-1)}(0)\right)$$
$$= \frac{m}{m+1} + \frac{1}{m^m}\sum_{k=1}^{\infty}\frac{B_k}{k!}(f^{(k-1)}(m) - f^{(k-1)}(0)).$$

Since $f^{(k-1)}(m) - f^{(k-1)}(0)$ is nonzero only for $k \leq m$, this yields

$$\sum_{j=1}^{m-1}\left(\frac{j}{m}\right)^m = \frac{m}{m+1} + \frac{1}{m^m}\sum_{k=1}^{m}\frac{B_k}{k!}f^{(k-1)}(m)$$
$$= \frac{m}{m+1} + \frac{1}{m^m}\sum_{k=1}^{m}\left[\frac{B_k}{k!}m^{m-k+1}\left(m(m-1)\cdots(m-k+2)\right)\right]$$
$$= \frac{m}{m+1} + \sum_{k=1}^{m}\left[\frac{B_k}{k!}m^{1-k}(m(m-1)\cdots(m-k+2))\right].$$

There are exactly $k-1$ factors in the expression $m(m-1)\cdots(m-k+2)$. Thus the resulting polynomial is m^{k-1} plus a polynomial of degree $k-2$. For our purposes, all that matters of the latter polynomial is its degree. We can therefore use "big-O" notation to express $m(m-1)\cdots(m-k+2)$ as $m^{k-1} + O(m^{k-2})$. Here, $O(m^{k-2})$ effectively means that the expression added to m^{k-1} is of order no larger than that of m^{k-2}. (For a more precise definition and a discussion of big-O notation, see *Concrete*

Mathematics [**2**, p. 471].) Multiplying through by m^{1-k} then yields the expression $1 + O(1/m)$. Substituting back in, we have

$$\sum_{j=1}^{m-1}\left(\frac{j}{m}\right)^m = \frac{m}{m+1} + \sum_{k=1}^{m}\frac{B_k}{k!}\left[1 + O\left(\frac{1}{m}\right)\right].$$

Now we take the limit to get

$$\begin{aligned}\lim_{m\to\infty}\sum_{j=1}^{m-1}\left(\frac{j}{m}\right)^m &= \lim_{m\to\infty}\left\{\frac{m}{m+1} + \sum_{k=1}^{m}\frac{B_k}{k!}\left[1 + O\left(\frac{1}{m}\right)\right]\right\}\\ &= 1 + \sum_{k=1}^{\infty}\frac{B_k}{k!} + \lim_{m\to\infty}\left\{O\left(\frac{1}{m}\right)\sum_{k=1}^{m}\frac{B_k}{k!}\right\}.\end{aligned}$$

The crucial question for both the second and third terms is the convergence of $\sum_{k=0}^{\infty} B_k/k!$. Fortunately, the infinite sum is a special case of the exponential generating function for the Bernoulli numbers,

$$\sum_{k=0}^{\infty} B_k\frac{x^k}{k!} = \frac{x}{e^x - 1},$$

valid for $|x| < 2\pi$ [**5**, p.147]. Therefore, $\sum_{k=1}^{m} B_k/k!$ is bounded by a constant, yielding

$$\lim_{m\to\infty}\left\{O\left(\frac{1}{m}\right)\sum_{k=1}^{m}\frac{B_k}{k!}\right\} = 0.$$

Since $B_0 = 1$, we have

$$\lim_{m\to\infty}\sum_{j=1}^{m-1}\left(\frac{j}{m}\right)^m = \sum_{k=0}^{\infty}\frac{B_k}{k!},$$

which gives us the simple limiting expression

$$\lim_{m\to\infty}\left[\left(\frac{1}{m}\right)^m + \left(\frac{2}{m}\right)^m + \cdots + \left(\frac{m-1}{m}\right)^m\right] = \frac{1}{e-1}.$$

Thus, in the limit, the sum $1^m + 2^m + \cdots + (m-1)^m$ will represent $(e-1)^{-1}$ (approximately 0.582) of m^m.

The interested reader may enjoy showing that the left-hand side of (2) actually increases to $1/(e-1)$. In addition, the excellent text *Concrete Mathematics* contains numerous further examples of the use of the Euler-Maclaurin summation formula [**2**, pp. 469–489].

Acknowledgment. The author would like to thank one of the referees for several helpful suggestions.

REFERENCES

1. D. M. Bloom, An old algorithm for the sums of integer powers. this MAGAZINE, **66** (1993), 304–305.
2. R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, Reading, MA, second edition, 1994.

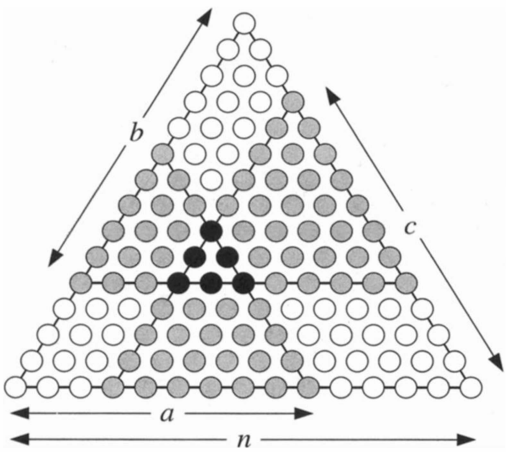
3. Vito Lampret, The Euler-Maclaurin and Taylor formulas: Twin, elementary derivations, this MAGAZINE, **74** (2001) 109–122.
4. Robert W. Owens, Sums of powers of integers, this MAGAZINE, **65** (1992), 38–40.
5. Kenneth H. Rosen, ed., *Handbook of Discrete and Combinatorial Mathematics*, CRC Press, Boca Raton, FL, 2000.
6. Gian-Carlo Rota, Combinatorial snapshots, *Math. Intelligencer*, **21** (1999), 8–14.

Proof Without Words: Inclusion-Exclusion for Triangular Numbers

THEOREM. Let $t_k = 1 + 2 + \cdots + k$, $t_0 = 0$. If $0 \le a, b, c \le n$ and $2n \le a + b + c$, then

$$t_a + t_b + t_c - t_{a+b-n} - t_{b+c-n} - t_{c+a-n} + t_{a+b+c-2n} = t_n$$

Proof.



NOTES:

1. If $0 \le a, b, c \le n$, $2n > a + b + c$, but $n \le \min(a + b, b + c, c + a)$, then the identity is $t_a + t_b + t_c - t_{a+b-n} - t_{b+c-n} - t_{c+a-n} + t_{2n-a-b-c-1} = t_n$, with a similar proof.
2. The following special cases are of interest:
   (a) If $(n; a, b, c) = (2k - j; k, k, k)$, then $3(t_k - t_j) = t_{2k-j} - t_{2j-k}$;
   (b) If $(n; a, b, c) = (a + b + c; 2a, 2b, 2c)$, then $t_{2a} + t_{2b} + t_{2c} = t_{a+b+c} + t_{a+b-c} + t_{a-b+c} + t_{-a+b+c}$;
   (c) If $(n; a, b, c) = (3k; 2k, 2k, 2k)$, then $3(t_{2k} - t_k) = t_{3k}$.

Illustrations for all of these cases can be found at the MAGAZINE website.

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

Disjoint Pairs with Distinct Sums

GERHARD J. WOEGINGER
Department of Mathematics and Computer Science
TU Eindhoven
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
gwoegi@win.tue.nl

Klamkin and Newman [**1**] investigated the maximum number $N(k)$ of disjoint pairs of positive integers with distinct sums at most k. For instance for $k = 7$ the two disjoint pairs $(1, 3)$ and $(2, 4)$ have the distinct sums 4 and 6, whereas there do not exist three such pairs with distinct sums. Whence $N(7) = 2$. Klamkin and Newman proved that $(2k - 13)/5 \leq N(k) \leq (2k - 1)/5$, thus leaving a small gap between the upper and the lower bound. We close this gap by showing that $N(k) = \lfloor (2k - 1)/5 \rfloor$ for all $k \geq 1$.

For the sake of completeness, let us briefly recapitulate Klamkin and Newman's upper bound argument [**1**]: Consider $N = N(k)$ disjoint pairs with distinct sums at most k, and let S be the sum of all $2N$ elements in these pairs. Since the pairs are disjoint, S is at least $1 + 2 + \cdots + 2N = (2N + 1)N$. Since the N pair-sums are distinct and at most k, S is at most $k + (k - 1) + (k - 2) + \cdots + (k - N + 1) = (2k - N + 1)N/2$. Combining these two inequalities yields $N(k) \leq (2k - 1)/5$.

For the lower bound, we first consider the three cases where $k = 5m + 3$, $k = 5m + 4$, or $k = 5m + 5$ for some integer m. Note that this implies $\lfloor (2k - 1)/5 \rfloor = 2m + 1$. We exhibit $2m + 1$ disjoint pairs where the first coordinates cover the range $1, 2, \ldots, 2m + 1$ and the second coordinates cover the range $2m + 2, \ldots, 4m + 2$:

- The first $m + 1$ pairs are $(j, 3m + j + 1)$,
  with sums $3m + 2j + 1$, for $1 \leq j \leq m + 1$;
- the last m pairs are $(m + j + 1, 2m + j + 1)$,
  with sums $3m + 2j + 2$, for $1 \leq j \leq m$.

Since the $2m + 1$ pair-sums are distinct and bounded by $5m + 3$, this demonstrates $N(k) \geq (2k - 1)/5$ for these three cases. In the remaining two cases, we have either $k = 5m + 1$ or $k = 5m + 2$ for some integer m. We exhibit $\lfloor (2k - 1)/5 \rfloor = 2m$ disjoint pairs where the first coordinates cover the range $1, 2, \ldots, 2m$ and the second coordinates cover the range $2m + 1, \ldots, 4m$:

- The first pair is $(1, 2m + 1)$ with sum $2m + 2$;
- the next m pairs are $(j + 1, 3m + j)$,
  with sums $3m + 2j + 1$, for $1 \leq j \leq m$;
- the last $m - 1$ pairs are $(m + j + 1, 2m + j + 1)$
  with sums $3m + 2j + 2$, for $1 \leq j \leq m - 1$.

These $2m$ pair-sums are distinct and bounded by $5m + 1$. Hence, also in these two cases we have $N(k) \geq (2k - 1)/5$ and the argument is complete.

REFERENCE

1. M. S. Klamkin and D. J. Newman, *Some combinatorial problems of arithmetic*, this MAGAZINE **42** (1969), 53–56.

PROBLEMS

ELGIN H. JOHNSTON, *Editor*
Iowa State University

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

Proposals

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

1736. *There was no name on this proposal.*

Is there a triangle all of whose vertices are points with rational coordinates on the circle $x^2 + y^2 = 1$ and whose vertex angles are 45°, 60°, and 75°?

1737. *Proposed by Michael Z. Spivey, The University of Puget Sound, Tacoma, WA.*

A Pythagorean triple is an ordered triple (a, b, c) of positive integers satisfying $a^2 + b^2 = c^2$. The number c is called the hypotenuse of the Pythagorean triple.

a) Prove that an even perfect number cannot be the hypotenuse of a Pythagorean triple.
b) Prove that if there is an odd perfect number, then it is the hypotenuse of a Pythagorean triple.

1738. *Proposed by José Luis Díaz-Barrero, Universitat Politécnica de Catalunya, Baracelona, Spain.*

Let n be a positive number and let $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ be complex numbers. Prove that

$$\operatorname{Re}\left(\sum_{k=1}^{n} a_k b_k\right) \leq \frac{1}{2n}\left(\sum_{k=1}^{n} |a_k|^2 + \frac{4n^2-1}{3}\sum_{k=1}^{n} |b_k|^2\right).$$

1739. *Proposed by Emeric Deutsch, Polytechnic University, Brooklyn, NY.*

An object moves in the plane, starting from the origin, and at each step moving one unit up, down, to the right, or to the left. Find the number of such paths that stay in the

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

Solutions should be written in a style appropriate for this MAGAZINE. Each solution should begin on a separate sheet.

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

quadrant $\{(x, y) : x, y \geq 0\}$, and consist of a total of n steps, exactly k of which are vertical (up or down).

1740. *Proposed by Michel Bataille, Rouen, France.*

Prove or disprove: there exists a scalar multiplication $(z, r) \to z * r$ from $\mathbb{C} \times \mathbb{R}$ into $\mathbb{R}$ such that $(\mathbb{R}, +, *)$ is a vector space over $\mathbb{C}$.

Quickies

Answers to the Quickies are on page 73.

Q957. *Proposed by John H. Jaroma, Austin College, Sherman, Texas.*

Let m and n be odd positive integers such that n is not a perfect square and such that $\sqrt{n} < m + 1 < \sqrt{n} + 2$. Prove that for positive integer k, $\lfloor (m + \sqrt{n})^{2k} + 1 \rfloor$ is divisible by 2^{k+1}.

Q958. *There was no name on this Quickie proposal.*

Let a, b, c be the side lengths of a nondegenerate triangle. Prove that

a) $a^4 + b^4 + c^4 < 2(a^2b^2 + b^2c^2 + c^2a^2)$.

b) $a^3 + b^3 + c^3 + 2abc < ab(a + b) + bc(b + c) + ca(c + a)$.

Solutions

Similarity and the Centroid **February 2005**

1711. *Proposed by Mowaffaq Hajja, Yarmouk University, Irbid, Jordan.*

In triangle ABC, let A' be on BC, B' on CA, and A' on BC, and suppose that the cevians AA', BB', and CC' meet at M. Prove that if $\triangle ABC$ is similar to $\triangle A'B'C'$, then M is the centroid of $\triangle ABC$.

Solution by Michel Bataille, Rouen, France.

Identify the points with complex numbers, with the origin O coinciding with M. Because M is interior to $\triangle ABC$, there are positive real numbers α, β, γ with $\alpha A + \beta B + \gamma C = 0$ and

$$\alpha + \beta + \gamma = 1. \tag{1}$$

Because A' is on $\overline{BC}$ and on line $OA(= MA)$, it follows that

$$A' = \frac{-\alpha}{1 - \alpha} A.$$

Similarly,

$$B' = \frac{-\beta}{1 - \beta} B \quad \text{and} \quad C' = \frac{-\gamma}{1 - \gamma} C.$$

Because $\triangle ABC$ is similar to $\triangle A'B'C'$, we have

$$\frac{A' - B'}{A - B} = \frac{B' - C'}{B - C}.$$

Substituting our expressions for A', B', and C' this yields

$$\frac{\gamma BC}{1-\gamma} + \frac{\beta AB}{1-\beta} + \frac{\alpha AC}{1-\alpha} - \frac{\alpha AB}{1-\alpha} - \frac{\beta BC}{1-\beta} - \frac{\gamma AC}{1-\gamma} = 0. \tag{2}$$

Taking (1) into account, equation (2) implies that

$$BC(\gamma^2 - \beta^2) + CA(\alpha^2 - \gamma^2) + AB(\beta^2 - \alpha^2) = 0,$$

and then that

$$(\gamma^2 - \beta^2)B(C - A) + (\alpha^2 - \gamma^2)A(C - B) = 0.$$

If $\gamma^2 - \beta^2 \neq 0$, then the cross-ratio

$$\frac{(C-A)/(C-B)}{(0-A)/(0-B)} = \frac{B(C-A)}{A(C-B)}$$

is a real number. But this implies that M is on the circumcircle of $\triangle ABC$, contradicting the fact that M is in the interior of the triangle. Thus $\gamma^2 = \beta^2$, and $\alpha^2 = \gamma^2$ as well. It follows that $\alpha = \beta = \gamma = \frac{1}{3}$, and hence, that M is the centroid of $\triangle ABC$.

Also solved by Sadi Abu-Saymeh (Jordan), Herb Bailey, Roy Barbara (Lebanon), Alper Cay (Turkey), Daniele Donini (Italy), Peter Gressis, Peter E. Nüesch (Switzerland), Michael Vowe (Switzerland), Paul Weisenhorn (Germany), Tom Zerger, Li Zhou, and the proposer. There were three incorrect submissions.

Sum and Product of Units **February 2005**

1712. *Proposed by William P. Wardlaw, U. S. Naval Academy, Annapolis, MD.*

For each integer $m > 1$, let $\mathbb{Z}_m = \{0, 1, \ldots, m-1\}$ be the ring of integers modulo m, and let Z_m^* be the (multiplicative) group of units in Z_m. Find the sum

$$S(m) = \sum_{u \in \mathbb{Z}_m^*} u$$

and the product

$$P(m) = \prod_{u \in \mathbb{Z}_m^*} u$$

of all elements in Z_m^*.

Solution by Jim Delany, Emeritus, California Polytechnic State University, San Luis Obispo, CA.

Clearly $S(2) = P(2) = 1$, so we assume that $m \geq 3$.

An element $u \in \mathbb{Z}_m$ is a unit if and only if u and m are relatively prime. Thus if $u \in \mathbb{Z}_m^*$, then $m - u \in \mathbb{Z}_m^*$, and $u \neq m - u$ because $2u \equiv 0 \pmod{m}$ is impossible with $m \geq 3$ and $\gcd(u, m) = 1$. Therefore in the sum

$$\sum_{u \in \mathbb{Z}_m^*} u,$$

each unit can be paired with its additive inverse, giving $S(m) = 0$.

Now let $T = \{u \in \mathbb{Z}_m^* | u^{-1} = u\}$. Note that T is a subgroup of $\mathbb{Z}_m^*$ and let $T' = Z_m^* \backslash T$. We have

$$P(m) = \left(\prod_{u \in T} u\right)\left(\prod_{u \in T'} u\right).$$

The second product is 1, because each factor can be paired with its multiplicative inverse. Thus we need only evaluate the first product. Because T is abelian and every nonidentity element of T is of order 2, T is a direct product of cyclic subgroups of order 2. Thus the order of T is a power of 2.

Now $1, -1 \in T$ and $u \in T$ if and only if $-u \in T$. By pairing each element of T with its additive inverse we find

$$P(m) = \prod_{u \in T} u = (-1)^{|T|/2} = \begin{cases} -1 & |T| = 2 \\ 1 & \text{otherwise} \end{cases}$$

If $m = 2^r$, $r \geq 3$, then $2^{r-1} \pm 1 \in T$. Thus $|T| \geq 4$ and $P(m) = 1$. Next consider the case with $m = ab$ with $a, b \geq 3$ and $\gcd(a, b) = 1$. Note that $1 \not\equiv -1 \pmod a$ and $1 \not\equiv -1 \pmod b$. By the Chinese Remainder Theorem, the system of equations

$$x \equiv 1 \pmod a, \qquad x \equiv -1 \pmod b$$

has a unique solution x_1 modulo ab. Similarly, the system

$$x \equiv -1 \pmod a, \qquad x \equiv 1 \pmod b$$

also has a unique solution x_2 modulo ab. The four elements $1, -1, x_1, x_2 \in T$ are all different modulo m, as can be seen by considering them modulo a and then modulo b. Thus if m has such a factorization, then $|T| \geq 4$ and $P(n) = 1$. The only $m \geq 3$ which are not accounted for are $m = 4$, $m = p^r$, and $m = 2p^r$, where p is an odd prime. But those are precisely the m for which $\mathbb{Z}_m^*$ is a cyclic subgroup; in the language of number theory, these are the values of m for which there exist primitive roots. Because a cyclic group has at most one element of order 2, $|T| = 2$ in these cases and $P(m) = -1$. In summary, for $m \geq 3$,

$$P(m) = \begin{cases} -1 & m = 4 \text{ or } m = p^r \text{ or } m = 2p^r,\ p \text{ an odd prime},\ r \geq 1 \\ 1 & \text{otherwise} \end{cases}$$

Note. Some readers calculated the sum and product in Z, obtaining

$$S(m) = \frac{\phi(m)}{2} \quad \text{and} \quad P(m) = m^{\phi(m)} \prod_{d|m} \left(\frac{f}{d!} d^d \right)^{\mu(m/d)}.$$

The result for $P(m)$ can also be found in Theorem 129 in G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 5th Edition, Oxford University Press, Oxford, 1979, and in Oystein Ore, Number Theory and Its History, McGraw-Hill Co., 1948. See also problem E 2080 in The American Mathematical Monthly, Vol. 76, (1969), 417–418.

Also solved by Hamza Ahmad, Tsehaye Andeberhan, Roy Barbara (Lebanon), S. Floyd Barger, Michel Bataille (France), Jean Bogaert (Belgium), Robert Calcaterra, Minh Can, Wenchang Chu and Pierluigi Magli (Italy), Con Amore Problem Group (Denmark), Robert L. Doucette, Gregory P. Dresden, Dmitry Fleischman, Michael Goldenberg and Mark Kaplan, G.R.A.20 Math Problems Group (Italy), Ralph P. Grimaldi and Kenneth W. McMurdy, Natalio H. Guersenzvaig (Argentina), Lee O. Hagglund, Hannah Harding, Mitch Harris, Russell Jay Hendel, Colonel Johnson Jr. and students V. Gant and Q. Jacobs and A. Paul and L. Raphael and L. Hydell, Steven Klee, Harris Kwong, Elias Lampakis (Greece), Daniel R. Patten, Manuel Reyes, Henry Ricardo, Rolf Richberg (Germany), Daniel Shapiro, Nicholas C. Singer, Doug Wilcock, Michael Woltermann, Bill Yankosky and the NCWC Mathematics Senior Seminar Class, Li Zhou, and the proposer.

Inverse Binomial Sum **February 2005**

1713. *Proposed by Shawn Hedman and David Rose, Florida Southern College, Lakeland, FL.*

Prove that

$$\sum_{n=4}^{\infty}\left(\sum_{k=2}^{n-2}\binom{n}{k}^{-1}\right)=\frac{3}{2}.$$

Many readers provided a solution along the following lines.

Interchanging the order of summation we have

$$\begin{aligned}\sum_{n=4}^{\infty}\left(\sum_{k=2}^{n-2}\binom{n}{k}^{-1}\right)&=\sum_{k=2}^{\infty}\left(\sum_{n=k+2}^{\infty}\binom{n}{k}^{-1}\right)\\&=\sum_{k=2}^{\infty}\left(\frac{k!}{k-1}\sum_{n=k+2}^{\infty}\frac{k-1}{P(n,k)}\right)\\&=\sum_{k=2}^{\infty}\left(\frac{k!}{k-1}\sum_{n=k+2}^{\infty}\left(\frac{1}{P(n-1,k-1)}-\frac{1}{P(n,k-1)}\right)\right)\\&=\sum_{k=2}^{\infty}\frac{k!}{k-1}\cdot\frac{1}{P(k+1,k-1)}\\&=\sum_{k=2}^{\infty}\left(\frac{1}{k-1}-\frac{1}{k+1}\right)=\frac{3}{2}.\end{aligned}$$

Note. This problem appeared in The College Mathematics Journal, Vol. 30, No. 5, (November 1999), 409–410, and, in slightly altered form, in the article "Infinite Series with Binomial Coefficients" by Courtney Moen, Mathematics Magazine, Vol. 64, No. 1, (Feb. 1991), 53–55.

Also solved by JPV Abad, Reza Akhlaghi, Tsehaye Andeberhan, Michael Andreoli, Michel Bataille (France), Jean Bogaert (Belgium), Paul Bracken and Nadi Cantroll, Brian Bradie, Alper Cay (Turkey), John Christopher, Wenchang Chu (Italy), Con Amore Problem Group (Denmark), Chip Curtis, Knut Dale (Norway), Charles R. Diminnie, Daniele Donini (Italy), Robert L. Doucette, Michael Goldenberg and Mark Kaplan, G.R.A.20 Math Problems Group (Italy), Mowaffaq Hajja (Jordan), Mitch Harris, Russell Jay Hendel, James C. Hickman, Houghton College Math Club, Harris Kwong, Elias Lampakis (Greece), Carl Libis, Peter W. Lindstrom, James Marengo, Kim McInturff, Rob Pratt, Muneer Ahmad Rashid (Australia), Henry Ricardo, Rolf Richberg (Germany), Volkhard Schindler (Germany), Heinz-Jürgen Seiffert (Germany), Nicholas C. Singer, Byron Siu, Christopher J. Smith, John W. Spellman and Ricardo M. Torrejón, Albert Stadler (Switzerland), Fernand R. Tessier (Kuwait), Michael Vowe (Switzerland), Paul Weisenhorn (Germany), Li Zhou, and the proposer.

A Lower Bound **February 2005**

1714. *Mohammed Aassila, Strasbourg, France.*

Let m, n, x, y, z be positive real numbers with $x+y+z=1$. Prove that

$$\frac{x^4}{(mx+ny)(my+nx)}+\frac{y^4}{(my+nz)(mz+ny)}+\frac{z^4}{(mz+nx)(mx+nz)}\geq\frac{1}{3(m+n)^2}.$$

Solution by Robert L. Doucette, McNeese State University, Lake Charles, LA.

Applying the inequality

$$\frac{1}{ab}\geq\frac{4}{(a+b)^2},\, a,b>0,$$

to each term on the left side of the inequality yields

$$\begin{aligned}
&\frac{x^4}{(mx+ny)(my+nx)}+\frac{y^4}{(my+nz)(mz+ny)}+\frac{z^4}{(mz+nx)(mx+nz)}\\
&\geq \frac{4x^4}{(mx+ny+my+nx)^2}+\frac{4y^4}{(my+nz+mz+ny)^2}+\frac{4z^4}{(mz+nx+mx+nz)^2}\\
&=\frac{4}{(m+n)^2}\left(\frac{x^4}{(x+y)^2}+\frac{y^4}{(y+z)^2}+\frac{z^4}{(z+x)^2}\right). \qquad (3)
\end{aligned}$$

Next use two applications of the Cauchy-Schwarz inequality to see

$$\begin{aligned}
1 &= x+y+z\\
&= \frac{x}{\sqrt{x+y}}\sqrt{x+y}+\frac{y}{\sqrt{y+z}}\sqrt{y+z}+\frac{z}{\sqrt{z+x}}\sqrt{z+x}\\
&\leq \left(\frac{x^2}{x+y}+\frac{y^2}{y+z}+\frac{z^2}{z+x}\right)^{1/2}(x+y+y+z+z+x)^{1/2}\\
&= \sqrt{2}\left(1\cdot\frac{x^2}{x+y}+1\cdot\frac{y^2}{y+z}+1\cdot\frac{z^2}{z+x}\right)^{1/2}\\
&\leq \sqrt{2}\sqrt[4]{3}\left(\frac{x^4}{(x+y)^2}+\frac{y^4}{(y+z)^2}+\frac{z^4}{(z+x)^2}\right)^{1/4},
\end{aligned}$$

so

$$\frac{x^4}{(x+y)^2}+\frac{y^4}{(y+z)^2}+\frac{z^4}{(z+x)^2}\geq\frac{1}{12}. \qquad (4)$$

Combining (1) and (2) gives the desired result.

Also solved by Arkady Alt, Tsehaye Andeberhan, Michel Bataille (France), Erhard Braüne (Austria), Minh Can, Alper Cay (Turkey), Wenchang Chu (Italy), Daniele Donini (Italy), Charles R. Diminnie, Elnur Emrah (Turkey), Michael Goldenberg and Mark Kaplan, Elias Lampakis (Greece), Phil McCartney, Michael G. Neubauer, Northwestern University Math Problem Solving Group, Rolf Richberg (Germany), Heinz-Jürgen Seiffert (Germany), Achilleas Sinefakopoulos, Paul Weisenhorn (Germany), Ding Ya-yuan and Guo Yao-Hong (China), Li Zhou, and the proposer.

What Did You Expect? **February 2005**

1715. *Proposed by Barthel Wayne Huff, Salt Lake City, UT.*

An urn contains 35 red balls, labeled $1, 2, \ldots, 35$, and k blue balls. Balls are drawn one at a time at random, identified by number, then replaced, until a blue ball is drawn. Some red balls may be drawn more than once before the first blue is drawn. What is the minimal value of k for which the expected number of repetitions is less than 1?

Solution by C. Ray Rosentrater, Westmont College, Santa Barbara, CA.

We analyze the more general problem in which there are m red balls and k blue balls. Suppose that n red balls are drawn (with replacement) and let D_n denote the number of distinct balls drawn. Then conditioning on whether the $(n+1)$st ball is different from each of the first n drawn, we have

$$E[D_{n+1}]=\sum_{j=1}^{m}\left(j\frac{j}{m}+(j+1)\left(1-\frac{j}{m}\right)\right)\cdot P[D_n=j]$$

$$= \sum_{j=1}^{m} jP[D_n = j] + \sum_{j=1}^{m}\left(1 - \frac{j}{m}\right)P[D_n = j]$$

$$= \left(1 - \frac{1}{m}\right)E[D_n] + 1.$$

Noting that $E[D_0] = 0$, it follows by induction that $E[D_n] = m(1 - q^n)$, where

$$q = 1 - \frac{1}{m}.$$

Now let

$$p = \frac{m}{m+k}$$

be the probability that a red ball is selected for a particular draw. Then the number n of red balls drawn before the first blue ball is selected has a geometric distribution with probability density function f given by $f(n) = p^n(1-p)$ and expected value $\frac{p}{1-p}$. Because the number of repetitions in n draws is $n - D_n$, the expected number of repetitions before drawing the first blue ball is

$$E[R] = \sum_{n=0}^{\infty}(n - E[D_n])p^n(1-p)$$

$$= \frac{p}{1-p} - \sum_{n=0}^{\infty} m(1-q^n)p^n(1-p)$$

$$= \frac{p}{1-p} - m + m(1-p)\frac{1}{1-pq} = \frac{m}{k(k+1)}.$$

This expected value is less than 1 if and only if $m < k(k+1)$, that is, if and only if

$$k \geq \frac{1}{2}\left(\sqrt{4m+1} - 1\right).$$

With $m = 35$, this becomes

$$k \geq \frac{1}{2}\left(\sqrt{141} - 1\right) \approx 5.44.$$

Thus, if there are 35 red balls, then the minimal number of blue balls for which the expected number of repetitions is less than 1 is $k = 6$.

Also solved by, Jack Abad, Michael Andreoli, Herb Bailey, Frederick H. Chen, Randall J. Covill, Daniele Donini (Italy), Robert L. Doucette, Jerrold W. Grossman, Houghton College Math Club, Kathleen E. Lewis, Rob Pratt, Dexter Senft, Nicholas C. Singer, Albert Stadler (Switzerland), and the proposer. There were six incorrect submissions.

Answers

Solutions to the Quickies from page 68.

A957. Observe that

$$(m \pm \sqrt{n})^{2k} = \left((m^2+n) \pm 2m\sqrt{n}\right)^k = (2L \pm 2m\sqrt{n})^k = 2^k M \pm 2^k N\sqrt{n},$$

for some positive integers L, M, N. Because n is not a perfect square and

$$\sqrt{n} < m + 1 < \sqrt{n} + 2,$$

it follows that $|m - \sqrt{n}| < 1$. Thus

$$2^{k+1}M = (m + \sqrt{n})^{2k} + (m - \sqrt{n})^{2k} = \lfloor (m + \sqrt{n})^{2k} + 1 \rfloor .$$

A958. Let $p = \frac{1}{2}(a + b + c)$ and note that because the triangle is nondegenerate, $p - a$, $p - b$, $p - c$ are all positive. Part a) follows by observing that

$$0 < p(p-a)(p-b)(p-c) = \frac{2(a^2b^2 + b^2c^2 + c^2a^2) - a^4 - b^4 - c^4}{16},$$

Part b) follows from

$$\begin{aligned} 0 &< (p-a)(p-b)(p-c) \\ &= \frac{ab(a+b) + bc(b+c) + ca(c+a) - a^3 - b^3 - c^3 - 2abc}{8}. \end{aligned}$$

The Eightfold Way[1], Lie Algebra, and Spider Hunting in the Dark[2]

Spider, spider, burning bright
Like a beacon in the night,
What algebraist in the skies
Gave you eight legs and then eight eyes?
Was Buddha watching you at play
When he conceived the Eightfold Way?
And did your eightness give a key
To Gell-Mann: quarks and SU3?

——J. D. MEMORY
PROFESSOR OF PHYSICS, EMERITUS
NORTH CAROLINA STATE UNIVERSITY
jmemory@nc.rr.com

1. Gell-Mann originally dubbed his quark theory of hadrons (strongly interacting particles) as the "Eightfold Way" in allusion to the Buddha's prescription for the path to enlightenment, because he was able to relate a family of eight hadrons to the eight generators of the Lie-group SU3 (special unitary group in three dimensions).

2. For those unfamiliar with this childhood pastime, one can locate a spider at a considerable distance at night by looking down the beam of a flashlight. They gleam brilliantly.

REVIEWS

PAUL J. CAMPBELL, *Editor*
Beloit College

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

Numb3rs. Television series on CBS, 10 P.M. ET & PT Fridays. 13 episodes in Spring 2005, 24 episodes 2005–06. Created by Cheryl Heuton and Nick Falacci, with Ridley Scott and Tony Scott as executive producers. Different writers and directors for each episode. Stars Rob Morrow, David Krumholtz, Judd Hirsch, and others. Home page: `http://www.cbs.com/primetime/numb3rs/index.shtml` . Episode summaries: `www.numb3rs.org/spoilers.htm`, `http://www.tv.com/numb3rs/show/25043/episode_guide.html` . List of all the mathematics and science concepts used in the series, with links to on-line sources: `http://www.redhawke.org/content/blogcategory/5/34/` . Mathematical comments on the epsisodes by Mark Bridger (Mathematics Dept., Northeastern University): `http://www.atsweb.neu.edu/math/cp/blog/`.

"We all use math every day: to predict the weather, to tell time, to handle money. Math is more than formulas and equations. It's logic, it's rationality, it's using your mind to solve the biggest mysteries we know." Every episode of *Numb3rs*, arguably the smartest show on TV (and not just because of the math), begins with this "public service announcement" on behalf of mathematics. At last there is a TV show with a mathematician as a lead character! It is a Nielsen top-30 show and number one on Friday nights. I missed it last spring while in Germany (where those episodes played this past fall, dubbed). The premise is that Charlie Eppes, young "genius" professor of mathematics at "CalSci" in L.A., helps his FBI-agent brother Don catch criminals. The mathematics of telling time and handling money may be the banal hook of familiarity to make the audience comfortable, but Charlie actually uses advanced mathematics—very briefly but cleverly exposited in the episodes—to help solve cases. Moreover, he is presented as an actual person with friends and family and as someone to be admired. The series intends to highlight different ways of thinking of the two brothers: "highly ordered logical intuition" (FBI agent) vs. "logic and evidence" (mathematician). Critics like the premise, the acting, and the story lines; and the mathematics is right—Gary Lorden (Chair, Mathematics Dept., Caltech) vets the mathematical content of the scripts. Texas Instruments provides an "inspirational" poster for teachers ("We all use," etc., in huge letters) plus high-school-level lesson plans in advance of episodes. Mark Bridger's blog, however, gives fuller exposition of the mathematics involved. What should mathematicians make of their "breakthrough" into mainline media, with a mid-season show popular enough to get renewed at least through this spring? That's hard for me to discern. So far, no one (friends, neighbors, acquaintances, co-workers) has asked what I think about the series, except for one student, who succeeds in rallying only non-math majors to watch episodes. (Do math majors have better things to do on Friday evenings? Do they see mathematics as relevant to their lives outside the classroom?) It is easy for a mathematician to scoff at glibness in Charlie's lines or to note that his physicist colleague Larry sometimes has quicker insights. But are we mathematicians making the most of what may be the opportunity of a lifetime? Dialogue with students about *Numb3rs* is a natural opportunity that could help increase the number of bachelor's degrees in the mathematical sciences, as the *CUPM Curriculum Guide 2004* (`http://www.maa.org/cupm/`) recommends. Sure, trying to explain on short notice some of the mathematics behind episodes might be demanding, but Bridger's blog is a helpful crutch. Why not try it, even on an informal basis, for the rest of this spring? (Thanks to Bill Nye the Science Guy for inspiring this series; may his science show for kids return to PBS and rerun forever. And thanks to Sarah Price for recording episodes.)

NEWS AND LETTERS

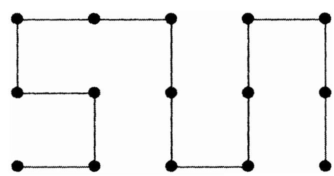
66th Annual William Lowell Putnam Mathematical Competition

Editors Note: Additional solutions will be printed in the *Monthly* later in the year.

PROBLEMS

A1. Show that every positive integer is a sum of one or more numbers of the form $2^r 3^s$, where r and s are nonnegative integers and no summand divides another. (For example, $23 = 9 + 8 + 6$.)

A2. Let $\mathcal{S} = \{(a, b) | a = 1, 2, \dots, n, b = 1, 2, 3\}$. A *rook tour* of $\mathcal{S}$ is a polygonal path made up of line segments connecting points $p_1, p_2, \dots, p_{3n}$ in sequence such that $(i) p_i \in \mathcal{S}$, $(ii) p_i$ and p_{i+1} are a unit distance apart, for $1 \le i < 3n$, (iii) for each $p \in \mathcal{S}$ there is a unique i such that $p_i = p$. How many rook tours are there that begin at $(1, 1)$ and end at $(n, 1)$? (An example of such a rook tour for $n = 5$ is depicted below.)



A3. Let $p(z)$ be a polynomial of degree $n \ge 1$, all of whose zeros have absolute value 1 in the complex plane. Put $g(z) = p(z)/z^{n/2}$. Show that all zeros of $g'(z)$ have absolute value 1.

A4. Let H be an $n \times n$ matrix all of whose entries are ± 1 and whose rows are mutually orthogonal. Suppose H has an $a \times b$ submatrix whose entries are all 1. Show that $ab \le n$.

A5. Evaluate $\int_0^1 \frac{\ln(x+1)}{x^2+1}\, dx$.

A6. Let n be given, $n \ge 4$, and suppose that $P_1, P_2, \dots, P_n$ are n randomly, independently and uniformly, chosen points on a circle. Consider the convex n-gon whose vertices are the P_i. What is the probability that at least one of the vertex angles of this polygon is acute?

B1. Find a nonzero polynomial $P(x, y)$ such that $P(\lfloor a \rfloor, \lfloor 2a \rfloor) = 0$ for all real numbers a. (Note: $\lfloor v \rfloor$ is the greatest integer less than or equal to v.)

B2. Find all positive integers $n, k_1, \dots, k_n$ such that $k_1 + \cdots + k_n = 5n - 4$ and

$$\frac{1}{k_1} + \cdots + \frac{1}{k_n} = 1.$$

B3. Find all differentiable functions $f : (0, \infty) \longrightarrow (0, \infty)$ for which there is a positive real number a such that

$$f'\left(\frac{a}{x}\right) = \frac{x}{f(x)}$$

for all $x > 0$.

B4. For positive integers m and n, let $f(m, n)$ denote the number of n-tuples $(x_1, x_2, \dots, x_n)$ of integers such that $|x_1| + |x_2| + \cdots + |x_n| \le m$. Show that $f(m, n) = f(n, m)$.

B5. Let $P(x_1, \dots, x_n)$ denote a polynomial with real coefficients in the variables $x_1, \dots, x_n$, and suppose that

(a) $\left(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}\right) P(x_1, \ldots, x_n) = 0$ (identically) and that

(b) $x_1^2 + \cdots + x_n^2$ divides $P(x_1, \ldots, x_n)$.

Show that $P = 0$ identically.

B6. Let S_n denote the set of all permutations of the numbers $1, 2, \ldots, n$. For $\pi \in S_n$, let $\sigma(\pi) = 1$ if π is an even permutation and $\sigma(\pi) = -1$ if π is an odd permutation. Also, let $\nu(\pi)$ denote the number of fixed points of π. Show that

$$\sum_{\pi \in S_n} \frac{\sigma(\pi)}{\nu(\pi) + 1} = (-1)^{n+1} \frac{n}{n+1}.$$

SOLUTIONS

Solution to A1. We argue by induction. Clearly $1 = 2^0 3^0$, so the number 1 is represented. If n is even, then by the induction hypothesis $n/2$ has such a representation, and it suffices to multiply each summand by 2. Suppose n is odd, and let k be the largest integer such that $3^k \le n$. If $n = 3^k$, then we are done. If $3^k < n < 3^{k+1}$, then $m = (n - 3^k)/2$ is a positive integer. Since $m < n$, it follows by the inductive hypothesis that we can write $m = \sum_i 2^{r_i} 3^{s_i}$. Then $n = 3^k + \sum_i 2^{r_i+1} 3^{s_i}$. It remains to show that no summand divides another. Using the inductive hypothesis, $2^{r_i+1} 3^{s_i}$ divides $2^{r_j+1} 3^{s_j}$ only when $i = j$. Also, $2^{r_i+1} 3^{s_i}$ does not divide 3^k because 2 is a factor of $2^{r_i+1} 3^{s_i}$ but not a factor of 3^k. Finally, $m = (n - 3^k)/2 < (3^{k+1} - 3^k)/2 = 3^k$. Thus, $2^{r_i+1} 3^{s_i} < 3^k$ and therefore 3^k does not divide $2^{r_i+1} 3^{s_i}$.

Solution to A2. Let $T(n)$ be the number of rook tours of $\mathcal{S}_n \equiv \mathcal{S}$ that begin at $(1, 1)$ and end at $(n, 1)$ and of these, let $T_H(n)$ and $T_V(n)$ be the number whose first move is horizontal or vertical respectively. Similarly, let $U(n)$ be the number of rook tours of $\mathcal{S}_n$ that begin at $(1, 1)$ and end at $(n, 3)$, and of these, let $U_H(n)$ and $U_V(n)$ be the number whose first move is horizontal or vertical respectively.

Clearly, for $n \ge 3$,

$$T(n) = T_H(n) + T_V(n) \quad \text{and} \quad U(n) = U_H(n) + U_V(n), \tag{1}$$

and

$$T_V(n) = U(n-1) \quad \text{and} \quad U_V(n) = T(n-1). \tag{2}$$

But in addition,

$$T_H(n) = T(n-1) \quad \text{and} \quad U_H(n) = U(n-1). \tag{3}$$

For the first of these claims, note that a rook tour counted by $T_H(n)$ must contain line segments connecting points (2, 2), (1, 2), (1, 3), (2, 3). Replace these with the single segment from (2, 2) to (2, 3) and remove the segment from (1, 1) to (2, 1) to obtain a rook tour counted by $T(n-1)$. This construction is reversible. The argument for the second claim is similar.

Substituting (2) and (3) into (1) yields $T(n) = U(n-1) + T(n-1) = U(n)$ and now, from the fact that $T(2) = U(2) = 1$, it easily follows that $T(n) = 2T(n-1) = 2^{n-2}$.

Solution to A3. Write $p(z) = c(z - z_1)(z - z_2) \cdots (z - z_n)$. If $|z| < 1$ (resp., $|z| > 1$), then

$$\operatorname{Re}\left(\frac{z}{z - z_k}\right) < \frac{1}{2} \quad \left(\text{resp.,} > \frac{1}{2}\right),$$

so that

$$\operatorname{Re}\left(\sum_{k=1}^{n} \frac{z}{z - z_k}\right) < \frac{n}{2} \quad \left(\text{resp.,} > \frac{n}{2}\right),$$

which is to say,

$$\operatorname{Re}\left(\frac{zp'(z)}{p(z)}\right) < \frac{n}{2} \quad \left(\text{resp.,} > \frac{n}{2}\right).$$

So, for $|z| \neq 1$,

$$\frac{zp'(z)}{p(z)} \neq \frac{n}{2}.$$

But $g(z) = p(z)z^{-n/2}$, so

$$\frac{g'(z)}{g(z)} = \frac{p'(z)}{p(z)} - \frac{n}{2z} = \frac{1}{z}\left(\frac{zp'(z)}{p(z)} - \frac{n}{2}\right).$$

So, from the first paragraph, $|z| \neq 1$ implies $g'(z) \neq 0$.

Solution to A4. Let M denote the $n \times n$ matrix obtained by dividing each element of H by $\sqrt{n}$. Thus M is an orthogonal matrix, which is to say that the rows $\mathbf{m}_1, \ldots, \mathbf{m}_n$ form an orthonormal basis for $\mathbf{R}^n$. Consequently, any $\mathbf{v} \in \mathbf{R}^n$ is uniquely expressed as a linear combination of the $\mathbf{m}_j$, $\mathbf{v} = c_1\mathbf{m}_1 + \cdots + c_n\mathbf{m}_n$, the c_i are given by the formula $c_i = \mathbf{v} \cdot \mathbf{m}_i$, and by Pythagoras $c_1^2 + \cdots + c_n^2 = \|\mathbf{v}\|^2$. Let I and J be intervals chosen so that the submatrix in question is obtained by restricting i to I and j to J. Take $\mathbf{v} = [v_j]$ so that $v_j = 1$ for $j \in J$ and $v_j = 0$ otherwise. Then $\|\mathbf{v}\|^2 = b$, and $c_i = b/\sqrt{n}$ for $i \in I$. Thus the left hand side above is $\geq ab^2/n$. Hence $ab^2/n \leq b$, which gives the result.

Solution to A5. We make the change of variable $u = \arctan x$. Then $du = dx/(1+x^2)$ and $x = \tan u$. Hence the integral in question is

$$\int_0^{\pi/4} \ln(1 + \tan u)\, du. \tag{1}$$

But $1 + \tan u = (\sin u + \cos u)/\cos u$, so the above is

$$= \int_0^{\pi/4} \ln(\sin u + \cos u)\, du - \int_0^{\pi/4} \ln \cos u\, du.$$

But

$$\cos(\pi/4 - u) = \cos(\pi/4)\cos u + \sin(\pi/4)\sin u = (\sin u + \cos u)/\sqrt{2}.$$

Hence (1) is

$$= \frac{\pi}{8}\ln 2 + \int_0^{\pi/4} \ln(\cos(\pi/4 - u))\, du - \int_0^{\pi/4} \ln\cos u\, du.$$

Here the last two integrals are equal, as we see by making the change of variable $v = \pi/4 - u$ in the first one. Hence

$$\int_0^1 \frac{\ln(x+1)}{x^2+1}\, dx = \frac{\pi}{8}\ln 2.$$

Solution to A6. We first prove a simple lemma.

LEMMA. For $n \geq 4$ and an n-gon Q inscribed in a circle, if the angles at two vertices of Q are acute, then the vertices must be adjacent in Q. Hence there are at most two vertices of Q with an acute angle.

Proof of Lemma. Suppose A and D are nonadjacent vertices of Q. Then proceeding counterclockwise around the circle starting at A, we must have a first vertex B after A and a last vertex C before reaching D, with possibly $B = C$. Continuing, there must be a first vertex E after D and a last vertex F before returning to A, with possibly $E = F$. Let $m(\overset{\frown}{XY})$ be the angle measure of the arc counterclockwise around the circle from X to Y. Then the points in counterclockwise order are A, B, C, D, E, and F possibly with $B = C$, where we take $m(\overset{\frown}{BC}) = 0$, or with $E = F$, where we take $m(\overset{\frown}{EF}) = 0$. Note that $\angle A$ is inscribed in the arc from B to F so has measure one-half of $m(\overset{\frown}{BF})$. Likewise $\angle D$ has measure one-half of $m(\overset{\frown}{EC})$. Hence

$$\begin{aligned} m(\angle A)+m(\angle D) &= \tfrac{1}{2}(m(\widehat{BF})+m(\widehat{EC})) \\ &= \tfrac{1}{2}\big(m(\widehat{BC})+m(\widehat{CD})+m(\widehat{DE})+m(\widehat{EF}) \\ &\quad + m(\widehat{EF})+m(\widehat{FA})+m(\widehat{AB})+m(\widehat{BC})\big) \\ &\geq \tfrac{1}{2}\left(m(\widehat{BC})+m(\widehat{CD})+m(\widehat{DE})+m(\widehat{EF})+m(\widehat{FA})+m(\widehat{AB})\right) \\ &= \pi. \end{aligned}$$

Since the sum of $\angle A$ and $\angle D$ is at least π, at most one of these can be less than $\pi/2$, that is, at most one is acute. If there were more than two vertices with acute angles in Q, then two of these would not be adjacent, contradicting the preceding analysis. Hence at most two vertices of Q can have acute angles, and then only if the two vertices having acute angles are adjacent.

For $n \geq 4$, let $X_1, X_2, \ldots, X_n$ be the n points. Let E_{ij} be the event that X_j is the next vertex counterclockwise from X_i, $\angle X_i$ is acute, but $\angle X_j$ is not. Then the events E_{ij} are mutually exclusive and include all of the cases where Q has an acute angle. If there are two acute angles on Q, say at X_i and X_k, then X_k is adjacent to X_i, and if X_k is clockwise from X_i, say, and X_j is counterclockwise from X_i, then E_{ij} holds.

By symmetry, the probability of the events E_{ij} are equal, so we may as well compute the probability of E_{12}. By symmetry, this probability is independent of the location of X_1, but given X_1, it is a function of $\theta = m(\widehat{X_1X_2})$, uniformly distributed between 0 and 2π. If $\theta \geq \pi$ and X_2 was the next point counterclockwise from X_1, then $\angle X_2$ is acute and E_{12} fails. Assume $\theta < \pi$ and let X_1' be the antipodal point to X_1 and X_2' the antipodal point to X_2. Then E_{12} holds if and only if the points X_3 to X_n all lie on the arc from X_2 counterclockwise to X_2', so that X_2 is the first point counterclockwise from X_1 and $\angle X_1$ is acute, but not all these lie on the arc counterclockwise from X_1' to X_2', so that $\angle X_2$ is acute. Since the X_3 to X_n are uniformly distributed independent of X_1 and X_2, the probability of E_{12} given angle θ is

$$P(E_{12}|\theta) = \left(\frac{m(\widehat{X_2X_2'})}{2\pi}\right)^{n-2} - \left(\frac{m(\widehat{X_1X_2'})}{2\pi}\right)^{n-2} = \frac{1}{2^{n-2}} - \left(\frac{\theta}{2\pi}\right)^{n-2}.$$

Hence the probability of E_{12} is

$$P(E_{12}) = \int_0^{\pi}\left[\frac{1}{2^{n-2}} - \left(\frac{\theta}{2\pi}\right)^{n-2}\right]\frac{d\theta}{2\pi} = \frac{1}{2^{n-1}} - \frac{1}{(n-1)2^{n-1}}.$$

Hence the probability P that some angle is acute is

$$P = n(n-1)P(E_{12}) = \frac{n(n-2)}{2^{n-1}}.$$

Solution to B1. One such polynomial is $P(x, y) = (2x - y)(2x - y + 1)$. To see this, let a be a real number and $n = \lfloor a \rfloor$. If $a \in [n, n + \frac{1}{2})$ then $\lfloor 2a \rfloor = 2\lfloor a \rfloor$ so $P(\lfloor a \rfloor, \lfloor 2a \rfloor) = 0$. If $a \in [n + \frac{1}{2}, n + 1)$ then $\lfloor 2a \rfloor = 2\lfloor a \rfloor + 1$, so once again $P(\lfloor a \rfloor, \lfloor 2a \rfloor) = 0$.

Solution to B2. The possible solutions are

$$n = 1, k_1 = 1,$$

$$n = 3, \{k_1, k_2, k_3\} = \{2, 3, 6\},$$

$$n = 4, k_1 = k_2 = k_3 = k_4 = 4.$$

From the arithmetic mean–harmonic mean (or Cauchy-Schwarz) inequality,

$$(k_1 + \cdots + k_n)\left(\frac{1}{k_1} + \cdots + \frac{1}{k_n}\right) \geq n^2, \tag{1}$$

with equality if and only if $k_1 = \cdots = k_n$. Hence $5n - 4 \geq n^2$, which is equivalent to $(n-1)(n-4) \leq 0$. Thus $n \in \{1, 2, 3, 4\}$. For $n = 1$ or $n = 4$, we have equality in (1), so

$k_1 = \cdots = k_n$. We obtain the solutions $n = 1$, $k_1 = 1$, and $n = 4$, $k_1 = k_2 = k_3 = k_4 = 4$. We are left with the cases $n = 2$ and $n = 3$. For $n = 2$, the system of equations $k_1 + k_2 = 6$, $1/k_1 + 1/k_2 = 1$ is not solvable in positive integers. For $n = 3$, we seek the triples (k_1, k_2, k_3) such that $k_1 + k_2 + k_3 = 11$ and $1/k_1 + 1/k_2 + 1/k_3 = 1$. Let $k_1k_2 + k_2k_3 + k_3k_1 = k_1k_2k_3 = q$. Then k_1, k_2, k_3 are positive integral solutions to the equation $x^3 - 11x^2 + qx - q = 0$. It follows that a solution x is an integer different from 1 and 11, and that

$$q = \frac{-x^3 + 11x^2}{x - 1} = -x^2 + 10x + 10 + \frac{10}{x - 1}.$$

Because q is a positive integer, $x - 1$ is a positive divisor of 10 and different from 10. Then $x - 1 \in \{1, 2, 5\}$, so $x \in \{2, 3, 6\}$. A simple case analysis shows that $\{k_1, k_2, k_3\} = \{2, 3, 6\}$.

Solution to B3. Taking the derivative of each side of $f'(x) = a/\big(xf(\frac{a}{x})\big)$ yields

$$\begin{aligned} f''(x) &= \frac{-a\left[f\left(\frac{a}{x}\right) + xf'\left(\frac{a}{x}\right)\left(-\frac{a}{x^2}\right)\right]}{x^2\left(f\left(\frac{a}{x}\right)\right)^2} \\ &= \frac{-a\left[\frac{a}{xf'(x)} + \frac{-a}{x}\left(\frac{x}{f(x)}\right)\right]}{x^2\left(\frac{a^2}{x^2(f'(x))^2}\right)} \\ &= \left[\frac{-f(x)}{xf(x)} + \frac{xf'(x)}{xf(x)}\right] f'(x) \end{aligned}$$

or equivalently,

$$\frac{f''(x)}{f'(x)} = -\frac{1}{x} + \frac{f'(x)}{f(x)}.$$

Thus $\ln f'(x) = -\ln x + \ln f(x) + \ln c$ for some $c > 0$, which is the same as

$$\frac{f'(x)}{f(x)} = \frac{c}{x}.$$

Another antiderivative gives $\ln f(x) = c \ln x + \ln d$ for some $d > 0$, and it follows that $f(x) = dx^c$. Substituting into the original equation yields

$$cd\left(\frac{a}{x}\right)^{c-1} = \frac{1}{dx^{c-1}}$$

or equivalently, $d^2 = \dfrac{1}{ca^{c-1}}$, and therefore

$$f(x) = \frac{1}{\sqrt{c\,a^{(c-1)}}}x^c = \sqrt{\frac{a}{c}}\left(\frac{x}{\sqrt{a}}\right)^c \qquad c > 0.$$

Solution to B4. Let $F_{(m,n)}$ denote the set of lattice points $(x_1, x_2, \ldots, x_n)$ in $\mathbf{Z}^n$ such that $|x_1| + |x_2| + \cdots + |x_n| \le m$, and define a mapping $\varphi_{(m,n)} : F_{(m,n)} \to F_{(n,m)}$ as follows. Let x be an arbitrary element of $F_{(m,n)}$, and reading from right to left in x, denote the nonzero coordinates by $a_1, a_2, \ldots, a_k$; that is,

$$x = (\underbrace{0, 0, 0, \ldots, 0}_{a}, \underbrace{a_k, 0, 0, \ldots, 0}_{|b_k|}, a_{k-1}, \ldots, \underbrace{a_2, 0, 0, \ldots, 0}_{|b_2|}, \underbrace{a_1, 0, 0, \ldots, 0}_{|b_1|})$$

where for each i, b_i is chosen to have the same sign as a_i, and a is chosen so that $a + |b_1| + |b_2| + \cdots |b_k| = n$ (note that $a \ge 0$ because x is an n-tuple). Now define $\varphi_{(m,n)}(x) = (y_1, y_2, \ldots y_m)$ to be

$$(\underbrace{0,0,0,\ldots,0}_{b},\underbrace{b_k,0,0,\ldots,0}_{|a_k|}),b_{k-1},\ldots,\underbrace{b_2,0,0,\ldots 0}_{|a_2|},\underbrace{b_1,0,0,\ldots,0}_{|a_1|}))$$

where b is chosen so that $b+|a_1|+|a_2|+\cdots+|a_k|=m$ (note that $b\geq 0$ because $|a_1|+|a_2|+\cdots+|a_k|=|x_1|+|x_2|+\cdots|x_n|\leq m$). Also, note that this element is in $F_{(n,m)}$ because it is an m-tuple and $|y_1|+|y_2|+\cdots|y_m|=|b_1|+|b_2|+\cdots+|b_k|\leq n$.

It is easy to check that $\varphi_{(m,n)}\circ\varphi_{(n,m)}$ and $\varphi_{(n,m)}\circ\varphi_{(m,n)}$ are identity functions, so the correspondence is one-to-one and onto. It follows that $f(m,n)=f(n,m)$.

Solution to B5. Write $P=\sum_k P_k$ where P_k is homogeneous of degree k. If P has the required properties, then each of the P_k does. Thus we may assume that P is homogeneous, say of degree d. We note that if $i_1+\cdots+i_n=j_1+\cdots j_n$, then

$$\frac{\partial^{i_1}}{\partial x_1^{i_1}}\cdots\frac{\partial^{i_n}}{\partial x_n^{i_n}}x_1^{j_1}\cdots x_n^{j_n}=\begin{cases}i_1!\cdots i_n! & \text{if } i_1=j_1,\ldots,i_n=j_n,\\ 0 & \text{otherwise.}\end{cases}$$

So if

$$P=\sum_{\substack{i_1,\ldots,i_n,i_k\geq 0\\ \sum i_k=d}} a(i)x_1^{i_1}\cdots x_n^{i_n},$$

then

$$P\left(\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_n}\right)P(x_1,\ldots,x_n)=\sum_{\substack{i_1,\ldots,i_n,i_k\geq 0\\ \sum i_k=d}} a(i)^2x_1^{i_1}\cdots x_n^{i_n}.$$

We note that the above is a strictly positive constant unless $a(i)=0$ for all i. Suppose that P is not identically zero, let $S=x_1^2+\cdots+x_n^2$, and suppose that S divides P, say $P=SQ$. Put

$$R=\left(\frac{\partial^2}{\partial x_1^2}+\cdots+\frac{\partial^2}{\partial x_n^2}\right)P(x_1,\ldots,x_n).$$

Then R is not identically zero, because

$$\begin{aligned}&Q\left(\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_n}\right)R(x_1,\ldots,x_n)\\ &=Q\left(\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_n}\right)\left(\left(\frac{\partial^2}{\partial x_1^2}+\cdots+\frac{\partial^2}{\partial x_n^2}\right)P(x_1,\ldots,x_n)\right)\\ &=P\left(\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_n}\right)P(x_1,\ldots,x_n)>0.\end{aligned}$$

Solution to B6. Let t be a real variable, and let $A(t)=[a_{ij}]$ denote the $n\times n$ matrix for which $a_{ii}=t$ and $a_{ij}=1$ for $i\neq j$. We show first that

$$\det A(t)=(t+n-1)(t-1)^{n-1}. \tag{1}$$

To see this, add rows 2 through n to the first row. This does not change the determinant. The resulting entries in the first row are all $t+n-1$. Factor this out. In the remaining matrix, all entries in the first row are 1. Subtract the first row from each of rows 2 through n. The result is that in rows 2 through n, all non-diagonal entries are 0, and the diagonal entries are $t-1$. Since this matrix is upper triangular, it is clear that it has determinant $(t-1)^{n-1}$. Thus we have (1).

In general, $\det[a_{ij}]=\sum_{\pi\in S_n}(-1)^{\sigma(\pi)}\prod_{i=1}^n a_{i\pi(i)}$. On applying this to $A(t)$, we find that

$$\sum_{\pi\in S_n}(-1)^{\sigma(\pi)}t^{\nu(\pi)}=(t-1)^n+n(t-1)^{n-1}.$$

Now integrate the above from 0 to 1 to obtain the stated identity.

CONTENTS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW
Washington, DC 20036