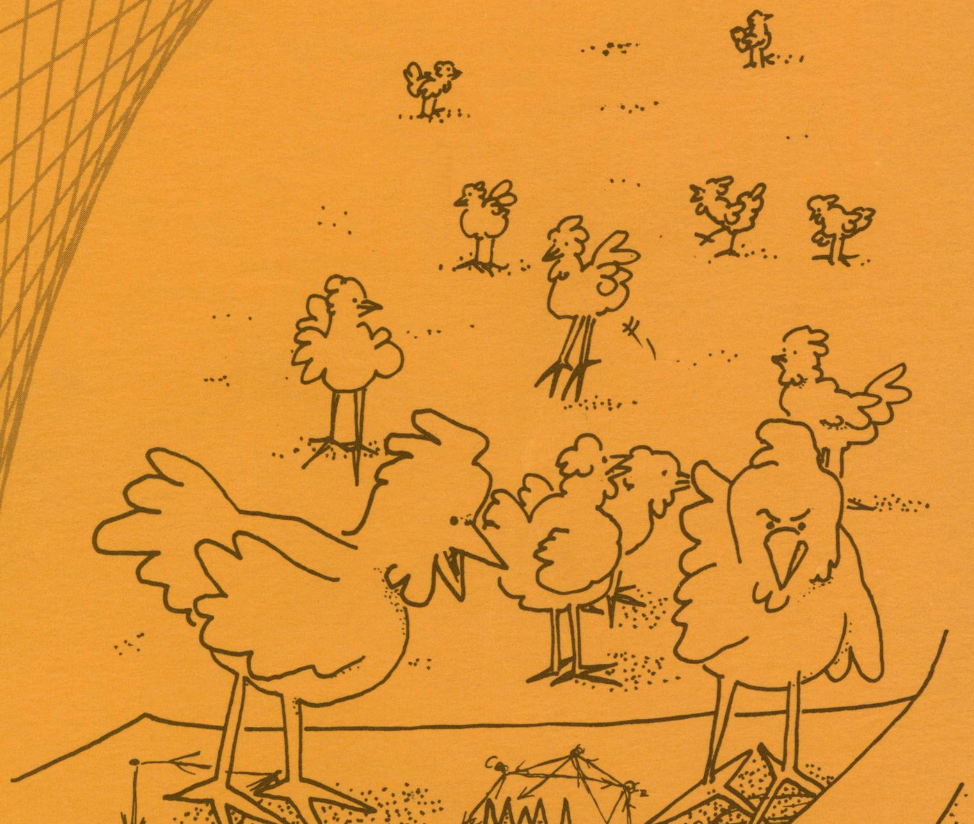
MATHEMATICS MAGAZINE



Vol. 53, No. 2
March, 1980

PECKING ORDERS • TOPOLOGY • PRETZEL TWISTING
LINEAR CONGRUENCES • MAGIC SQUARES AND CUBES

MATHEMATICS MAGAZINE

Volume 53, Number 2, March 1980
(ISSN—0025-570X)

COVER: If chickens are vertices and pecking makes an edge, then some models of this social structure lay an egg (p. 67).

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

EDITORIAL POLICY

Mathematics Magazine is a journal of collegiate mathematics designed to enrich undergraduate study of the mathematical sciences. The *Magazine* should be an inviting, informal journal emphasizing good mathematical exposition of interest to undergraduate students. Manuscripts accepted for publication in the *Magazine* should be written in a clear and lively expository style. The *Magazine* is not a research journal, so papers written in the terse "theorem-proof-corollary-remark" style will ordinarily be unsuitable for publication. Articles printed in the *Magazine* should be of a quality and level that makes it realistic for teachers to use them to supplement their regular courses. The editors especially invite manuscripts that provide insight into applications and history of mathematics. We welcome other informal contributions, for example, brief notes, mathematical games, graphics and humor.

New manuscripts should be sent to: Doris Schattschneider, Editor-Elect, Mathematics Magazine, Department of Mathematics, Moravian College, Bethlehem, Pennsylvania 18018. Manuscripts should be prepared in a style consistent with the format of *Mathematics Magazine.* They should be typewritten and double spaced on $8\frac{1}{2}$ by 11 paper. Authors should submit the original and one copy and keep one copy as protection against possible loss. Illustrations should be carefully prepared on separate sheets of paper in black ink, the original without lettering and two copies with lettering added; the printers will insert printed letters on the illustration in the appropriate locations.

ABOUT OUR AUTHORS

Stephen B. Maurer ("The King Chicken Theorems") grew up in Silver Spring, Maryland. He received his B.A. from Swarthmore College in 1967 and his Ph.D. from Princeton in 1972. He recently returned to Swarthmore as an associate professor, having taught previously at the Phillips Exeter Academy, the University of Waterloo, the Hampshire College Summer Studies in Mathematics, and Princeton. His research interests are in combinatorics, especially graphs, matroids, and partially ordered sets. The present article came to be because some of his Princeton students, intrigued by his mention of an old application of graph theory to chickens, started asking questions and getting some answers.

Donald E. Sanderson ("Advanced Plane Topology from an Elementary Standpoint") studied topology under R. H. Bing at the University of Wisconsin and is currently at Iowa State University. It was Professor Bing's comment, during a lecture on "The Elusive Fixed Point Property", on the lack of a simple "dog chasing a rabbit" type proof of the Brouwer Fixed Point Theorem in dimensions higher than one, that started Donald Sanderson on this chase. After catching the rabbit in the plane (on the Phragmen-Brouwer Property), the chase broadened and produced the "rabbit stew" presented here.

BUSINESS INFORMATION. Mathematics Magazine is published by the Mathematical Association of America at Washington, D.C., five times a year in January, March, May, September, and November. Ordinary subscriptions are $18 per year. Members of the Mathematical Association of America or of Mu Alpha Theta may subscribe at special reduced rates. Colleges and university mathematics departments may purchase bulk subscriptions (5 or more copies to a single address) for distribution to undergraduate students.

Subscription correspondence and notice of change of address should be sent to A. B. Willcox, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Back issues may be purchased, when in print, from P. and H. Bliss Co., Middletown, Connecticut 06457.

Advertising correspondence should be addressed to Raoul Hailpern, Mathematical Association of America, SUNY at Buffalo, Buffalo, New York 14214.

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

The King Chicken Theorems

A flock of results about pecking orders, describing possible patterns of dominance.

STEPHEN B. MAURER
Swarthmore College
Swarthmore, PA 19081

Fact: In any barnyard of chickens, in every pair of chickens, one is dominant over the other. The dominant chicken asserts its dominance by pecking the other chicken on the head and neck, whence the phrase "pecking order." However, it rarely happens that this pecking order is linear. That is, it is rare that there is a first chicken who pecks all the others, a second chicken who pecks all but the first, and so on. Question: Given that the pecking order may not be linear, is there still a reasonable way to designate a most dominant chicken, i.e., a king? In this paper, we analyze at length one promising king definition, and look briefly at several others.

This material illustrates nicely the delights and pitfalls of applied mathematics. One delight is that a model set up to describe one situation may well describe several others as well. A pitfall is that the mathematics done with the model, though good mathematics, may not really be of much use in the situation being modeled. It is not merely a matter of irrelevant theorems. It may be a matter of the wrong mathematical assumptions or even the wrong definitions. This last is a very subtle matter because it may be far from obvious that the definitions are wrong. Specifically, we wish to define a king chicken in such a way that the definition implies a strong sense of dominance and also implies that every flock of chickens has a king. We give such a definition and such a theorem—a definition and a theorem due to the mathematical sociologist H. G. Landau in 1953 [5]. But a king should also be unique, or at least rare. It is necessary to ask if our definition assures this. We are led to discover and prove many theorems which say that kings are all too common. In short, the theorems have negative implications about the usefulness of the definition. But we would not even have discovered this if we had not done some mathematics with the definition, and done it with this sort of testing in mind.

This lesson, that a model needs to be judged by the theorems it keeps, is perhaps the main point of this paper. I must confess, however, that for me the theorems themselves are also a main point; the pure mathematician in me finds them interesting in their own right. Also, their proofs illustrate some interesting techniques, for instance, duality and induction from n to $n+2$. In any event, good modeling does not stop with wrong definitions, no matter how pretty the theorems they give. One must look for better definitions and/or better mathematical assumptions, and then start proving theorems again. We say a bit about this second round towards the end of the paper.

This paper is written primarily for undergraduates, both in the style of proofs and in the amount of interpretation of theorems. There are problems stated later which could lead to good undergraduate projects. Indeed, the paper developed from work begun with undergraduates—my 1977 Graph Theory class at Princeton. Some of the results appear to be new.

A word about sex: there are female chickens, called hens, and there are male chickens, called roosters. To this natural division mankind has also added neuter chickens, called capons. These divisions are important to chickens. Chickens of the same sex residing together always form a pecking order, but roosters and hens rarely peck each other. Thus the "Fact" stated at the beginning of this article is not quite correct; one must restrict one's attention to a single sex. Since a traditional barnyard has several hens but only one or two roosters, pecking is usually associated with hens, both in common parlance ("henpecked") and in scholarly articles. For this reason several people have suggested rather heatedly that I should revise this article by replacing "king" throughout with "queen." However, since all chickens peck, I have stuck with king. I use this as a general term for leader. I have also avoided all use of either male or female pronouns for chickens.

For a delightful and informative essay on chicken society in general and pecking in particular, one should read the original article on the subject by Schjelderup-Ebbe **[16]**.

The Model

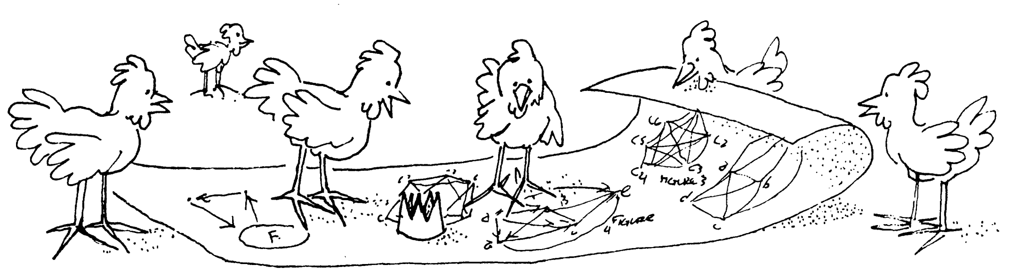
By a **flock** F of chickens we mean a nonempty (finite) set of chickens which has the property that for any two chickens in it, exactly one pecks the other. Formally, we should write $F=(C,P)$, that is, a flock is not just the set C of chickens, but a binary pecking relation P as well, where $(c,d)\in P$ iff c pecks d. However, it is more natural, and should not be confusing, if we write "chicken c is in F," rather than "chicken c is in the chicken set C of flock F." Likewise, we often write $c\in F$ rather than $c\in C$.

We will sometimes want to refer to P alone, so let us call it the **pecking order**. Also, let us call any individual ordered pair in P a **pecking pair**. Sociologists use the term "peck right," but in what I have read it seems to mean pecking order or pecking pair or both, depending on context.

In general, chickens will be denoted by c,d,a,b or sometimes $c_1,c_2,\ldots,c_n$. Flocks will be F and G. Usually there will be exactly one flock under consideration, F, and all chickens mentioned will be understood to belong to F.

It is helpful to think of a flock as a **directed graph**. A directed graph consists of a set of vertices and a set of directed edges between various pairs of vertices. In our case the chickens become the vertices, and there is a directed edge from vertex c to vertex d if and only if chicken c pecks chicken d. Clearly, any binary relation can be represented by a directed graph in this way. The graph of a flock has an additional property: between every pair of vertices there is exactly one directed edge. Such a directed graph is often called **complete**. Flocks of chickens are not the only objects modeled by such graphs. Suffice it to say for now that the usual name in the literature for such a complete directed graph is "tournament." The theorems we will prove are good for all complete directed graphs, and hence for all the situations they model. Perhaps then we should use standard graph-theoretic terms like "graph" and "vertex." However, I have decided to stick to chicken-theoretic terms like "flock" and "chicken." It helps keep firmly in mind the origins of our study, and it's more fun.

Now, how shall we define king? Surely, if chicken c pecks all other chickens in its flock, it should be called a king, but such a c does not always exist. For instance, suppose there are n



chickens and they peck each other cyclically; that is c_1 pecks $c_2, \ldots, c_{n-1}$ pecks c_n, and c_n pecks c_1. Then no matter what the other pecking pairs are, every chicken is pecked by at least one other. So this first suggestion for a definition is too narrow. Yet, if a chicken is to be called a king, surely it should dominate every other chicken, at least indirectly, and not too indirectly. Landau's idea was to allow domination in two steps as well as directly in one. Landau did not introduce a term for his concept, but we will.

DEFINITION. Chicken c in flock F is a **king** if for every other chicken d in F, either c pecks d, or there is some third chicken b such that c pecks b and b pecks d.

Since we will eventually find that this king concept is not a good one, it is important to say a bit more about what Landau did with it. His main concern [**3**, **4**] was to measure how hierarchical a pecking order will be, based on various assumptions about how the individual pecking pairs are determined. He introduced the idea of domination in one or two steps, almost as an afterthought, as the last topic in a third paper [**5**] dealing with some related but more purely mathematical ideas. By introducing this 2-step concept and proving Theorem 1 below, Landau surely intimated that the concept might be useful for studying levels of dominance. However, he never explicitly discussed the problem of designating a most dominant chicken. Moreover, he did remark that his 2-step concept need not pick out a unique chicken. Thus Landau, who is dead, might well have objected to our use of the work "king" for his concept. So the negative results below about kings can in no way be regarded as a criticism of Landau's work.

Existence Theorems

The miraculous thing about Landau's definition is that there is always such a king, and that the proof is easy. First, a few further conventions: For $c \in F$, let S_c be the set of chickens c pecks (Submissive to c) and let D_c be the set which peck c (Dominates c). Clearly, S_c, D_c, and $\{c\}$ partition F. The partition is illustrated in FIGURE 1. In this figure and henceforth, a "balloon"

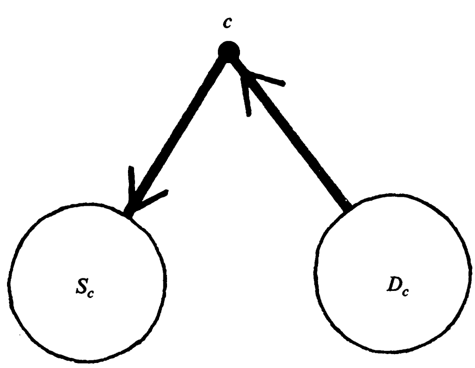


FIGURE 1.

represents a set of vertices, and a thick arrow between a single vertex and a balloon indicates that all edges between the single vertex and those in the balloon are directed like the thick arrow.

THEOREM 1 (Landau [**5**], p. 148). *Every flock of chickens has a king.*

Proof. Let $s(c)$ be the number of chickens c pecks. Let c be a chicken for which $s(c)$ is maximum. We will prove that c is a king. Suppose not. Then some chicken $d \in D_c$ is not dominated in two steps by c. That is, no chicken in S_c pecks d. Thus d pecks all chickens in S_c. But d also pecks c. So $s(d) > s(c)$, a contradiction.

COROLLARY 2. *Let s be the maximum number of chickens pecked by any chicken in F. Then every chicken in F who pecks s others is a king.*

This Corollary answers the question, "Can there be more than one king?" If we have a flock of 3 chickens which peck cyclically, then $s(c)=1$ for each chicken c. Thus all 3 are kings. Perhaps it is not unreasonable for all 3 to be called kings in this case; the flock is small and symmetric. In a large flock, though, we would want it to be unlikely, if not impossible, for all chickens to be kings. Thus we have not really answered the question about the existence of multiple kings until we have comprehensive results about how many kings there can be and how likely it is that there be that many.

Incidentally, for c to be a king it is not necessary that $s(c)$ be maximum. Can you give an example? If there are n chickens, can you determine how small $s(c)$ can be for a king?

The following lemma enables us to prove the next several results more easily. First note that every nonempty subset of a flock is actually a **subflock**; that is, the subset and its pecking pairs form a flock themselves. By picking a certain subflock and applying Theorem 1, we gain useful knowledge about the whole flock.

LEMMA 3. *In a flock, every chicken who is pecked is pecked by a king.*

Proof. Suppose $c \in F$ is pecked, that is, $D_c \neq \varnothing$. Since D_c is a flock, by Theorem 1 it contains a chicken d who is a king of D_c. We will show that d is actually a king of F. It suffices to show that d dominates every chicken in $F - D_c$ in one or two steps. This is true because $F - D_c = S_c \cup \{c\}$, d pecks c, and c pecks all chickens in S_c.

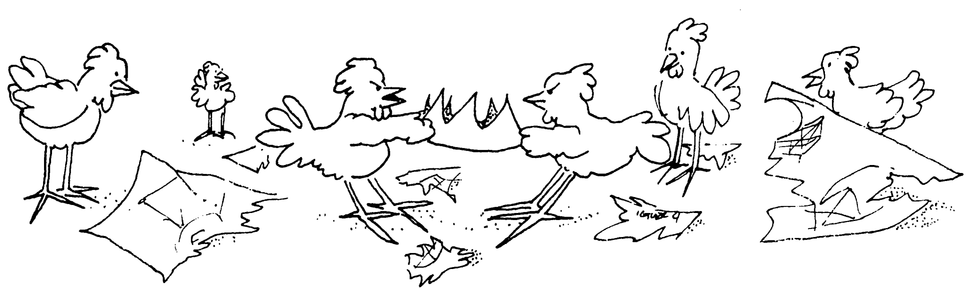
DEFINITION. Chicken c is an **emperor** of flock F if c pecks every other chicken in F.

THEOREM 4. *A flock has exactly one king if and only if that king is an emperor.*

Proof. "If" is clearly true (Why?). As for "Only if," suppose c were the only king but not an emperor. Then c is pecked. Hence F has another king by Lemma 3.

THEOREM 5. *No flock has exactly two kings.*

Proof. Suppose F has exactly two kings, c and d. Since d is a king, d dominates c in one or two steps. In particular, c is pecked. Thus c is pecked by a king. This must be d. Repeating the proof so far with d and c switched, we conclude that d is pecked by c. But, by definition of a flock, no pair can peck each other.



What Theorem 4 tells us is disappointing. Ideally, we would like a definition of king which implies that the king is always unique. However, we have already seen that Landau's definition does not assure uniqueness. Thus our next hope is that the definition assures uniqueness often, or at least in many cases where it isn't intuitively obvious that there's a best choice. Alas, what Theorem 4 tells us is that the only case in which Landau's definition gives a unique king is a case where it is already obvious.

Theorem 5, on the other hand, is promising. What is perhaps the least workable form of oligopoly, duopoly, can never occur. Perhaps there are lots of other numbers of kings which cannot occur. At the least, Theorem 5 prompts us to attack this question more fully. How shall we make this question more precise? Here is one way.

DEFINITION. Let n and k be integers, $n \geqslant k \geqslant 1$. We say that F is an **(n,k) flock** if F has n chickens and exactly k of them are kings. If F has n chickens, it is an **n-flock**.

The question now becomes, "For which n and k do there exist (n,k) flocks?" We will answer this completely. We begin by answering a special case: "For what n are there (n,n) flocks?" That is, when can every chicken be a king? The answer is surprising—at least it was to me.

THEOREM 6. *For every positive integer n except 2 and 4, there exists an n-flock in which every chicken is a king.*

Our proof of this follows by induction from the following Lemmas.

LEMMA 7. *If there exists an (n,n) flock, then there exists an $(n+2,n+2)$ flock.*

LEMMA 8. *There is a $(1,1)$ flock.*

LEMMA 9. *There does not exist a $(4,4)$ flock.*

LEMMA 10. *There is a $(6,6)$ flock.*

The proof of Theorem 6 from the lemmas is by two separate inductions. By Lemmas 7 and 8, one proves the existence of an (n,n) flock for all odd positive integers. By Lemmas 7 and 10 one proves the existence of an (n,n) flock for all even integers $\geqslant 6$. Finally, (4,4) flocks are excluded by Lemma 9, hence (2,2) flocks are excluded by Lemma 7.

Proof of Lemma 7. This beautiful argument is contained in FIGURE 2. Suppose F is an (n,n) flock. Create two new chickens c and d, and let their pecking relations to each other and to the

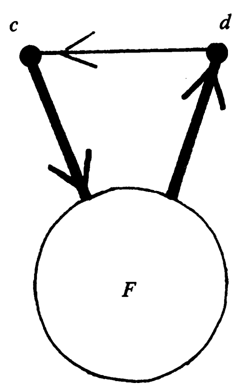


FIGURE 2.

chickens of F be defined as in the figure. Call the augmented flock G. Then c is a king of G (Why?). Also d is a king of G (Why?). Finally, we already know that each $b \in F$ is a king of F. Hence b is a king of G (Why?).

Proof of Lemma 8. This is true vacuously. That is, in any 1-flock the one chicken pecks every other chicken because there are no other chickens.

As for Lemma 9, there are only 64 ways to assign dominance between 4 chickens 2 at a time (Why?). By symmetry the number can be much reduced. In any event, it is clear that the truth value of Lemma 9 can be determined by brute force examination of cases. However, it is esthetic, as well as less tedious, to give a more incisive, conceptual proof. I offer the following sketch. If F is to be (4,4), no chicken either pecks or is pecked by all the others. Since there are 6 pecking pairs, some two chickens, call them a and b, peck exactly two chickens each, while the remaining two chickens peck exactly one each. One of the first two, say a, uses one of his pecks on the other. These facts uniquely determine F (draw the graph!) and it is not (4,4).

Proof of Lemma 10. To prove this, we need merely display one (6,6) flock. Checking whether a proffered 6-flock works is routine. To devise candidate flocks, it is very helpful to think in terms of graphs. Since all 6 chickens are to be kings, it is reasonable to arrange the arrows as symmetrically as possible. Also, for each c, the chickens that c pecks should be "spread out," that is, they shouldn't "waste" their pecks on each other since c must peck all other chickens through them. FIGURE 3 works.

This completes the proof of Theorem 6.

THEOREM 11. *There exist (n,k) flocks for all integers $n \geqslant k \geqslant 1$, with the following exceptions*: $k=2$ *with n arbitrary, and* $n=k=4$.

Proof. We already know that $k=2$ with n arbitrary, and (4,4) are excluded. We must show that nothing else is. The case $k \neq 2,4$ is easy. Start with a (k,k) flock and create $n-k$ new chickens, each of which is pecked by all k old chickens. Arrange the pecking order among the $n-k$ new chickens arbitrarily. Call the augmented flock G. It is easy to see that all the old chickens are still kings in G and that no new chickens are kings in G. Thus G is (n,k).

For $k=4$, it suffices to show that a (5,4) flock exists, for then an $(n,4)$ flock with $n>5$ can be created by adding $n-5$ new subservient chickens to the (5,4) flock just as above. A (5,4) flock is shown in FIGURE 4.

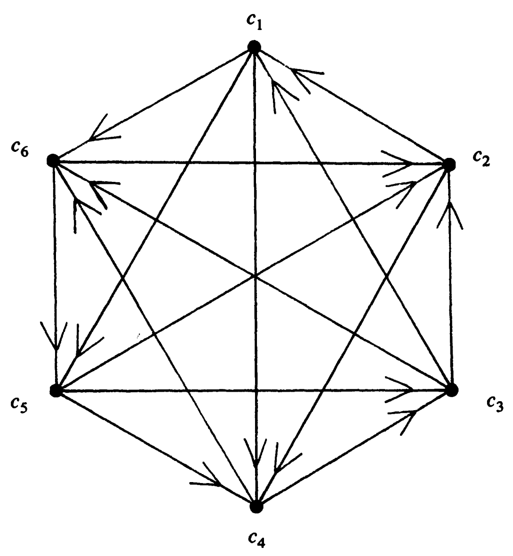


FIGURE 3.

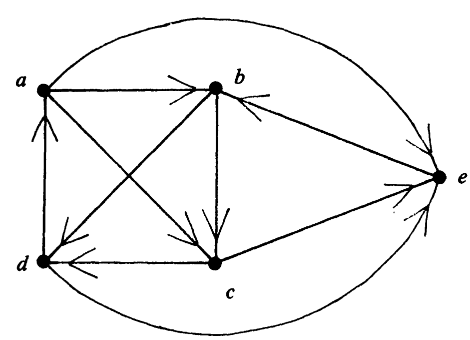


FIGURE 4.

Probabilistic Theorems

The results above about the existence of (n,k) flocks may seem to be evidence enough that our definition of king is not a good one. However, this conclusion is premature. It may be that (n,k) flocks for k large are extremely rare. It may also be that emperors are extremely common. If either of these things is true, the definition would be, in practice, a useful way to determine power in a flock.

How shall we assign probabilities to the various pecking orders a flock can have? With no prior information to go on, it seems reasonable to assign equal probability to each possible pecking pair. That is, for every unordered pair $\{c,d\}$, assign probability $1/2$ to the event that c

pecks d, and $1/2$ to the event that d pecks c. Since an n-flock has $\binom{n}{2}$ unordered pairs, there are $2^{\binom{n}{2}}$ equally likely pecking orders. In what follows we use this probability assumption, and we refer to an n-flock chosen using this assumption as a **random n-flock**. We also use basic rules of probability. In particular, let $\Pr(E)$ be the probability that event E occurs. Let $\{E_i\}$ be a set of events, and let $\cup E_i$ be the event that at least one of the E_i occurs. Then

$$\Pr(\cup E_i) \leqslant \sum_i \Pr(E_i), \tag{1}$$

with equality iff the E_i are mutually exclusive.

THEOREM 12. *The probability that a random n-flock has an emperor is* $n(1/2)^{n-1}$.

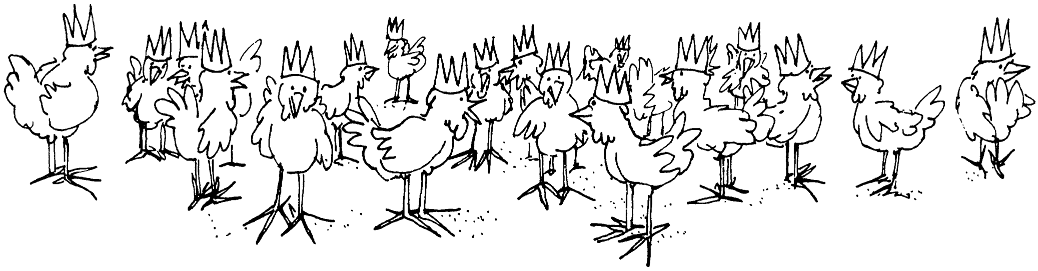
Proof. Let E_i be the event that chicken c_i is emperor. Since no two chickens can be emperors simultaneously, and since by symmetry $\Pr(E_i)=\Pr(E_j)$ for all i and j, the probability we seek is $n\Pr(E_1)$. Moreover, c_1 is emperor iff he pecks all $n-1$ others. Hence $\Pr(E_1)=(1/2)^{n-1}$.

COROLLARY 13. *The probability that a random n-flock has an emperor approaches* 0 *as* $n\to\infty$.

Proof. We need one fact from analysis: Exponential functions grow much faster than polynomials. Precisely, if $P(x)$ is any polynomial, and $a>1$, then

$$\text{as } x\to\infty,\ P(x)/a^x \to 0. \tag{2}$$

One way to prove this is by repeated use of L'Hospital's Rule. Since the probability $p_E(n)$ that a random n-flock has an emperor is $2n/2^n$ (by Theorem 12), it must approach 0 as $n\to\infty$. (In fact, $p_E(n)$ goes to 0 very fast. For $n \geqslant 12$, $p_E(n) < .01$.)



That it is very rare for only one chicken to be king did not surprise me, but I was shocked to learn that the opposite extreme—every chicken a king—is very common. I would never have guessed such a thing, let alone proved it, had not a student reported to my class the results of a computer program he ran which counted the number of kings in 100 random flocks of 16 chickens. No flock had fewer than 8 kings, and most had 14, 15, or 16!

THEOREM 14. *The probability that every chicken in a random n-flock is a king approaches* 1 *as* $n\to\infty$.

Proof. Let $p(n)$ denote this probability. I do not know an exact expression for $p(n)$, so the proof proceeds by estimates. Let $\bar{p}(n)=1-p(n)$ be the probability that *not* every chicken is a king. I will overestimate $\bar{p}(n)$ with a certain computable quantity $q(n)$ and will then show $q(n)\to 0$ as $n\to\infty$. Thus $\bar{p}(n)\to 0$, so $p(n)\to 1$.

Not every chicken is a king if and only if there exists an ordered pair (c,d) such that c does not peck d and no chicken which c does peck pecks d. Let $p(c,d)$ be the probability that the particular pair (c,d) has this non-dominating property. By symmetry all the $p(c,d)$ are equal. Since there are $n(n-1)$ such ordered pairs, $\bar{p}(n) \leqslant n(n-1)p(c,d)$ by (1). This last product is the $q(n)$ referred to above. It is easy to compute because $p(c,d)$ is easy to compute. The probability that c does not peck d is $1/2$. The probability that, for a specific third chicken b, c does not dominate d through b is $3/4$ (Why?). Now, who pecks whom between c and d is independent of their mutual relation to any one b, and their mutual relation to one b is independent of their relation to any other b. Thus $p(c,d)=(1/2)(3/4)^{n-2}$. So

$$q(n)=n(n-1)(1/2)(3/4)^{n-2}=\frac{(8/9)n(n-1)}{(4/3)^n}.$$

By (2), $q(n)\to 0$. (In fact, $q(n)\leqslant .01$ when $n\geqslant 42$. Thus for these same n (and no doubt some lower ones) $p(n)\geqslant .99$.)

COROLLARY 15. *Let $k(n)$ be the average number of kings in a random n-flock. As $n\to\infty$, $k(n)-n\to 0$.*

Proof. Let $p_K(n,i)$ be the probability that a random flock of n chickens has exactly i kings. By definition, $k(n)=\sum_i ip_K(n,i)$. Thus, using the notation $\bar{p}(n)$ as in the proof of Theorem 14,

$$\begin{aligned} n\geqslant k(n)&\geqslant np_K(n,n)=n(1-\bar{p}(n))\\ &\geqslant n\left[1-\frac{(8/9)n(n-1)}{(4/3)^n}\right]\\ &=n-\frac{(8/9)n^2(n-1)}{(4/3)^n}. \end{aligned}$$

By (2), the latter term goes to 0. Thus $k(n)-n\to 0$.

The results of this section are much more damaging to our king concept than the previous existence results. Yet, as we point out later, there may still be hope for the definition in terms of its utility for applications to real-world flocks. For now, let us be content that the definition has led to some surprising theorems.

Submissive Chickens and Duality

DEFINITION. Chicken $c\in F$ is a **slave** if c is pecked by all other chickens in F. Chicken c is a **serf** if every other chicken either pecks c or pecks another chicken who pecks c.

Since the definitions of serf and slave are just those of king and emperor "turned around," it seems clear that all the results about king and emperor go through for serf and slave. The only question is how to justify this without repeating everything.

What we have here is a simple example of **duality**. Duality is a correspondence between concepts so that for every true statement in one's theory, if one substitutes corresponding concepts throughout, one gets another true statement. For instance, in our theory the correspondence includes: pecks↔is pecked by; king↔serf; emperor↔slave; chicken↔chicken; and flock↔flock. Using this duality (justified in a moment), one gets from Theorem 1, for instance, that every flock of chickens has a serf, and from Lemma 2, that every chicken who pecks pecks a serf.

To justify this duality, note that the fundamental concept in our theory is the pecking order P. Everything else has been defined in terms of this binary relation. For instance a flock is a set C such that for every $c,d\in C$, exactly one of the ordered pairs (c,d), (d,c) is in P.

Thus we see that every theorem in our theory is really of the form "If P satisfies hypotheses A, then it also satisfies conclusions B." (Sometimes the only hypothesis is the tacit one that P is the relation of some flock.) Now, the crux is: the name P we have used for the relation is irrelevant. We could substitute any other name P^* in the theorem and the proof would still be valid. Of course any defined terms used in the (usual) statement of the theorem and its proof would have to be replaced by terms defined identically, but using P^* instead of P.

In particular, we could replace $P=$ "pecks" with $P^*=$ "is pecked by." Making this substitution in the definitions of king and emperor clearly gives serf and slave. Since a chicken is an element in the set on which "pecks" is defined, after the substitution a chicken is still a chicken. Also a flock is still a flock, for there is no difference in meaning between saying "for every pair, exactly one pecks the other," and "for every pair exactly one is pecked by the other." Thus the claimed duality is justified.

In addition to giving a formal boost to one's theory by providing an analogous theorem for every theorem already proved, duality also gives a psychological boost by suggesting situations in which to look for new types of theorems. Specifically, it gets us to ask about situations in which concepts and their dual concepts interact. For instance, since we know about the probability that every chicken is a king, and dually, about the probability that every chicken is a serf, it occurred to me to ask about the probability that every chicken is both. The answer is another theorem.

THEOREM 16. *Let $p_{KS}(n)$ be the probability that, in a random n-flock, every chicken is both a king and a serf. As $n \to \infty$, $p_{KS}(n) \to 1$.*

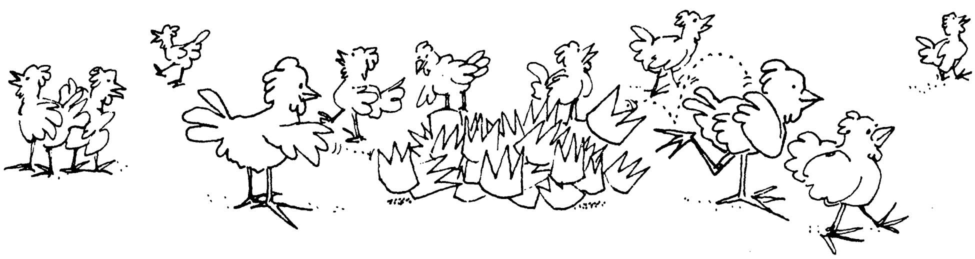
Proof. Let $p_K(n)$ be the probability that all n chickens are kings, and let $p_S(n)$ be the probability that all n are serfs. By Theorem 14, $p_K(n) \to 1$. By its dual, $p_S(n) \to 1$. Now, if $E \cap E'$ is the event that events E and E' both occur, then

$$\Pr(E \cap E') \geqslant \Pr(E) + \Pr(E') - 1. \quad (3)$$

(This is shown by applying (1) to the events not-E and not-E'.) So let E be "all n chickens are kings" and E' be "all n chickens are serfs." Then (3) becomes $p_{KS}(n) \geqslant p_K(n) + p_S(n) - 1$. Since $p_{KS}(n) \leqslant 1$ in any event, the theorem follows.

We ask the reader to prove the next theorem, and to use it to give a much simpler proof of Theorem 16.

THEOREM 17. *If every chicken of F is a king, then every chicken of F is a serf.*



Redefining Kings

Clearly we should consider other definitions of king. I begin with two alternative definitions suggested by students who have heard me lecture on this topic.

According to the current definition, a king dominates every other chicken in one or two steps. Why not allow 3 steps, or 4? To be precise, let us say c dominates d **in s steps** if there is a sequence of chickens $c = c_0, c_1, \ldots, c_{s-1}, c_s = d$ such that for all $i = 0, 1, \ldots, s-1$, c_i pecks c_{i+1}. Then c is an **s-king** if c dominates every other chicken in F in at most s steps. In particular, a 1-king is an emperor and a 2-king is a king.

We may now ask: do 3-kings always exist? For which (n, k) do there exist n-flocks with k 3-kings? How often is every chicken a 3-king? Is the definition of 3-king more satisfactory than that of king? What about 4-kings? Etc. Actually, most of these questions are very easy to answer, either by mimicking the proofs of the theorems for kings, or, often enough, by simply using the statements of those theorems. We leave all this to the reader. The upshot is: s-kings are not more satisfactory.

Another approach suggested by students hinges on the idea that, though a flock rarely has just one king, often the set of all kings is at least a proper subset of the flock. Thus, given flock

$F = F_1$, let F_2 be the subflock of all kings of F_1. In general, for $i \geqslant 1$, let F_{i+1} be the subflock of all kings of flock F_i. Then $F_1, F_2, \ldots$ is a nonincreasing sequence of subsets of F_1. Since F_1 is finite, there must be a first i for which $F_{i+1} = F_i$; thereafter the sequence is constant. Let us call any c in this F_i a **king-of-kings**.

The idea is that the number of kings-of-kings should be much less than the number of kings. Alas, Theorem 14 tells us that it isn't much less. As soon as n is large enough, most of the sequences $F_1, F_2, \ldots$ are constant right from the start. Almost every chicken is a king-of-kings! It also turns out that most of the existential questions we have asked about kings can be answered easily for kings-of-kings too: review the proofs for kings. One intriguing question may appear more difficult: characterize those flocks which have a unique king-of-kings. (A hint on how to answer this question is given later.)

We now turn to definitions of king suggested by studies in the literature of other situations modeled by complete directed graphs. We have already mentioned that such graphs are usually called tournaments. Consider a round-robin tournament among n teams $t_1, \ldots, t_n$. Round-robin means that each pair $\{t_i, t_j\}$ plays exactly once and there are no ties. If we create a vertex i for each t_i, and put an edge from i to j iff t_i beats t_j, then this construction gives us a complete directed graph for each round-robin tournament. Conversely, every complete directed graph represents a possible round-robin tournament.

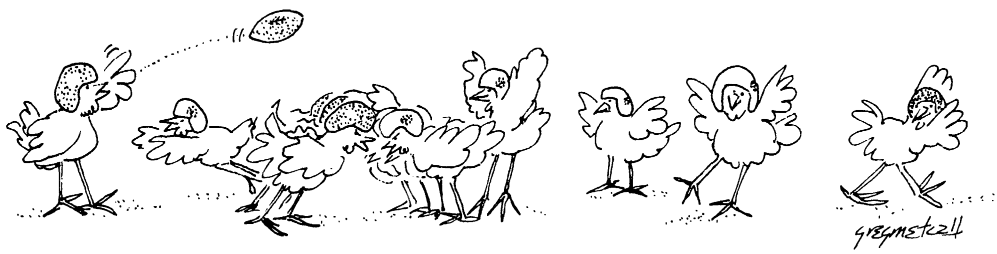
In a real-world tournament, often one wants not only to determine the best team but also to rank them all in descending order. Surely, the basic idea of best team is akin to that of king chicken, so any method for ranking teams ought to have inherent in it a plausible method for choosing a king.

A concept used in many proposed ranking methods is the **score vector**. Let T be a tournament with n teams. Let s_i be the number of victories of team t_i in T. Renumber the teams so that $s_1 \geqslant s_2 \geqslant \cdots \geqslant s_n$. Then the score vector is defined to be $(s_1, \ldots, s_n)$. It seems natural to rank the teams in this order, and to declare team t_1 to be the champion. By analogy, we could redefine king to be a chicken who pecks the most others. Indeed, by Corollary 2, this new definition is at least as selective as the original. (What is called s in Corollary 2 is what we now call s_1.)

However, there may be ties in score vectors. If a tie occurs for top scorer, then this ranking method does not pick a unique king either. In fact, it is possible for *every* team to have the same score. It is not hard to show that this is possible if and only if n is odd.

Another objection is that a team which wins more games may not really be better. What if some t_i wins fewer games than t_1 but wins them all against tougher teams than t_1 wins against? We might be inclined to rate t_i higher than t_1. Only, how do we judge which teams are tough unless we already have a ranking? Here is one way which has been proposed. We make an initial ranking of the teams based on their score vectors as above. Now, for each team we compute a revised score vector by summing the scores of all the teams it beat. We now have a revised score vector $(s_1', \ldots, s_n')$, where it may well be that we have to rerank the teams in order that the revised scores are in descending order. We can repeat the process: Compute a rerevised score for each team by summing the revised scores of all the teams it beat. The rerevised scores may require a further reordering. In theory, we could repeat this process endlessly. If the reader feels intuitively that, after a while, little new information is gained by revising the current score vector again, he is right. It is a theorem that after a finite number of revisions, no further reranking will be caused by further revision. The proof of this uses the theory of eigenvalues of matrices! For precise statements and proofs, see Moon [8] Section 15. Suffice it to say here that this method of choosing a champion (or a king) is not free from objections either.

Another definition of champion is suggested by the following theorem: In every tournament there is a directed path which passes through each vertex once. In terms of chickens in an n-flock, this means it is always possible to number the chickens so that c_1 pecks c_2, c_2 pecks $c_3, \ldots$, and c_{n-1} pecks c_n. We could redefine a king to be the head of such a chain of command. Unfortunately, this path is rarely unique; in fact, it is unique if and only if the pecking order is



linear. Moreover, as $n \to \infty$, the probability that in a random n-flock every chicken is a king in this sense goes to 1 too. For more details about this approach, see Roberts [**14**], Section 3.2. (Roberts does not give a proof of the last, probabilistic assertion. It can be proved using the fact that each so-called strong tournament contains a cycle which passes through all the vertices and the fact that most tournaments are strong. See Moon [**8**], Theorems 1 and 3.)

Many other methods have been proposed for ranking tournaments, none of them entirely satisfactory. Perhaps this explains why most leagues have special play-offs to determine the champion.

Other Models

Our model produces too many kings, yet changes in the definition of king don't seem to help. So maybe the real problem is with our assumptions about pecking orders rather than with our definitions. To explain this suggestion without undue complications, let us do so within the context of Landau's original definition. We have shown that this definition is unreasonable if we allow a flock to have any pecking order, or even if we assume each pecking order on n chickens is equally likely. But these assumptions were based, to put it politely, on mathematical esthetics and convenience, not on knowledge of chickens. Landau himself was a theoretician of chickens rather than an observer of chickens. But he had read the observational literature. As he notes, although the pecking order in real-world flocks is rarely linear, it is also rarely very far from linear. Thus, for real-world flocks, it may be that his definition works much better than we so far have led ourselves to suspect.

In short, the real problem may be that our assumptions about pecking are too broad. Maybe we should restrict flocks to tournaments which are "almost linear." Maybe we need to use a probability distribution which is very far from random.

Let us pursue the first of these proposed alterations just a bit further. Before we can prove anything about kings in "almost linear" flocks, we need to have a precise mathematical definition of that concept. One way to define it is suggested by the following facts. Proofs may be found in Moon [**8**], Sections 5-7, but the reader might try to prove them himself. (1) A tournament is linear iff it contains no 3-cycles; (2) the number of 3-cycles in a tournament depends only on the score vector; (3) in a random n-tournament, the expected number of 3-cycles is $(1/4)\binom{n}{3}$. Thus the number of 3-cycles is a measure of how linear a tournament is. According to the observational literature Landau quotes, a typical flock of 10 to 20 chickens contains only 2 or 3 3-cycles. Thus a good topic for further study is this: Are there bounds for the number of kings in terms of the score vector or the number of 3-cycles? If the number of 3-cycles is severely bounded, is the number and nature of the kings severely restricted? If not, what does the number of kings depend on? For more information about the number of 3-cycles as a measure of linearity, see Harary, Norman and Cartwright [**2**], pp. 300–304. Landau himself proposed a different measure of linearity, which he called the hierarchy index [**3, 4**].

A word of caution: if the goal is to designate a unique king, then no amount of restriction on pecking orders, either absolute or probabilistic, will cause Landau's definition to be the right one. This is because of Theorem 4. In our present context, what it says is this: Any restriction which leads to Landau kings' being unique is a restriction which makes the concept unnecessary; use the simpler concept of emperor instead.

However, it may well be reasonable to allow a few kings. Then Landau's definition would still be in the running, especially if one can find mathematical restrictions on pecking orders which result in few kings in his sense and which correspond well to the empirical facts about flocks. On the other hand, it may be that the combination of some restrictions on pecking orders and some alternative definition of king will turn out to be best. In any event, this is not something a mathematician can decide by pure cogitation. A thorough knowledge of the facts about chickens is necessary. To pursue this seriously, one might begin by reading the many articles on chickens in the anthology on social hierarchy **[15]**.

Dominance structures occur in many human and animal societies. They were first studied among chickens, where they are especially apparent, quite stable, very hierarchical, and involve every pair; but now many other instances have been documented (see **[15]** or Chapter 13 of **[18]**). The point here is that there is considerable variety among species as to the nature of the dominance structure. Mathematical restrictions which are right for chickens may not be right for others. There is room here for a lot of mathematical modeling. Also, a lot of reading of empirical studies and a lot of interaction with social scientists will be necessary if this modeling is to be accurate and useful.

Further Problems

The previous section presents one big open problem. One reason the problem is big is because it is vague; a lot of the work would be in determining appropriate models. Proving theorems would only come later, and might forever be subsidiary, say, to running and evaluating computer simulations.

In contrast, the problems in this section concern our original model and definitions. Thus they are precise and mathematical, but more narrow. Because of the unrealistic implications of this model for kings, these problems are better regarded as pure mathematics than as applied. When I first drafted this article, these problems were all open. Problems 1, 2 and 4 have since been solved, and no doubt more solutions will be found by the time this article appears in print. However, this should not discourage anyone from working on any of them. Different solutions often involve different techniques, are more or less difficult or complete, answer or suggest different generalizations, etc. Problem 4 was solved by an undergraduate. The solution of Problem 1 is announced in **[13]**.

PROBLEM 1. For what 4-tuples (n,k,s,k_s) does there exist an n-flock with exactly k kings and s serfs and such that exactly k_s of the kings are also serfs?

PROBLEM 2. What are the possible structures for subflocks of kings? That is, for which n-flocks F does there exist an m-flock G such that G contains n kings and F is the subflock of kings of G?

It is not true that every flock F is the king subflock of some G. A necessary condition on F is implicit in Lemma 3. Problem 2 asks for necessary and sufficient conditions. Incidentally, Lemma 3 can be used to answer a question asked in the previous section: When is there a unique king-of-kings?

PROBLEM 3. When do the various definitions of king agree? For instance, when is the set of kings-of-kings the same as the set of kings? When is the set of chickens with the highest score the same as the set of kings? Etc.

PROBLEM 4. Find a proof of Lemma 7 using induction from n to $n+1$ instead of from n to $n+2$. Such an induction can only work starting at $n=5$ (Why?) so it cannot be entirely straightforward.

PROBLEM 5. In a random n-flock, what is the probability that exactly k chickens are kings? That at least k chickens are kings? That at least a particular set S of chickens are kings?

Implicit in our theorems are answers to the "exactly" question for $k=0,1,2$ and to the at-least question for $k=0,1,2,3$. The answer to the "set" question is known when S is a singleton. Answers for a few more cases ought to be obtainable by careful brute force, but matters rapidly get complicated: kingships are not independent events. The real question is: are there formulas for general k and S? Fortunately, if any one of the three questions can be answered completely, the other two follow by the general combinatorical principle of Inclusion-Exclusion. (For discussion of this principle, see any probability or combinatorics book.)

If exact formulas can't be found, what about bounds and asymptotics? We already have an asymptotic result for $k=n$, namely that the probability goes to 1. (Note that the "exactly" and the "at-least" problems are the same for $k=n$.) Thus we automatically have asymptotic results for all $k<n$, namely that the probability goes to 0 for the exactly problem. However, it might be interesting to determine how fast it goes to 0.

PROBLEM 6. In real-world flocks, pecking orders are not permanent. Occasionally the order between two chickens switches. Call one such change a **single switch**. Landau's second paper **[4]** was devoted to the question of how repeated single switches affect the nature of the pecking order over time, under various assumptions about the nature of those single switches. Landau did not, however, analyze what single switches do to the set of kings. Problem 6 is to analyze this. Specifically, suppose F and G are flocks on the same set of chickens. Let their number of kings be k_F and k_G, where we may assume that $k_F \leqslant k_G$. There are many sequences of single switches which turn F into G. Can a sequence be found in which the number of kings always stays the same or increases? If S is the set of chickens who are kings in both F and G, can a sequence be found so that at each state every chicken in S is still a king? These are examples of questions about "minimally disruptive" sequences of single switches from F to G. Still other questions can be made up if one does not specify G. For instance, for every flock F is there a single switch which does not change the set of kings, or at least does not change the number of kings? Does every flock have a single switch which changes the number of kings by at most one? Does every emperorless flock have a single switch which changes the number of kings by exactly one?

Credits

The idea of studying chicken flocks mathematically seems to have originated with Rapoport **[10, 11, 12]**, but the first significant general results are Landau's. Landau succeeded in analyzing the expected value of his "hierarchy index" for arbitrary n-flocks under several different probabilistic assumptions about pecking pairs. As an aside, he proved Theorem 1 and Corollary 2. Landau attributes independent and earlier proofs of Theorem 1 to F. E. Hohn and H. E. Vaughan.

Theorems 4 and 5 appear implicitly in a problem posed in this MAGAZINE by D. L. Silverman **[17]** and solved there by Moon **[9]**. The context there is that of debtors in a club instead of henpecked chickens. Lemma 9 appears as half of problem 5 on page 317 of **[2]**. Theorem 14 is essentially a restatement of part of a result of Moon and Moser, as given by Moon **[8]**, pages 32-34.

It surprises me that more has not been done with Landau's ideas in the 25 years since he wrote **[3, 4, 5]**. Dominance is an important concept in biology, sociology, and psychology. One wants to know why various groups of living things develop dominance structures, why one particular living thing dominates another, and why dominance structures of whole groups tend to be hierarchical. Landau's hierarchy index is well-known among quantitatively minded social scientists; they compute it as a statistic in analyzing their data. However, little has been done in these 25 years to further develop a mathematical theory of hierarchy formation. Landau himself wrote another paper on this subject in 1965 **[6]**. He also wrote a survey of all his work on this subject, published posthumously in 1968 **[7]**. (I recommend this survey as a first paper to read on this subject.) It seems that the only substantial article to appear since on the mathematics of hierarchy formation is Chase [**1**].

What have mathematicians done with Landau's ideas? The King Chicken Theorem is only a secondary result in [**5**]. The main theorem is a characterization of those n-tuples which are the score vectors of tournaments. This is what has interested mathematicians most. This score vector characterization has been reproved several ways and generalized; see Moon [**8**], Sections 21 and 22.

Landau's king concept seems to have interested mathematicians only to the limited extent already mentioned, and to have interested social scientists not at all. If we limit ourselves to chickens, there are several possible explanations for this lack of interest among social scientists. It has already been noted that the pecking order in chicken flocks is so close to hierarchy that subtle methods of designating top chickens have not seemed necessary. Moreover, the observational literature indicates that even when there is an evident most dominant chicken, this chicken does not *lead* the flock. It merely exercises certain perquisites, e.g., it gets the best food or the best roost. Domestic flocks do not seem to have real leaders.

On the other hand, since the variety of dominance structures in human and animal societies is immense, it is possible that Landau's king concept, or some other mathematical king concept, could still prove useful to social scientists.

In closing, I would like to thank all the mathematicians and students, especially Tim Bock, who have kept me thinking about this subject.

Work on this paper was done while the author was in the Mathematics Department at Princeton University, supported by National Science Foundation Grant MCS 76-19670.

References

[1] I. D. Chase, Models of hierarchy formation in animal societies, Behavioral Sci., 19 (1974) 374–382.

[2] F. Harary, R. Norman and D. Cartwright, Structural Models: an Introduction to the Theory of Directed Graphs, Wiley, New York, 1965.

[3] H. G. Landau, On dominance relations and the structure of animal societies: I. effect of inherent characteristics, Bull. Math. Biophys., 13 (1951) 1–19.

[4] H. G. Landau, On dominance relations and the structure of animal societies: II. some effects of possible social factors, Bull. Math. Biophys., 13 (1951) 245–262.

[5] H. G. Landau, On dominance relations and the structure of animal societies: III. the condition for a score structure, Bull. Math. Biophys., 15 (1953) 143–148.

[6] H. G. Landau, Development of structure within a society with a dominance relation when new members are added successively, Bull. Math. Biophys., 27 (special issue) (1965) 151–160.

[7] H. G. Landau, Models of Social Structure, Bull. Math. Biophys., 30 (1968) 215–224.

[8] J. Moon, Topics on Tournaments, Holt, Rinehart and Winston, New York, 1968.

[9] J. Moon, Solution to problem 463, this Magazine, 34 (1961) 425.

[10] A. Rapoport, Outline of a probabilistic approach to animal sociology: I, Bull. Math. Biophys., 11 (1949) 183–196.

[11] A. Rapoport, Outline of a probabilistic approach to animal sociology: II, Bull. Math. Biophys., 11 (1949) 273–281.

[12] A. Rapoport, Outline of a probabilistic approach to animal sociology: III, Bull. Math. Biophys., 12 (1950) 7–17.

[13] K. B. Reid, Dominance in tournaments: kings and serfs, Proceedings of the Tenth Southeastern Conference on Combinatorics, Graph Theory, and Computing 1979, Utilitas Mathematica Publishing, to appear.

[14] F. S. Roberts, Discrete Mathematical Models, Prentice-Hall, Englewood Cliffs, 1976.

[15] M. W. Schein, ed., Social Hierarchy and Dominance, Dowden, Hutchinson and Ross, Stroudsburg, Pa., 1975.

[16] T. Schjelderup-Ebbe, Beiträge zur Sozialpsychologie des Haushuhns, Z. Psychol., 88 (1922) 225–252, reprinted with an English translation in reference [**15**] 7–49.

[17] D. L. Silverman, Problem 463, this Magazine, 34 (1961) 425.

[18] E. O. Wilson, Sociobiology: The New Synthesis, Belknap Press of Harvard University Press, Cambridge, Mass., 1975.

Advanced Plane Topology From an Elementary Standpoint

A simple combinatorial approach to subtle topological premises of basic undergraduate analysis.

DONALD E. SANDERSON
Iowa State University
Ames, IA 50010

The title of Felix Klein's classic attempt to smooth the discontinuities between secondary and university mathematics [7] is paraphrased here in the hope of conveying the spirit of this paper and not for adding any guilt by association. Klein's objective was to show how university mathematics could enrich and elucidate secondary material. Turning half-circle from this point of view, our goal here is to show how undergraduate mathematics can reveal classical results of advanced plane topology which impinge on undergraduate mathematics but are seldom fully treated until graduate school, if ever.

At least three examples come to mind. First, most mathematics majors are exposed to Green's Theorem, which equates an integral on a simple closed curve to an integral over the region it bounds. The statement itself implicitly involves the Jordan Curve Theorem which guarantees that such a curve indeed bounds a unique finite region (see [2] page 243). Secondly, in elementary calculus the Intermediate Value Theorem is often used to show that $f(x)=0$ has a solution between a and b if $f(a)<0<f(b)$ and f is continuous. Elementary coordinate changes show that this is equivalent to existence of a solution of an equation $g(x)=x$ (a "fixed point" of g) where g maps an interval into itself. This is the one-dimensional Brouwer Fixed Point Theorem, a special instance of more general fixed point theorems that are useful for showing the existence of solutions to differential equations, economic models, etc. Finally, the implicit function theorem, necessary for justifying implicit differentiation, entails some form of Brouwer's Invariance of Domain Theorem (see footnote on page 281 of [4]). Students need not be burdened by complete proofs of all such topological preliminaries, but for those interested they should be more accessible.

Perhaps the simplest complete treatment of the classical results discussed here is in M. H. A. Newman's book [8]. The very last section of Hocking and Young [6] gives a modern version of J. W. Alexander's combinatorial approach [1] to some of these topics. However, each of these sources involve at least a modicum of algebraic topology. The approach taken here uses none, but could be used to motivate an introduction to algebraic topology. Little more is used than the barest facts about continuous functions and their behavior with regard to closed, compact (closed bounded) and connected sets. Most references are for the record (or the very curious), not for suggested reading, as many are relatively inaccessible. No apology is offered as that situation is the motivation for this paper!

The author wishes to acknowledge a considerable debt to Albrecht Dold for his excellent paper [5] which, somewhat in the spirit of Klein, was presented at a conference fostering interaction between universities and "Hochschulen" in Germany. Many of the basic ideas in what follows originated there and inspired this attempt to publicize and expand on his results.

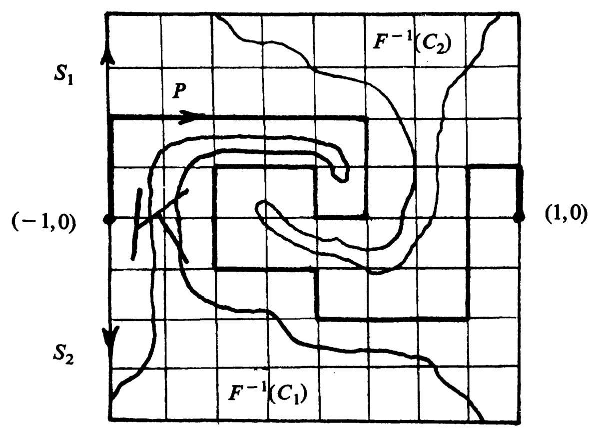
A Fundamental Result

A great deal of important topology of n-dimension Euclidean space E^n can be derived with surprising ease from what is sometimes called Alexander's Addition Theorem. A special case, sufficient for the topology of E^2, can be established by a simple parity (even-odd) argument similar to that used by Euler to solve the Königsberg bridge problem. For simplicity we establish some standard terminology and notation.

We call a continuous function a **map**. If $f: I \to X$ is a map of an interval I with ends a, b, then the image, $f(I)$, is called a **path** in X from $f(a)$ to $f(b)$. Observe that the continuous image of a path is again a path. A map f is called an **embedding** or **homeomorphism** onto its image if it is $1-1$ (i.e., its **inverse** f^{-1} exists) and f^{-1} is a map. Let R be the square region in E^2 with corners $(\pm 1, \pm 1)$. The region R has a square boundary $S = S_1 \cup S_2$ where S_1 is the upper and S_2 the lower half, each being a path joining $(-1,0)$ to $(1,0)$. We can now state and prove our key theorem which says, roughly, that if a path from x to y missing one closed set can be continuously deformed into one missing a second closed set without ever touching a point in both sets, then some path from x to y misses both sets.

Alexander Addition Theorem. *For $i = 1,2$, suppose C_i is closed in a space X, P_i is a path in $X - C_i$ from x to y and $F: R \to X - (C_1 \cap C_2)$ is a map with $F(S_i) = P_i$. Then some path $F(P)$ in $X - (C_1 \cup C_2)$ joins x to y.*

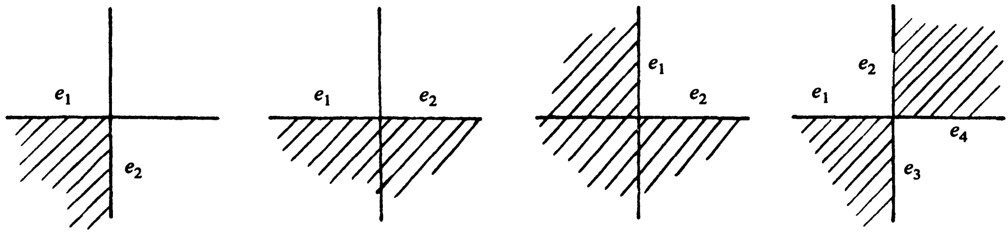
Proof. A picture of the square domain R of F appears in Figure 1; the map F is the continuous deformation referred to in the description of the theorem. Note first that $F^{-1}(C_2)$ has no points in common with $F^{-1}(C_1)$ or S_2. Since these sets are closed, R can be divided into



Pick a path P so $F(P)$ misses $C_1 \cup C_2$.

Figure 1.

finitely many squares by horizontal and vertical lines (including the axes) so no square of R meets both $F^{-1}(C_2)$ and $F^{-1}(C_1) \cup S_2$. Let K be the collection of squares meeting the latter and B the set of edges of the little squares lying on exactly one square of K. Then only an even number (two or four) of such edges have a common point (see Figure 2). All of S_2 is in B so $B - S_2$ has finitely many edges, with an even number meeting at each endpoint except for the ends, $(\pm 1,0)$, of S_2. Starting at $(-1,0)$, we can trace out a path P along edges of $B - S_2$ which must terminate at $(1,0)$ since the even order condition assures that any other endpoint arrived at can be left by a different edge (if no edge is retraced, the path must indeed terminate). Each edge of P is an edge of a square of K meeting $F^{-1}(C_1) \cup S_2$ and not $F^{-1}(C_2)$, so P misses $F^{-1}(C_2)$. To show P also misses $F^{-1}(C_1)$, note that each edge e of P lies in $B - S_2$. Thus if



Edges e_i of B meet by twos or fours.

FIGURE 2.

$e \subset S = S_1 \cup S_2$, then $e \subset S_1$ and $F(e) \subset F(S_1) = P_1 \subset X - C_1$. If $e \not\subset S$, then e lies on two squares of R, both of which belong to K if e meets $F^{-1}(C_1)$, so then e could not be in B. Thus P misses $F^{-1}(C_1)$ and $F(P)$ is the desired path.

Before going on let us see what we have really accomplished. Suppose R is any collection of square regions which intersect, if at all, in a common edge or vertex (call such an R a **surface**). Let S_1 and S_2 be paths in R meeting only at their ends p, q (replacing $(\pm 1, 0)$ above). If each S_i is a union of odd order edges of R (edges lying on an odd number of squares of R) and no other such edge meets $F^{-1}(C_1 \cup C_2)$, then *the Alexander Addition Theorem still holds for surfaces*. Only minor changes are required in the proof: K is the collection of squares meeting $F^{-1}(C_1) \cup S_2$ as before and B is the set of odd order edges of K. Each vertex of B has even order. That is, if $e_1, \ldots, e_m$ are the edges of B meeting at a vertex v, then m is even. To see this, let $e'_1, \ldots, e'_n$ be the remaining edges of K meeting at v. If we count the two edges of each square having v as a vertex, then the total count is even. But each e_i is counted $2n_i + 1$ times (being of odd order) and e'_i is counted $2n'_i$ times, so $\sum_{i=1}^{m}(2n_i + 1) + \sum_{i=1}^{n} 2n'_i = m + 2(\sum_{i=1}^{m} n_i + \sum_{i=1}^{n} n'_i)$ is even and m must be even. The path P from p to q is constructed as before and the only other change needed is editorial (in the next to last sentence of the proof change "two" to "an even number of" and "both" to "all").

Extension to higher dimensions is analogous. Suppose, for example, that R is the cubical region in E^3 with corners $(\pm 1, \pm 1, \pm 1)$ and $S = S_1 \cup S_2$ its boundary surface where S_1 is the upper and S_2 the lower half. Instead of paths spanning the points x and y, the $P_i = F(S_i)$ are now images of the surfaces S_i each of whose boundaries is $S_1 \cap S_2$. We say S_i **spans** $S_1 \cap S_2$. Slice up R parallel to coordinate planes and define K and B as before (except K is a collection of cubes, B is a set of odd order faces of K and we now seek a surface P **spanning** the square $S_1 \cap S_2$ in the sense that $S_1 \cap S_2$ is the union of the odd order edges of P). Counting as in our discussion of surfaces, we find that every edge of $B - S_2$ except for those in $S_1 \cap S_2$ has even order in $B - S_2$. Remove (or blacken) a face of $B - S_2$ having an odd order edge and continue doing so as long as what is left (unblackened) of $B - S_2$ has a face with an odd order edge (i.e., an edge lying on an odd number of unblackened faces of $B - S_2$). When there are none left, the black faces form a surface $P \subset B - S_2$ whose odd order edges (those on an odd number of black faces) form $S_1 \cap S_2$ and whose image $F(P)$ misses $C_1 \cup C_2$.

Three Important Applications

We now apply the Alexander Addition Theorem to get three famous results. In the first case choose for the space X of the theorem the set S (the boundary of the square region R of FIGURE 1), then choose C_1 and C_2 to be the bottom and top edges of R respectively and take $P_i = S_i$. If there is a map $F: R \to X - (C_1 \cap C_2) = X$, which is the identity when restricted to $X \subset R$, then the Alexander Addition Theorem gives a path $F(P)$ joining $(\pm 1, 0)$ in $X - (C_1 \cup C_2)$, which is impossible. Thus no such extension of the identity (called a **retraction** of R to $X = S$) exists. We state this as a theorem.

NO RETRACTION THEOREM. *No map of R to S maps each point of S to itself.*

In short, S is not a **retract** of R. It is in fact an easy exercise to show that the same is true for any disk (space homeomorphic to R). For example, the unit circle $C=\{(x,y)|x^2+y^2=1\}$ is not a retract of the unit disk $D=\{(x,y)|x^2+y^2\leqslant 1\}$. We also leave as an exercise the proof of the No Retraction Theorem for surfaces spanning S and for higher dimensions.

Another variant of the Intermediate Value Theorem of calculus says that a map of an interval into the reals which maps the endpoints onto themselves takes on every value in the interval. A two-dimensional version of this follows from the No Retraction Theorem. Also, as in one-dimension, Brouwer's Fixed Point Theorem in two dimensions (every map of a disk into itself has a fixed point) is an easy consequence.

INTERMEDIATE VALUE THEOREM. *If f maps the unit disk D into E^2 and is the identity on the unit circle C, then $D\subset f(D)$, that is, f takes on every "value" in D.*

Proof. If $p\in D-f(D)$, then f followed by projection from p to C retracts D to C. But, as we just observed, C is not a retract of D.

BROUWER FIXED POINT THEOREM. *If f maps the unit disk D into E^2, then either $f(p)=p$ for some point p in D or $f(p)=\lambda p$ for some p on the unit circle C and $\lambda>1$. Thus if $f(D)\subset D$ or even $f(C)\subset D$, then f has a fixed point.*

Proof. Define, as usual, $\|v\|=\sqrt{x_1^2+x_2^2}$ if $v=(x_1,x_2)$ in E^2; then set $g(v)=2v-f(2v)$ if $\|v\|\leqslant 1/2$ and $g(v)=(v/\|v\|)-2(1-\|v\|)f(v/\|v\|)$ if $1/2\leqslant\|v\|\leqslant 1$. Then $g:D\to E^2$ is a map which is the identity on C and by the Intermediate Value Theorem $g(v)=(0,0)$ for some v in $D-C$. If $\|v\|\leqslant 1/2$, $f(2v)=2v$ and $p=2v$ is the required point of D. If $\|v\|>1/2$, $f(v/\|v\|)=(2-2\|v\|)^{-1}v/\|v\|=\lambda v/\|v\|$ and $p=v/\|v\|$ is the required point of C since $\|v\|>1/2$ implies $\lambda=(2-2\|v\|)^{-1}>1$.

Connectivity and Components

At this point we have successfully investigated the fixed point theorem which is implicit in the intermediate value theorem. The topological ideas implicit in Green's Theorem and the implicit function theorem require some further development. We will proceed next to a study of the Jordan Curve Theorem which is buried in Green's Theorem. The Jordan Curve Theorem, which we shall state precisely and prove later, says that a simple closed curve divides the plane into exactly two pieces, one inside and one outside. The appropriate topological ideas with which to study "pieces" are the concepts of connectedness and components. Our purpose in this section is to present and develop these ideas.

We begin by noting that the nonexistence of the function F in the Alexander Addition Theorem just yielded three important results with marvelous ease. Our study of connectedness and components will show what can be done when such maps F do exist. It turns out that the hypotheses of the Alexander Addition Theorem are commonly satisfied. Each P_i is a continuous image of an interval $[a,b]$ which is homeomorphic to S_i so a map of $S=S_1\cup S_2$ taking S_i to P_i always exists. However, extending this to a map F of R into X is the problem. Existence of such extensions is important for many other reasons in topology and analysis, so we codify this with a definition. If every map of S into a space extends to a map of R, we call the space **simply connected**. Thus the Alexander Addition Theorem applies whenever $X-C_1\cap C_2$ is simply connected. (Even when it is not, our comments on surfaces may apply.) In particular the Alexander Addition Theorem applies if $X-C_1\cap C_2$ is "starlike", i.e., homeomorphic to a set $Y\subset E^n$ having a point p joinable to every point $y\in Y$ by a straight line segment $py\subset Y$. For if $F:S\to Y$ is a map, it extends to a map of R into Y which takes the origin O to p and each segment Oq, $q\in S$, linearly onto the segment $pF(q)$. Note that by the No Retraction Theorem, S itself is not simply connected, since an extension of the identity would be a retraction.

Turning our attention to the complementary open sets $U_i=X-C_i$, we can discover a simple and useful method for finding and counting the "connected parts" of $U_1\cap U_2$. Euclidean spaces have arbitrarily small neighborhoods whose points are joined by paths within the neighborhood. For such **locally path connected** spaces, an open set U is **connected** iff it is **path connected**, i.e., iff any two points of U are ends of a path in U. A **component** ("connected part") of an open set V is the set of all points of V joinable to a given point by a path in V.

COMPONENTS OF INTERSECTIONS THEOREM. *Let C_1 and C_2 be closed in a space X and let $U_i = X - C_i$. If $U_1 \cup U_2$ is both simply and locally path connected, then each component of $U_1 \cap U_2$ is the intersection of a component of U_1 and one of U_2.*

Proof. If $x \in K$, a component of $U_1 \cap U_2$, and K_i is the component of U_i containing x, then $K \subset K_1 \cap K_2$ since a path joining y to x in $U_1 \cap U_2$ does so in U_1 and U_2 as well. Conversely, if $y \in K_1 \cap K_2$, then $y \in K_i$ so $U_i = X - C_i$ contains a path P_i joining the points x, y of $U_1 \cap U_2 = X - (C_1 \cup C_2)$. By the Alexander Addition Theorem there is a path $F(P)$ in $X - (C_1 \cup C_2) = U_1 \cap U_2$ between x and y. But then $y \in K$ and $K_1 \cap K_2 \subset K$ which means $K_1 \cap K_2 = K$. This completes the proof.

Let $k(U)$ denote the number of components of U. The result just proved implies, as we show next, the very useful formula:

$$k(U_1) + k(U_2) = k(U_1 \cup U_2) + k(U_1 \cap U_2). \tag{1}$$

The observant student of mathematics will notice this same formula in many other contexts, e.g., where $k(U)$ denotes cardinality of finite sets, dimension of linear subspaces, measure, Euler characteristic (Vertices − Edges + Faces), etc.

COUNTING COMPONENTS THEOREM. *The formula $k(U_1) + k(U_2) = k(U_1 \cup U_2) + k(U_1 \cap U_2)$ holds if the U_i are open and $U_1 \cup U_2$ is both simply and locally path connected.*

Proof. Let $W_1, \ldots, W_n$ be a list of all components of U_1 and U_2 and p_m the number of intersecting pairs from each initial list $W_1, \ldots, W_m$. Then $k(U_1) + k(U_2) = n$ and, by the Components of Intersections Theorem, $k(U_1 \cap U_2) = p_n$ since each nonempty $W_i \cap W_j$ is a component of $U_1 \cap U_2$ as well as of any subset of $U_1 \cap U_2$ containing $W_i \cap W_j$. Thus our formula can be written $k(\cup_{i=1}^n W_i) = n - p_n$. To prove it, we use induction on n. Since $k(W_1) = 1 = 1 - p_1$ is trivial, we assume $k(\cup_{i=1}^m W_i) = m - p_m$ for $1 \leqslant m < n$. By definition of p_m, W_{m+1} meets $p_{m+1} - p_m$ of the components $W_1, \ldots, W_m$. If no component K of $\cup_{i=1}^m W_i$ contains two (or more) of these $p_{m+1} - p_m$ W_i, then W_{m+1} meets exactly $p_{m+1} - p_m$ such components K and unites them into one component of $\cup_{i=1}^{m+1} W_i$ so that $k(\cup_{i=1}^{m+1} W_i) = (m - p_m) - (p_{m+1} - p_m) + 1 = (m+1) - p_{m+1}$, completing the proof. But a component K of $\cup_{i=1}^m W_i$ must lie in a component K' of $U_1' = \cup_{i \neq m+1} W_i$. The Components of Intersections Theorem applies to U_1' and $U_2' = W_{m+1}$ since $U_1' \cup U_2' = U_1 \cup U_2$ and the U_i' are open because the components W_i of open locally path connected sets are open. So, unless it is empty, $K' \cap U_2'$ is a component of $U_1' \cap U_2'$. Thus $W_{m+1} = U_2'$ meets only one W_i in K', hence only one in $K \subset K'$ (at most). So our formula is established.

The ***n*-sphere** S^n is the set of points in E^{n+1} a unit distance from the origin: S^0 is a two-point set, S^1 a circle, S^2 an ordinary sphere, etc. Path connectedness is often called **0-connectedness** since it requires that each map of S^0 extend to a map of the interval $[-1, 1]$ spanning S^0. Somewhat more general than simple connectedness is the property of **1-connectedness** which requires that each map of a square boundary, or equivalently of S^1, extend to a surface spanning S^1. Extension to 2-connected and beyond is more complex since there are boundary surfaces other than the 2-sphere (e.g., torus, double torus, etc.). There is an interesting and important relation between this "dimensional" connectedness of a suitable set and of its complement in E^n or S^n called Alexander's Duality Theorem. Our final application of the Alexander Addition Theorem is the proof of this relation for an arc in S^2. An arc is 0- and 1-connected (it is starlike) and the Alexander Duality Theorem amounts to saying the same is true for its complement. This result implies that the addition formula (1) is valid for open sets $U_i \subset S^2$ if the complement of

$U_1 \cup U_2$ is an arc. This can be seen by using our remarks about surfaces in place of the Alexander Addition Theorem to extend our two results about components. The same is true with E^2 in place of S^2 since adding a point "at infinity" to E^2 gives S^2 and removing a point from S^2 does not affect connectedness of open sets.

ALEXANDER DUALITY THEOREM. *The complement of an arc in the plane or 2-sphere is path connected.*

Proof. From the comment preceding the theorem it suffices to prove it for S^2, so let $A = h([-1,1])$ be an arc embedded in S^2 by a homeomorphism h. Two points x, y of $S^2 - A$ are joined by a path P in $S^2 - h(s)$ for any s in $[-1,1]$ since $S^2 - h(s)$ is homeomorphic to E^2. In fact, since $h^{-1}(P)$ is a closed set in $[-1,1]$ not containing s, P misses $h([r,t])$ for some $r \leqslant s \leqslant t$ with $r < s$ unless $s = -1$, and $s < t$ unless $s = 1$. Thus if we choose s to be the least upper bound of all numbers n in $[-1,1]$ such that $h([-1,n])$ misses a path from x to y, then $s > -1$ and some path P_1 from x to y misses $C_1 = h([r,t])$ where $r < s \leqslant t$. By our choice of s there is a path P_2 from x to y missing $C_2 = h([-1,r])$. If $X = S^2$, then $X - (C_1 \cap C_2) = S^2 - h(r)$ and $F: R \to X - (C_1 \cap C_2)$ with $F(S_i) = P_i$ for $i = 1,2$ exists by the construction at the end of the paragraph following the proof of Brouwer's Fixed Point Theorem ($S^2 - h(r)$, being homeomorphic to E^2, is starlike). Thus the Alexander Addition Theorem gives a path from x to y in $X - (C_1 \cup C_2) = S^2 - h([-1,t])$. But if $s < 1$, then $s < t$ contrary to the choice of s, so $s = 1 = t$ and $S^2 - h([-1,t]) = S^2 - A$ is path (or 0-) connected.

The same pattern of proof, using the Alexander Addition Theorem in higher dimensions, shows that the complement of an arc A in S^2 is also 1-connected. Instead of joining given points x, y in $S^2 - A$ by paths, we extend a given map of the square $S_1 \cap S_2$ into $S^2 - A$ to a surface spanning $S_1 \cap S_2$. The P and P_i in the preceding proof are now the images of these surfaces. All extensions needed to complete the proof in the same pattern exist by the construction following Brouwer's Fixed Point Theorem, since, as noted before, S^2 minus a point is starlike.

Indeed, the same type argument proves duality for a disk $D = H(R)$ in S^2 (or E^2) where H is a homeomorphism and R is the square region in E^2 with corners $(\pm 1, \pm 1)$. For example, to show $S^2 - D$ is connected, observe that for $-1 \leqslant s \leqslant 1$, instead of a point $h(s)$ in the above proof, we have an arc $h(s) = H(s \times [-1,1])$. But duality for an arc in S^2 guarantees the extensions needed to complete the proof as before using the Alexander Addition Theorem for surfaces.

The Jordan Curve Theorem

Mathematicians and others long assumed that a simple closed curve (the embedding of a circle) in the plane separates the plane, as does a circle, into two connected pieces and is the boundary of each. In 1865 the German mathematician, Carl Neumann, in a book on integration asked for an explicit proof of this. Over twenty years later in 1887 a French mathematician, Camille Jordan, published a "proof" which was not valid even for a simple closed polygon! We commemorate this pioneering but shaky mathematics by continuing to call what he attempted to prove the Jordan Curve Theorem. It was almost another twenty years before the American topologist Oswald Veblen gave a complete valid proof in 1905. The Dutch mathematician L. E. J. Brouwer extended it to n-dimensional space in 1912 and in 1916 Alexander announced his Duality Theorem which extended it further.

The solution of the original plane problem is made simple by the Alexander Duality Theorem and the formula (1). It amounts to applying this formula to cases where $U_1 = E^2 - C$ and U_2 is any connected open set known to be separated into two components by the curve C. (Dold's proof, mentioned in the introduction, is based on the assumption that the entire curve C lies in such a set U_2. Thus Dold shows a global bisection of space assuming a global bisection of a neighborhood of the curve whereas we obtain a global bisection of space by using a known local bisection by certain auxiliary curves.)

JORDAN CURVE THEOREM. *If C is a simple closed curve in the plane E^2, then E^2-C has two components and C is the boundary of each.*

Proof. If K is a component of E^2-C and U an open set containing a point x of C, then $C-U$ is contained in a subarc A of C. By the Alexander Duality Theorem, E^2-A is connected and so contains a path $P=f([-1,1])$ from $x=f(-1)$ to a point $y=f(1)$ of K. The closed set $f^{-1}(C)$ has a maximum m, $-1\leqslant m<1$ and $f(m)\in U$ so by continuity some interval about m maps into U. If n is a point of this interval and $n>m$, then $f(n)\in U$ and $f([n,1])$ is a path in E^2-C joining y to $f(n)$. Thus $f(n)\in K$ and x is in the closure, $\bar{K}$, of K. Since components of E^2-C are open, $C=\bar{K}-K$ is the boundary of K.

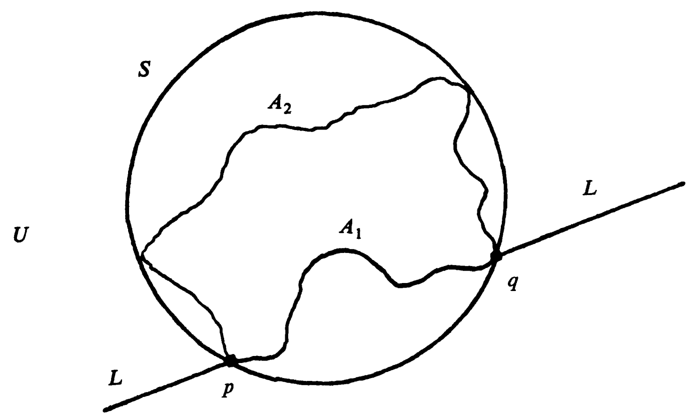


FIGURE 3.

To prove separation, let S be a circle containing at least two points p, q of C and whose exterior U contains no point of C (see FIGURE 3). One can imagine how S may be found by thinking of shrinking a circle inside which the bounded set C lies until it first touches C at a point p and further shrinking it with p fixed until it touches a second point q. The existence of p and q is assured since C is closed. Let A_i $(i=1,2)$ be the arcs of C with ends p and q, L the part of the line through p and q not lying between them and $L_i=L\cup A_i$. Three calculations of component numbers complete the proof:

(i) If $U_1=U$ and $U_2=E^2-L_i$, then $U_1\cup U_2=E^2-A_i$ and, as observed prior to the Alexander Duality Theorem, our component formula (1) applies giving $k(E^2-L_i)=k(E^2-A_i)+k(U-L)-k(U)=1+2-1=2$.

(ii) If $U_i=E^2-L_i$, then $U_1\cup U_2=E^2-L$ is simply and locally path connected since it is open and starlike (in fact it is homeomorphic to E^2). Thus by the Counting Components Theorem, $k(E^2-L_1-L_2)=k(U_1\cap U_2)=k(E^2-L_1)+k(E^2-L_2)-k(E^2-L)=2+2-1=3$.

(iii) If $U_1=E^2-C$ and $U_2=E^2-L_1$, then $U_1\cup U_2=E^2-A_1$. As in (i), the component formula (1) applies and gives $k(E^2-C)=k(E^2-A_1)+k(E^2-L_1-L_2)-k(E^2-L_1)=1+3-2=2$ as required.

Proving the Jordan Curve Theorem seems to require knowing that an arc does not separate the plane. Thus, although there are other proofs at least as elementary (see pp. 100–104 of **[10]**), some form of the Alexander Duality Theorem appears to be unavoidable and it was this that caused the most complication in the proof given here. Most of the complication and clutter is caused by the necessity of extensions to surfaces and higher dimensions which at the same time suggest the methods may generalize to any finite dimension. In fact this is so, and an oft-quoted principle of George A. Polya is at work here: a more general problem may have a simpler solution. The same methods, devoid of the necessity for extensions to surfaces and higher dimensions, are used in **[9]** to prove all these results in n-dimensional space. However, the latter is less intuitive, of course, since it is difficult to visualize beyond three dimensions.

Many applications of the Jordan Curve Theorem require only the polygonal case for which an easy intuitive proof can be found in [3]. An even simpler and rigorous proof (using induction on the number of edges) can be given if the edges are all horizontal or vertical. Try it!

The Invariance of Dimension

There are other important and easy applications of the Alexander Addition Theorem. For example, a half-dozen properties known as Phragmen-Brouwer Properties (see p. 359 of [6]) are easily verified for spaces which are both simply and locally path connected. We close by showing how the Jordan Curve Theorem and the Alexander Duality Theorem imply invariance of domain and dimension.

The inverse image of an open set with respect to any map is open (in fact this is often taken as the definition of continuity) and the image of an open set under a homeomorphism is open (since its inverse is continuous) but not necessarily under an embedding (homeomorphism with a subset). For example, no open subset of the reals maps onto an open set in E^2 under the natural embedding of the reals onto the x_1-axis. However, Euclidean spaces are somewhat unusual in that any embedding of an open subset of E^n (or S^n) in E^n (or S^n) is open. This is known as the Invariance of Domain of Euclidean spaces. Besides useful consequences in analysis, it has the reassuring topological consequence, called Invariance of Dimension, that E^m is not homeomorphic to E^n if $m \neq n$. This fact, first proved by Brouwer in 1911, was especially reassuring then because in 1890 Giuseppe Peano had destroyed the then current concept of dimension by showing that a 2-dimensional square disk is the continuous image of a 1-dimensional interval (in fact, any n-dimensional disk is) and previously Georg Cantor had shown that the points of a line can be put in 1-1 correspondence with those of a plane! The 2-dimensional version of this invariance is easily proved using the Jordan Curve Theorem and the Alexander Duality Theorem for disks.

INVARIANCE OF DOMAIN THEOREM. *If U is open in E^2 and h embeds U in E^2, then $h(U)$ is open.*

Proof. If p is a point of U, then U contains a circle C about p along with its interior I. It suffices to show that $h(I)$ is open. By the Jordan Curve Theorem, $E^2 - h(C)$ has two (open) components. Let V be the one containing $h(p)$ and W the other. Then the connected set $h(I)$ lies in V, and W contains no point of the disk $h(C \cup I)$. The complement of this disk contains W, is contained in $V \cup W$, is connected by the Alexander Duality Theorem for disks, and so must equal W. Then $V = h(I)$ and $h(I)$ is open.

INVARIANCE OF DIMENSION THEOREM. *The real line E^1 is not homeomorphic to the plane E^2.*

Proof. If $g: E^2 \to E^1$ is a homeomorphism and $f: E^1 \to E^2$ is the natural embedding of E^1 onto the x_1-axis of E^2, then $f \circ g = h$ embeds the open set $U = E^2$ and $h(U)$ is not open, contrary to the Invariance of Domain Theorem.

Thus a 1-1 map of E^1 onto E^2 must have a discontinuous inverse. In fact such a map cannot exist but a slightly more sophisticated argument is required. In summary, Cantor exhibited a 1-1 function from E^1 onto E^2. Peano's example yields a continuous function from E^1 onto E^2, but no such function can be both 1-1 and continuous.

A portion of this paper was presented at the April, 1975, meeting of the MAA (Iowa Section) at Iowa State University. Research supported in part by Iowa State University SHRI funds. The author appreciates the interest and assistance of the editors in this paper.

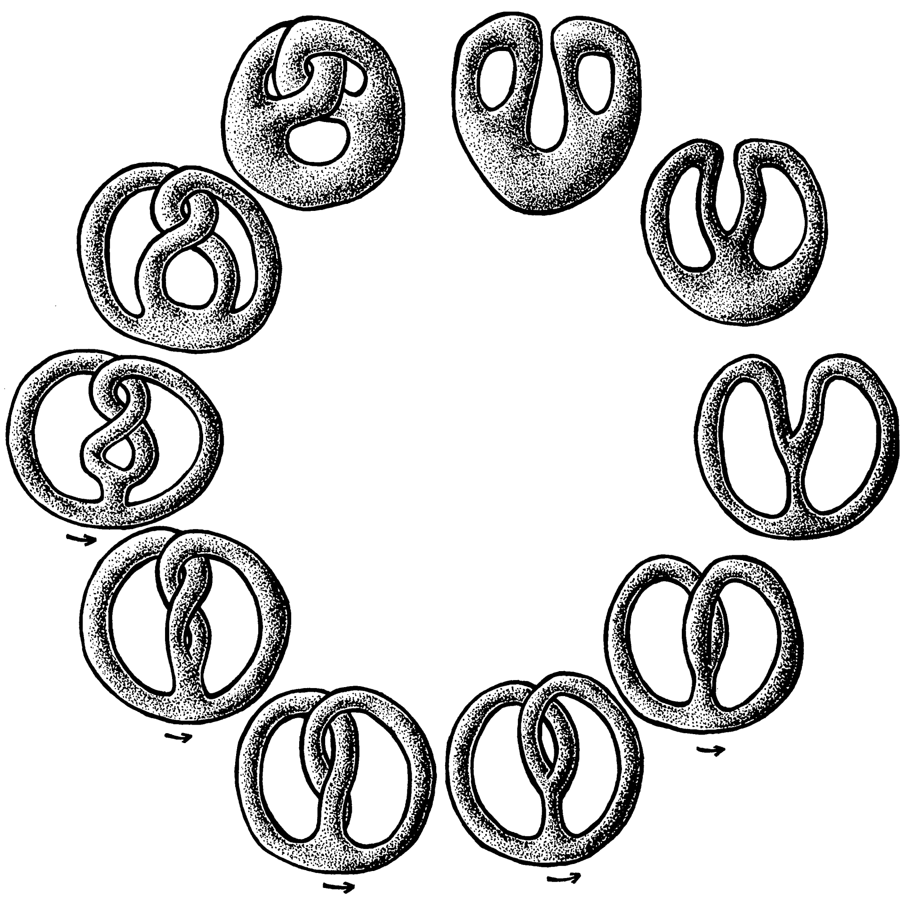
References

[1] J. W. Alexander, A proof and extension of the Jordan-Brouwer separation theorem, Trans. Amer. Math. Soc., 23 (1922) 333–349.

[2] T. M. Apostol, Calculus, vol. 2, Blaisdell, New York, 1962.

[3] B. H. Arnold, Intuitive Concepts in Elementary Topology, Prentice-Hall, Englewood Cliffs, 1962.
[4] R. C. Buck, Advanced Calculus, McGraw-Hill, New York, 1965.
[5] Albrecht Dold, Einfacher Beweis des Jordanschen Kurvensatzes unter zusätzlichen Voraussetzungen, Math.-Phys. Semesterber., Göttingen, 17 (1970) 23–32.
[6] J. G. Hocking and G. S. Young, Topology, Addison-Wesley, Reading, 1961.
[7] Felix Klein, Elementary mathematics from an advanced standpoint, Macmillan, New York, 1939.
[8] M. H. A. Newman, Elements of the Topology of Plane Sets of Points, Cambridge Univ. Press, Cambridge, 1939.
[9] D. E. Sanderson, An elementary romp through some advanced Euclidean topology, Topology Proc., 1 (1976) 79–90.
[10] G. T. Whyburn, Analytic Topology, Amer. Math. Soc. Colloq. Publ., vol. 28, New York, 1942.

If mathematicians made pretzels...



DAN KALMAN
University of Wisconsin
Green Bay, WI 54302

NOTES

A Matrix Method for Solving Linear Congruences

JOHN R. SILVESTER
King's College, University of London
Strand, London WC2R 2LS

A familiar method for solving a diophantine equation such as $12x+41y=1$ is to apply the division algorithm repeatedly until a remainder of 1 (the constant term) is obtained:

$$\begin{aligned} 41 &= (3)(12)+5, \\ 12 &= (2)(5)+2, \\ 5 &= (2)(2)+1; \end{aligned}$$

then reverse the process,

$$\begin{aligned} 1 &= 5-(2)(2) = 5-2(12-(2)(5)) = (5)(5)-(2)(12) \\ &= 5(41-(3)(12))-(2)(12) = (5)(41)-(17)(12). \end{aligned}$$

Hence $x=-17, y=5$ is a solution. (See [1, p. 64] for an explanation of the general method.) The alternative method we describe below, which closely resembles the row-operation method for inverting matrices, in effect performs the two halves of the above calculation simultaneously, and extends to a number of related problems.

To solve the diophantine equation $ax+by=1$, where $(a,b)=1$, we write the matrix

$$A=\begin{pmatrix} a & 1 & 0 \\ b & 0 & 1 \end{pmatrix}$$

and then perform *integral* row-operations: add or subtract an integral multiple of one row from another, until the first column contains $+1$. Thus, to solve $12x+41y=1$ we perform the following sequence of integral row operations:

$$A=\begin{pmatrix} 12 & 1 & 0 \\ 41 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 12 & 1 & 0 \\ 5 & -3 & 1 \end{pmatrix} \to \begin{pmatrix} 2 & 7 & -2 \\ 5 & -3 & 1 \end{pmatrix} \to \begin{pmatrix} 2 & 7 & -2 \\ 1 & -17 & 5 \end{pmatrix}.$$

The solution $x=-17, y=5$ now appears in the row that starts with a 1. Note that at each step we chose the row with the entry of smaller absolute value in the first column, and subtracted from the other row an appropriate multiple so as to reduce the absolute value of its first column entry as much as possible. A useful check is to carry the calculation one step further, to obtain $\binom{1}{0}$ or $\binom{0}{1}$ as the first column, thus:

$$A \to \cdots \to \begin{pmatrix} 2 & 7 & -2 \\ 1 & -17 & 5 \end{pmatrix} \to \begin{pmatrix} 0 & 41 & -12 \\ 1 & -17 & 5 \end{pmatrix}.$$

Then $5\cdot 41 - 17\cdot 12$ appears as the evaluation of the determinant of the matrix with the first column deleted. This is, as desired, equal to 1.

Here is a second example. To solve $8x+13y=1$, we write

$$\begin{pmatrix} 8 & 1 & 0 \\ 13 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 8 & 1 & 0 \\ -3 & -2 & 1 \end{pmatrix} \to \begin{pmatrix} -1 & -5 & 3 \\ -3 & -2 & 1 \end{pmatrix}$$

giving, from the first row, $x=5, y=-3$. (Note the change of signs.)

Let us justify the method. Note that each integral row-operation is equivalent to multiplying on the left by an elementary matrix $\left(\begin{smallmatrix}1 & n\\ 0 & 1\end{smallmatrix}\right)$ or $\left(\begin{smallmatrix}1 & 0\\ n & 1\end{smallmatrix}\right)$ for some integer n, so the sequence of operations multiplies A on the left by some matrix $B=\left(\begin{smallmatrix}p & q\\ r & s\end{smallmatrix}\right)$, where p,q,r,s are integers such that $ps-qr=1$. Hence

$$BA=\begin{pmatrix}p & q\\ r & s\end{pmatrix}\begin{pmatrix}a & 1 & 0\\ b & 0 & 1\end{pmatrix}=\begin{pmatrix}pa+qb & p & q\\ ra+sb & r & s\end{pmatrix}.$$

If $pa+qb=1$, then $x=p,y=q$ solves our equation. If further $ra+sb=0$, then $r=db$ and $s=-da$, for some integer d, since $(a,b)=1$. But now d divides $ps-qr=1$, so $d=\pm 1, r=\pm b$, and $s=\mp a$. A similar argument holds if $pa+qb=-1$, or if $pa+qb=0$ and $ra+sb=\pm 1$.

We leave it to the reader to extend the method to the solution of $ax+by=c$, where a,b are not necessarily coprime, but (a,b) divides c, as well as to the solution of linear diophantine equations in three or more variables.

The method readily adapts to give a quick solution of a linear congruence. One merely writes down that part of the matrix A that is actually needed. So, to solve $ax\equiv b \pmod{n}$, where $(a,n)=1$, we perform integral row operations on the matrix $\left(\begin{smallmatrix}a & b\\ n & 0\end{smallmatrix}\right)$ until a 1 appears in the first column. The solution is then the other entry in the same row. Thus, to solve $12x\equiv 1 \pmod{41}$, we write

$$\begin{pmatrix}12 & 1\\ 41 & 0\end{pmatrix}\to\begin{pmatrix}12 & 1\\ 5 & -3\end{pmatrix}\to\begin{pmatrix}2 & 7\\ 5 & -3\end{pmatrix}\to\begin{pmatrix}2 & 7\\ 1 & -17\end{pmatrix},$$

giving $x\equiv -17 \pmod{41}$.

The method also extends to Chinese remainder theorem calculations, that is, to the solution of simultaneous congruences

$$x\equiv a \pmod{n},\ x\equiv b \pmod{m},$$

where $(n,m)=1$. The solution is $x\equiv p(ma)+q(nb) \pmod{nm}$, where $pm+qn=1$, so we just need to perform our usual calculation using the matrix $\left(\begin{smallmatrix}m & ma\\ n & nb\end{smallmatrix}\right)$. For example, to solve

$$x\equiv 2 \pmod{5}$$
$$x\equiv 11 \pmod{17},$$

we write

$$\begin{pmatrix}17 & 34\\ 5 & 55\end{pmatrix}\to\begin{pmatrix}2 & -131\\ 5 & 55\end{pmatrix}\to\begin{pmatrix}2 & -131\\ 1 & 317\end{pmatrix},$$

giving $x\equiv 317\equiv 62 \pmod{85}$.

To solve

$$ax\equiv b \pmod{n},$$
$$cx\equiv d \pmod{m},$$

where $(n,m)=(a,n)=(c,m)=1$, we use the matrix $\left(\begin{smallmatrix}ma & mb\\ nc & nd\end{smallmatrix}\right)$, provided $(a,c)=1$; if $(a,c)>1$ we use instead the matrix $\begin{pmatrix}ma & mb\\ nc & nd\\ nm & 0\end{pmatrix}$. For example, to solve

$$4x\equiv 3 \pmod{7},$$
$$6x\equiv 1 \pmod{11},$$

we write

$$\begin{pmatrix}44 & 33\\ 42 & 7\\ 77 & 0\end{pmatrix}\to\begin{pmatrix}2 & 26\\ 42 & 7\\ 77 & 0\end{pmatrix}\to\begin{pmatrix}2 & 26\\ 42 & 7\\ 1 & -988\end{pmatrix},$$

giving $x\equiv -988\equiv 13 \pmod{77}$. We leave it to the reader to show that the simultaneous congruences

$$a_1x \equiv b_1 \pmod{n_1},$$
$$a_2x \equiv b_2 \pmod{n_2},$$
$$a_3x \equiv b_3 \pmod{n_3},$$

where $(a_i, n_i) = 1$, all i, and $(n_i, n_j) = 1$, $i \neq j$, can be solved by applying our method to the matrix

$$\begin{bmatrix} n_2n_3a_1 & n_2n_3b_1 \\ n_1n_3a_2 & n_1n_3b_2 \\ n_1n_2a_3 & n_1n_2b_3 \\ n_1n_2n_3 & 0 \end{bmatrix}.$$

Clearly this is becoming quite cumbersome. It is probably easier to solve each congruence separately and then solve the simultaneous congruences two at a time.

Finally, we note that similar calculations can be used in the ring of polynomials in one indeterminate over a field (or indeed in any Euclidean ring). One example will suffice: to find a polynomial $f(x)$ such that $(x^2+1)f(x) \equiv x \pmod{x^3+x^2+1}$, we write

$$\begin{pmatrix} x^2+1 & x \\ x^3+x^2+1 & 0 \end{pmatrix} \to \begin{pmatrix} x^2+1 & x \\ x^2-x+1 & -x^2 \end{pmatrix} \to \begin{pmatrix} x & x^2+x \\ x^2-x+1 & -x^2 \end{pmatrix} \to \begin{pmatrix} x & x^2+x \\ 1 & -x^3-x^2+x \end{pmatrix},$$

so $f(x) \equiv -x^3-x^2+x \equiv x+1 \pmod{x^3+x^2+1}$.

Reference

[1] P. J. Higgins, A first course in abstract algebra, Van Nostrand (New Mathematics Library, No. 7), 1975.

Sums of Powers of Integers via the Binomial Theorem

Barbara Turner

California State University

Long Beach, CA 90840

Many intriguing relationships hold among the sums of powers of positive integers, $s_k(n) = \sum_{j=1}^{n} j^k$ for $k = 1, 2, 3, \ldots$. Beyond the usual interest in expressing the sums $s_k(n)$ explicitly as functions of n, there are recursive identities relating the sums $s_k(n)$ for different k, that give the calculations of the formulas for $s_k(n)$ for any particular k an unusual coherence. Some of the recursive identities derived in this paper have special geometrical significance; they furnish geometrical proofs of the familiar formulas for the sum of positive integers, $s_1(n)$, the sum of the squares of positive integers, $s_2(n)$, and the sum of the cubes of positive integers, $s_3(n)$. Our presentation gives elementary proofs of fundamental identities for sums of powers of positive integers, alternating sums of powers, and sums of powers of odd positive integers. Our approach uses the recurring motif of the binomial theorem, both to clarify and to unite these identities.

We start with the fundamental identity

$$\sum_{k=0}^{r-1} \binom{r}{k} s_k(n) = (n+1)^r - 1, \tag{1}$$

a well-known result which appears in J. Riordan [1]; other interesting proofs can be found in J. L. Paul [2], S. L. Gupta [3], and M. J. A. Sharkey [4]. Riordan's proof depends on the complicated use of exponential generating functions, but the other proofs cited above depend only on an elementary combinatorial argument or on the binomial theorem. The simplest is the

$$a_1x \equiv b_1 \pmod{n_1},$$
$$a_2x \equiv b_2 \pmod{n_2},$$
$$a_3x \equiv b_3 \pmod{n_3},$$

where $(a_i, n_i) = 1$, all i, and $(n_i, n_j) = 1$, $i \neq j$, can be solved by applying our method to the matrix

$$\begin{bmatrix} n_2n_3a_1 & n_2n_3b_1 \\ n_1n_3a_2 & n_1n_3b_2 \\ n_1n_2a_3 & n_1n_2b_3 \\ n_1n_2n_3 & 0 \end{bmatrix}.$$

Clearly this is becoming quite cumbersome. It is probably easier to solve each congruence separately and then solve the simultaneous congruences two at a time.

Finally, we note that similar calculations can be used in the ring of polynomials in one indeterminate over a field (or indeed in any Euclidean ring). One example will suffice: to find a polynomial $f(x)$ such that $(x^2+1)f(x) \equiv x \pmod{x^3+x^2+1}$, we write

$$\begin{pmatrix} x^2+1 & x \\ x^3+x^2+1 & 0 \end{pmatrix} \to \begin{pmatrix} x^2+1 & x \\ x^2-x+1 & -x^2 \end{pmatrix} \to \begin{pmatrix} x & x^2+x \\ x^2-x+1 & -x^2 \end{pmatrix} \to \begin{pmatrix} x & x^2+x \\ 1 & -x^3-x^2+x \end{pmatrix},$$

so $f(x) \equiv -x^3-x^2+x \equiv x+1 \pmod{x^3+x^2+1}$.

Reference

[1] P. J. Higgins, A first course in abstract algebra, Van Nostrand (New Mathematics Library, No. 7), 1975.

Sums of Powers of Integers via the Binomial Theorem

Barbara Turner
California State University
Long Beach, CA 90840

Many intriguing relationships hold among the sums of powers of positive integers, $s_k(n) = \sum_{j=1}^{n} j^k$ for $k = 1, 2, 3, \ldots$. Beyond the usual interest in expressing the sums $s_k(n)$ explicitly as functions of n, there are recursive identities relating the sums $s_k(n)$ for different k, that give the calculations of the formulas for $s_k(n)$ for any particular k an unusual coherence. Some of the recursive identities derived in this paper have special geometrical significance; they furnish geometrical proofs of the familiar formulas for the sum of positive integers, $s_1(n)$, the sum of the squares of positive integers, $s_2(n)$, and the sum of the cubes of positive integers, $s_3(n)$. Our presentation gives elementary proofs of fundamental identities for sums of powers of positive integers, alternating sums of powers, and sums of powers of odd positive integers. Our approach uses the recurring motif of the binomial theorem, both to clarify and to unite these identities.

We start with the fundamental identity

$$\sum_{k=0}^{r-1} \binom{r}{k} s_k(n) = (n+1)^r - 1, \tag{1}$$

a well-known result which appears in J. Riordan [1]; other interesting proofs can be found in J. L. Paul [2], S. L. Gupta [3], and M. J. A. Sharkey [4]. Riordan's proof depends on the complicated use of exponential generating functions, but the other proofs cited above depend only on an elementary combinatorial argument or on the binomial theorem. The simplest is the one suggested by Sharkey, who observed that (1) is actually equivalent (see below) to the identity $\Sigma_{j=1}^{n}\Sigma_{k=0}^{r}\binom{r}{k}=\Sigma_{j=1}^{n}(j+1)^r$, which in turn is immediate from a special case of the binomial theorem: $\Sigma_{k=0}^{r}\binom{r}{k}j^k=(j+1)^r$. (This use of the binomial theorem is the key to our proofs as well.)

The proof of (1) suggested by Sharkey **[4]** (a discussion of the proof appears in H. W. Gould **[5]**) is based on the calculation

$$s_r(n)+\sum_{k=0}^{r-1}\binom{r}{k}s_k(n)=\sum_{k=0}^{r}\binom{r}{k}s_k(n)=\sum_{k=0}^{r}\binom{r}{k}\sum_{j=1}^{n}j^k=\sum_{j=1}^{n}\sum_{k=0}^{r}\binom{r}{k}j^k=$$
$$\sum_{j=1}^{n}(j+1)^r=s_r(n)+(n+1)^r-1.$$

This same technique can also be extended to provide elementary proofs of several other identities that appear on pages 159–160 of Riordan **[1]**, where the suggested proofs require the use of exponential generating functions. For example, to prove

$$s_r(n)-\sum_{k=0}^{r}(-1)^{r-k}\binom{r}{k}s_k(n)=n^r, \tag{2}$$

use the binomial theorem in the form $\Sigma_{k=0}^{r}(-1)^{r-k}\binom{r}{k}j^k=(j-1)^r$, as follows:

$$\sum_{k=0}^{r}(-1)^{r-k}\binom{r}{k}s_k(n)=\sum_{j=1}^{n}\sum_{k=0}^{r}(-1)^{r-k}\binom{r}{k}j^k=\sum_{j=1}^{n}(j-1)^r=s_r(n-1)=s_r(n)-n^r.$$

Hence, $s_r(n)-\Sigma_{k=0}^{r}(-1)^{r-k}\binom{r}{k}s_k(n)=n^r$, which is identity (2).

There are two important and closely related recursive identities for the sums $s_k(n)$:

$$s_k(n)=n^{k+1}-\sum_{j=0}^{k-1}\binom{k}{j}s_{j+1}(n-1) \qquad \text{for} \quad n\geq 2 \tag{3}$$

and

$$s_k(n)=\frac{(n+1)^{k+1}-(n+1)^k-\sum_{j=0}^{k-2}\binom{k}{j}s_{j+1}(n)}{k+1} \qquad \text{for} \quad n\geq 1, \tag{4}$$

where the sum in (4) is defined to be 0 when the upper summation limit is smaller than the lower summation limit. Identity (3) has important geometrical interpretations, which, when combined with the equivalence of (3) and (4), lead to geometrical demonstrations of the familiar formulas for $s_1(n)$, $s_2(n)$, and $s_3(n)$. Identity (4) is also important in its own right since it can be used to provide explicit formulas for $s_k(n)$ for arbitrary k as soon as one knows the formulas for $s_j(n)$ for all strictly smaller j. Identity (4) also yields, by induction on k, that $s_k(n)$ is a polynomial of degree $k+1$ in n.

We first prove (3). Observe that by (1), $\Sigma_{j=0}^{k-1}\binom{k}{j}s_j(n-1)=n^k-1$. Hence

$$\begin{aligned}
s_k(n)=n^k+s_k(n-1)&=\sum_{j=0}^{k-1}\binom{k}{j}s_j(n-1)+1+s_k(n-1)\\
&=s_0(n-1)+\sum_{j=1}^{k-1}\binom{k}{j}s_j(n-1)+1+s_k(n-1)\\
&=s_0(n-1)+\sum_{j=1}^{k}\binom{k}{j}s_j(n-1)+1\\
&=s_0(n-1)+\sum_{j=1}^{k}\left[\binom{k+1}{j}-\binom{k}{j-1}\right]s_j(n-1)+1,
\end{aligned}$$

where the last equality follows from the identity for binomial coefficients:

$$\binom{k}{j}+\binom{k}{j-1}=\binom{k+1}{j}$$

for $j>0$. Thus,

$$s_k(n)=s_0(n-1)+\sum_{j=1}^{k}\left[\binom{k+1}{j}-\binom{k}{j-1}\right]s_j(n-1)+1$$

$$=\sum_{j=0}^{k}\binom{k+1}{j}s_j(n-1)+1-\sum_{j=1}^{k}\binom{k}{j-1}s_j(n-1).$$

Now, applying (1) to the first sum, in the form $\Sigma_{j=0}^{k}\binom{k+1}{j}s_j(n-1)=n^{k+1}-1$, we obtain $s_k(n)=n^{k+1}-\Sigma_{n=0}^{k-1}\binom{k}{j}s_{j+1}(n-1)$, which completes the proof of (3).

Identity (4) may also be proved in a similar manner. However, we shall instead derive (4) from (3), since our geometrical demonstrations of the familiar formulas for $s_1(n)$, $s_2(n)$, and $s_3(n)$ depend on this proof. Identity (3) for $s_k(n+1)$ gives

$$s_k(n+1)=(n+1)^{k+1}-\sum_{j=0}^{k-1}\binom{k}{j}s_{j+1}(n).$$

Since $s_k(n+1)=s_k(n)+(n+1)^k$, it follows that

$$s_k(n)+(n+1)^k=(n+1)^{k+1}-\sum_{j=0}^{k-1}\binom{k}{j}s_{j+1}(n).$$

Transposing the term $\binom{k}{k-1}s_k(n)=ks_k(n)$ from the right side yields

$$(k+1)s_k(n)+(n+1)^k=(n+1)^{k+1}-\sum_{j=0}^{k-2}\binom{k}{j}s_{j+1}(n),$$

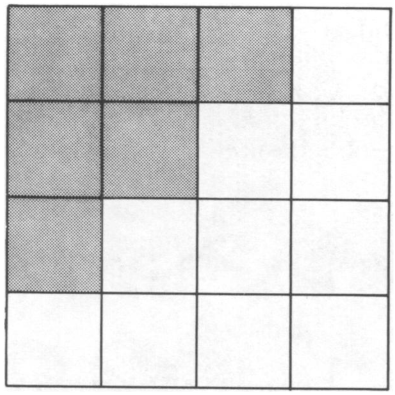
where the sum on the right is defined to be 0 when the upper summation limit is smaller than the lower summation limit. Solving this equation for $s_k(n)$ yields (4). In a similar fashion, (3) can be derived from (4), so that (3) and (4) are equivalent.

We turn now to the geometrical interpretations of the recursive identity (3) which, for the cases $k=1$, 2, and 3, provide geometrical proofs of the familiar formulas for $s_1(n)$, $s_2(n)$, and $s_3(n)$. For the case $k=1$, the recursive identity (3) states that $s_1(n)=n^2-s_1(n-1)$. This equation has a simple geometrical interpretation (FIGURE 1) which is the basis for a popular proof of the familiar formula for $s_1(n)$. The familiar formula for the sum of the first n positive integers, namely $s_1(n)=n(n+1)/2$, may be obtained as follows. Starting with $s_1(n)=n^2-s_1(n-1)$ with $(n+1)$ substituted for n, the pattern of the steps in the proof of (4) from (3) in the case $k=1$, yields $s_1(n+1)=(n+1)^2-s_1(n)$, so $s_1(n)+(n+1)=(n+1)^2-s_1(n)$ and thus $2s_1(n)+(n+1)=(n+1)^2$, hence $s_1(n)=[(n+1)^2-(n+1)]/2$. The last identity is (4) for the case $k=1$, and the familiar formula for $s_1(n)$ follows at once.

The geometrical interpretation of (3) for the case $k=1$ given in FIGURE 1 can be thought of as "completing the square." In B. Turner [6], I used an analogous 3-dimensional process, "completing the cube," to give a new 3-dimensional proof of the familiar formula for $s_2(n)$, the sum of the squares of the first n positive integers, which interpreted geometrically the recursive identity (3) for the case $k=2$, namely $s_2(n)=n^3-[s_1(n-1)+2s_2(n-1)]$. The familiar formula for $s_2(n)$, $s_2(n)=n(n+1)(2n+1)/6$, is then obtained as follows. Carry out the steps of the proof of (4) from (3) in the case $k=2$ and obtain

$$s_2(n)=\frac{(n+1)^3-(n+1)^2-s_1(n)}{3},$$

which is identity (4) for the case $k=2$. Hence, using $s_1(n)=n(n+1)/2$, we obtain



A Geometric Interpretation of $\sum_{i=1}^{n} i = n^2 - \sum_{i=1}^{n-1} i$ for $n=4$.

FIGURE 1.

$$s_2(n) = \frac{(n+1)^3 - (n+1)^2 - \frac{n(n+1)}{2}}{3},$$

which yields the familiar formula for $s_2(n)$.

A similar technique can also be used to give a new geometrical proof of the familiar formula for the sum of the cubes of the first n positive integers, namely $s_3(n) = \left[\frac{n(n+1)}{2}\right]^2$.

The techniques used in establishing the previous recursive identities are also useful in developing the identities for alternating sums of powers of positive integers, namely $a_k(n) = \sum_{j=1}^{n}(-1)^{j+1}j^k$, as well as identities for sums of powers of odd integers. The following two identities are for alternating sums the analogs of identities (1) and (2) (their proofs are similar):

$$\sum_{k=0}^{r}\binom{r}{k}a_k(n) = -a_r(n) + (-1)^{n+1}(n+1)^r + 1 \tag{5}$$

and

$$a_r(n) + \sum_{k=0}^{r}(-1)^{r-k}\binom{r}{k}a_k(n) = (-1)^{n+1}n^r. \tag{6}$$

To prove (5), we use the binomial theorem as before in the form $(j+1)^r = \sum_{k=0}^{r}\binom{r}{k}j^k$.

$$\sum_{k=0}^{r}\binom{r}{k}a_k(n) = \sum_{k=0}^{r}\binom{r}{k}\sum_{j=1}^{n}(-1)^{j+1}j^k = \sum_{j=1}^{n}(-1)^{j+1}\sum_{k=0}^{r}\binom{r}{k}j^k$$

$$= \sum_{j=1}^{n}(-1)^{j+1}(j+1)^r = \sum_{j=1}^{n-1}(-1)^{j+1}(j+1)^r + (-1)^{n+1}(n+1)^r$$

$$= -\sum_{j=1}^{n-1}(-1)^{j+2}(j+1)^r + (-1)^{n+1}(n+1)^r = -\sum_{j=2}^{n}(-1)^{j+1}j^r + (-1)^{n+1}(n+1)^r$$

$$= -[a_r(n) - 1] + (-1)^{n+1}(n+1)^r = -a_r(n) + (-1)^{n+1}(n+1)^r + 1.$$

This proves (5). The proof of (6) is similar to the proof of (2).

Finally we derive a useful recursive formula for the sum of the kth powers of the first m odd positive integers, $d_k(m) = \sum_{j=1}^{m}(2j-1)^k$. This identity is

$$\sum_{k=0}^{r-1}\left[1-(-1)^{r-k}\right]\binom{r}{k}d_k(m)=2^r m^r. \qquad (7)$$

Since $\sum_{j=1}^{m}(2j)^r=2^r m^r+\sum_{j=1}^{m}(2j-2)^r$, we have

$$\sum_{j=1}^{m}((2j-1)+1)^r-(2j-1)^r=2^r m^r+\sum_{j=1}^{m}((2j-1)-1)^r=(2j-1)^r.$$

The binomial theorem gives

$$\sum_{j=1}^{m}\left(\sum_{k=0}^{r}\binom{r}{k}(2j-1)^k-(2j-1)^r\right)=2^r m^r+\sum_{j=1}^{m}\left(\sum_{k=0}^{r}(-1)^{r-k}\binom{r}{k}(2j-1)^k-(2j-1)^r\right),$$

or

$$\sum_{j=1}^{m}\sum_{k=0}^{r-1}\binom{r}{k}(2j-1)^k=2^r m^r+\sum_{j=1}^{m}\sum_{k=0}^{r-1}(-1)^{r-k}\binom{r}{k}(2j-1)^k.$$

Thus,

$$\sum_{k=0}^{r-1}\binom{r}{k}\sum_{j=1}^{m}(2j-1)^k=2^r m^r+\sum_{k=0}^{r-1}(-1)^{r-k}\binom{r}{k}\sum_{j=1}^{m}(2j-1)^k.$$

Hence,

$$\sum_{k=0}^{r-1}\binom{r}{k}d_k(m)=2^r m^r+\sum_{k=0}^{r-1}(-1)^{r-k}\binom{r}{k}d_k(m).$$

From this identity, (7) is immediate.

References

[1] J. Riordan, Combinatorial Identities, Wiley, New York, 1968, pp. 159–160.

[2] J. L. Paul, On the sum of the kth powers of the first n integers, Amer. Math. Monthly, 78 (1971) 271–272.

[3] S. L. Gupta, An identity involving the sum of the kth powers of the first n natural numbers, Math. Gaz., 56 (1972) 128–129.

[4] M. J. A. Sharkey, An identity involving sums of powers, Math. Gaz., 57 (1973) 131–133.

[5] H. W. Gould, Sums of Powers of Integers, Number Theory Class Notes, Part I, West Virginia University, 1974–1975, pp. 3–4.

[6] B. Turner, Completing the cube, Math. Teacher, 70 (1977) 67–70.

The Unique Factorization Theorem: From Euclid to Gauss

MARY JOAN COLLISON

Molloy College

Rockville Centre, NY 11570

"Strangely obscure" is how Harold Davenport [3, p. 19] described the history of the Unique Factorization Theorem—that every composite integer greater than one has a unique expression as a product of primes (apart from the order in which the factors appear). An examination of the historical documents indicates that the confusion concerning the theorem's development is partly a consequence of notational deficiencies in Euclid's era and partly due to an unquestion-

$$\sum_{k=0}^{r-1}\left[1-(-1)^{r-k}\right]\binom{r}{k}d_k(m)=2^rm^r. \tag{7}$$

Since $\sum_{j=1}^{m}(2j)^r=2^rm^r+\sum_{j=1}^{m}(2j-2)^r$, we have

$$\sum_{j=1}^{m}((2j-1)+1)^r-(2j-1)^r=2^rm^r+\sum_{j=1}^{m}((2j-1)-1)^r=(2j-1)^r.$$

The binomial theorem gives

$$\sum_{j=1}^{m}\left(\sum_{k=0}^{r}\binom{r}{k}(2j-1)^k-(2j-1)^r\right)=2^rm^r+\sum_{j=1}^{m}\left(\sum_{k=0}^{r}(-1)^{r-k}\binom{r}{k}(2j-1)^k-(2j-1)^r\right),$$

or

$$\sum_{j=1}^{m}\sum_{k=0}^{r-1}\binom{r}{k}(2j-1)^k=2^rm^r+\sum_{j=1}^{m}\sum_{k=0}^{r-1}(-1)^{r-k}\binom{r}{k}(2j-1)^k.$$

Thus,

$$\sum_{k=0}^{r-1}\binom{r}{k}\sum_{j=1}^{m}(2j-1)^k=2^rm^r+\sum_{k=0}^{r-1}(-1)^{r-k}\binom{r}{k}\sum_{j=1}^{m}(2j-1)^k.$$

Hence,

$$\sum_{k=0}^{r-1}\binom{r}{k}d_k(m)=2^rm^r+\sum_{k=0}^{r-1}(-1)^{r-k}\binom{r}{k}d_k(m).$$

From this identity, (7) is immediate.

References

[1] J. Riordan, Combinatorial Identities, Wiley, New York, 1968, pp. 159–160.

[2] J. L. Paul, On the sum of the kth powers of the first n integers, Amer. Math. Monthly, 78 (1971) 271–272.

[3] S. L. Gupta, An identity involving the sum of the kth powers of the first n natural numbers, Math. Gaz., 56 (1972) 128–129.

[4] M. J. A. Sharkey, An identity involving sums of powers, Math. Gaz., 57 (1973) 131–133.

[5] H. W. Gould, Sums of Powers of Integers, Number Theory Class Notes, Part I, West Virginia University, 1974–1975, pp. 3–4.

[6] B. Turner, Completing the cube, Math. Teacher, 70 (1977) 67–70.

The Unique Factorization Theorem: From Euclid to Gauss

MARY JOAN COLLISON
Molloy College
Rockville Centre, NY 11570

"Strangely obscure" is how Harold Davenport [3, p. 19] described the history of the Unique Factorization Theorem—that every composite integer greater than one has a unique expression as a product of primes (apart from the order in which the factors appear). An examination of the historical documents indicates that the confusion concerning the theorem's development is partly a consequence of notational deficiencies in Euclid's era and partly due to an unquestioning acceptance of "self-evident" propositions by 18th century mathematicians. The history of the Unique Factorization Theorem is a good example of "nonlinearity" in the development of mathematics.

The Unique Factorization Theorem has been the subject of many conflicting reports. The mathematical historian Morris Kline indicated [**9**, pp. 79–80, 817] that Proposition 14 of Book IX of Euclid's *Elements* gives the Unique Factorization Theorem. Heath's edition of the *Elements*, Kline's major source, restated the proposition as, "A number can be resolved into prime factors in only one way" [**8**, p. 403]. Yet Hardy and Wright in [**7**], one of the foremost textbooks in number theory, claimed that the theorem was not stated until 1801 when Gauss presented it in Article 16 of his *Disquisitiones Arithmeticae* [**7**, pp. 10, 181–182, 189]. Hardy and Wright attributed their information to remarks made by Solomon Bochner. Bochner recalled this incident in [**1**, p. 827]:

> The nearest that Euclid himself came to this theorem was his proposition (vii, 24): 'If two numbers be prime to any number, their product also will be prime to the same.'

Bochner also quoted from his book [**2**, p. 16]:

> ...there is no substance to assertions that the 'fundamental theorem' had been consciously known to mathematicians before Gauss, but that they had neglected to make the fact known. We think that the 17th and even the 18th century was not yet ready for the peculiar kind of mathematical abstraction which the 'fundamental theorem' involves....

Bochner's underestimation of Euclid, of which Davenport was also guilty, has been perpetuated by Hardy and Wright through many editions and printings. Actually, as Heath recognized, Euclid came quite close to a statement of the Unique Factorization Theorem in Proposition 14 of Book IX of his *Elements* [**8**, p. 402]:

> If a number be the least that is measured by prime numbers, it will not be measured by any other prime number except those originally measuring it.

In Euclid's terminology, "measured by" meant "divisible by." There was no system for expressing exponents (with the consequence that square numbers and cube numbers had to be explicitly defined to be discussed), and rarely was a number concretely specified as having more than three divisors. Thus, it is not surprising that Euclid's proof of Proposition 14 begins by assuming that A is the smallest number measured by the primes B, Γ, and Δ. Then A cannot have any other prime divisor, for if another prime E measured A, where E is different from B, Γ, and Δ, then A would have another factor Z so that (in modern terminology) $A = E \cdot Z$. Then each of the primes B, Γ, and Δ measures the product $E \cdot Z$ which forces each prime to measure Z. This follows from Proposition 30 of Book VII [**8**, p. 331]:

> If two numbers by multiplying one another make some number, and any prime number measure the product, it will also measure one of the original numbers.

Because E is also a prime, Z is shown to be measured by the primes B, Γ, and Δ. However, Z is smaller than A which was assumed to be the least number measured by the primes B, Γ, Δ. This contradiction shows that A cannot have any other prime factors.

Considering the restrictions imposed by a system of notation which precluded exponents and did not allow the expression of numbers as having an arbitrary number of factors, Euclid's Proposition 14 must be accepted as containing the essence of the Unique Factorization Theorem. The only qualification which can be made is that Euclid does not actually state that every integer is expressible as a product of primes. (LeVeque carefully makes this observation in his book [**11**, p. 30].) However, in Proposition 31 of Book VII, Euclid does prove that "any composite number is measured by some prime number." [**8**, p. 332] The method he gives is the same as the one he later used to prove Proposition 14: A sequence of divisors (each dividing the previous one) is formed which must eventually give a prime factor; their decreasing size as integers precludes an infinite number of such factors. Once this proposition has been proved, the repeated application of the procedure of finding divisors would give a prime product representation. This last step is never mentioned in the *Elements*.

The absence of a statement by Euclid involving representation is not surprising, however. In the terminology of his time, the closest one could get to a statement concerning a product was "E by multiplying Z has made A" [**8**, p. 403], or "the number produced from" two or three specified numbers (which Heath then translates as "product") [**8**, p. 326]. No word existed for "product," but that is obviously the concept Euclid was forming when he let A be "the smallest number measured by B, Γ, Δ" in his proof of the Unique Factorization Theorem.

This same concept was utilized in Euclid's proof of the existence of infinitely many primes. The proposition itself (20 in Book IX) states [**8**, p. 412]: "Prime numbers are more than any assigned multitude of prime numbers." Euclid begins the proof: "Let the least number measured by the primes A, B, C be taken." The proof, paraphrased in more modern terminology and notation, continues:

> Take the product ABC and add 1. Then $ABC+1$ is either a prime or not. If it is a prime, we have added another prime number to those already given. If it is not, it must be measured by some prime number (VII, Prop. 31), say G. Now G cannot be identical with any of the prime numbers A, B, C. For, if it is, it will divide ABC. Therefore, since it divides $ABC+1$ also, it will divide the difference, which is 1: This is absurd. Thus G is a new prime number; the prime numbers A, B, C, G have been found which are more than the assigned multitude of A, B, C. Q.E.D.

This proof is considered an example of the elegance of Euclid, yet no one questions the fact that a product of primes is here being formed.

It is interesting to note that a reverse phenomenon takes place in Euler's and Legendre's work of the late 18th and early 19th centuries. In his *Elements of Algebra* (1770), Euler does not state the Unique Factorization Theorem at all, but he does give a method for resolving any number into its simple factors. The method is simply to divide by the different primes until all prime factors are found. He then states [5, pp. 17–18, 27]:

> When, therefore, we have represented any number assumed at pleasure by its simple factors, it will be very easy to shew all the numbers by which it is divisible.

He is clearly assuming that any factorization thus obtained will be unique.

Legendre, in his *Théorie des Nombres*, presents the same information but in a more detailed form. A literal translation of his statement would be [**10**, p. 5]:

> Any number N, if it is not a prime, can be represented by the product of several prime numbers α, β, γ, etc., each raised to a power, in a manner that one can always assume $N = \alpha^m \beta^n \gamma^p$, etc.

He gives the division method for finding the representation and then states [**10**, pp. 6–7]:

> Having reduced a number N to the form $\alpha^m \beta^n \gamma^p$, etc., every divisor of this number will also be of the form $\alpha^\mu \beta^\nu \gamma^\pi$, etc., where the exponents μ, ν, π, etc., cannot surpass m, n, p, etc.

No proof of this is given. He seems to consider it self-evident.

Thus, both Euler and Legendre discuss the representation of a composite N as a product of primes, but do not state or prove anything concerning the uniqueness of such a representation. Their assumption that *all* divisors of N can be found by taking the primes in the representation raised to the different exponents is only valid under the additional assumption that no other prime representation exists. In view of the fact that they never even stated that the representation was unique, Euclid was clearly much more rigorous on this point.

Ultimately, we must turn to Gauss' writings to find the exact statement of the Unique Factorization Theorem, nearly as we present it today. In his *Disquisitiones Arithmeticae*, Gauss states [**6**, p. 15]: "Numerus compositus quicumque unico tantum modo in factores primos resolvi potest." ("Any composite number can be resolved into prime factors in only one way.") As proof, Gauss utilizes his preceding theorem which translates as:

> If none of the numbers a, b, c, d, etc. is divisible by the prime p, then neither will the product $abcd$, etc. be divisible by p.

Thus, if a number is expressed as a product of prime powers, it follows that the same primes must appear in every factorization. Gauss then shows that if a number has two different factorizations in which the same prime appears with unequal exponents, then the division of both factorizations by that prime raised to the smaller power produces a contradiction of the preceding theorem. In general, the many forms of "cancelling out" proofs may be considered as derived from Gauss' work.

It seems clear from the relevant writings of Euclid, Euler, Legendre, and Gauss that Euclid deserves the credit for the Unique Factorization Theorem. His Proposition 31 (Book VII) indicates that a prime divisor can always be found for any composite. According to Proposition 14 (Book IX), once a certain product of primes is given, no other factorization is possible. The statement that every composite has a prime product representation is omitted because of vocabulary limitations and the absence of a system of notation permitting a representation as a product of arbitrary factors. Yet the proof of this representation statement requires merely the repeated application of Proposition 31. It seems unreasonable to withhold credit from Euclid only because he omitted a phrase which was inexpressible in the vocabulary of his time.

This position is supported by an examination of Gauss' work. Gauss had the advantage of a good notation and knowledge of the *Elements*, as well as of Euler's writings (but he specified that he had already completed most of his work on the *Disquisitiones Arithmeticae* before obtaining a copy of Legendre's *Essai sur la Théorie des Nombres*) [**6**, p. 5]. Before proving the Unique Factorization Theorem, Gauss gave the equivalent of Euclid's Proposition 30, and then said [**6**, p. 15]:

> Euclid had already proved this theorem in his *Elements*.... However, we do not wish to omit it because many modern authors have employed vague computations in place of the proof or have neglected the theorem completely....

Gauss began the proof of the Unique Factorization Theorem by saying [**6**, p. 16]:

> It follows from the elementary theory that any composite number can be resolved into prime factors, but it is tacitly assumed, and generally without proof, that this cannot be done in many different ways.

This shows that he considered the crux of the matter to be the proof of the uniqueness of the representation. In fact, Gauss did not even bother to prove that the representation existed, and he seemed to consider the Unique Factorization Theorem to be something already known, although not rigorously dealt with by his contemporaries.

Euler and Legendre's acceptance of the uniqueness of the prime representation would not have been contradicted by their mathematical experience. It is often with great reluctance that even today's beginning students in number theory are led to accept the fact that a proof of unique factorization is needed. The students are likely to consider it "obvious." For this reason, undergraduate teachers of number theory are forever finding clever elementary examples of mathematical systems in which unique factorization fails. This mirrors the experience of the centuries. It was not until such systems arose historically that the Unique Factorization Theorem was singled out as being "fundamental" and requiring proof. In fact, as late as 1843, Ernst Kummer was still capable of mistakenly assuming unique factorization in a certain class of algebraic numbers, according to a story told by Hensel. (See [**4**] for a discussion of the historical facts involving this possibly inaccurate story.) So natural did the property seem that this same error was made by Augustin Cauchy, another outstanding 19th century mathematician [**9**, p. 819].

Prior to the investigation of systems of complex numbers composed of higher roots of unity in conjunction with the search for higher reciprocity laws, the question of uniqueness had never arisen in applications. Hence, Euclid deserves enormous credit for even considering the property of uniqueness in the first place, especially since he did not possess a suitable vocabulary and notation to state the full theorem. Euler and Legendre had the notation but were too willing to accept without question or proof the uniqueness portion of the theorem.

Perhaps the geometric nature of Euclid's vocabulary did not make uniqueness seem so self-evident as it appeared to be in the algebraic atmosphere of the 17th and 18th centuries. Gauss' comments and the continued assumption of unique factorization (even when it did not apply) on the part of later mathematicians undermine Bochner's judgment concerning the lower level of mathematical abstraction in the 17th and 18th centuries as compared with the 19th century. Thus mathematics, like life itself, does not always proceed linearly.

References

[1] Salomon Bochner, Mathematical reflections, Amer. Math. Monthly, 81 (1974) 827–852.

[2] ______, The Role of Mathematics in the Rise of Science, Princeton Univ. Press, Princeton, 1966.

[3] H. Davenport, The Higher Arithmetic, 2nd ed., Hutchinson, London, 1962; 1st ed., 1952.

[4] Harold M. Edwards, The background of Kummer's proof of Fermat's last theorem for regular primes, Arch. History Exact Sci., 14 (1975) 219–236.

[5] Leonard Euler, Elements of Algebra, translator not identified, J. Johnson, St. Paul's Churchyard, London, 1797 (originally 1770).

[6] Carl F. Gauss, Disquisitiones Arithmeticae, Werke, Bd. I, Koniglichen Gesellschaft der Wissenschaften, Gottingen, 1863, pp. 1–466 (originally 1801). Also available in an English translation by Fr. Arthur Clarke, S.J., Yale Univ. Press, New York, 1966.

[7] G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Oxford Univ. Press, 1960; 1st ed., 1938.

[8] Thomas Heath, The Thirteen Books of Euclid's Elements, vol. 2, Cambridge Univ. Press, Cambridge, 1926.

[9] Morris Kline, Mathematical Thought from Ancient to Modern Times, Oxford Univ. Press, New York, 1972.

[10] Adrien-Marie Legendre, Théorie des Nombres, vol. 1, Librairie Scientifique et Technique A. Blanchard, Paris, 1955, 1st ed., 1830; the Essai originally appeared in 1798.

[11] William LeVeque, Fundamentals of Number Theory, Addison-Wesley, Reading, Mass., 1977.

A Family of Ninth Order Magic Squares

CHARLES W. TRIGG

2404 Loring Street

San Diego, CA 92109

A family of 134,217,728 ninth order magic squares can be generated from the standard nine-digit third order square displayed in part (1) of FIGURE 1. First add nine to each element of the square to form a second square labeled (2) in FIGURE 1; then repeat the operation until the eight derived squares of FIGURE 1 have been formed. Each of these squares is magic and

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
8 1 6	17 10 15	26 19 24	35 28 33	44 37 42	53 46 51	62 55 60	71 64 69	80 73 78
3 5 7	12 14 16	21 23 25	30 32 34	39 41 43	48 50 52	57 59 61	66 68 70	75 77 79
4 9 2	13 18 11	22 27 20	31 36 29	40 45 38	49 54 47	58 63 56	67 72 65	76 81 74

Nine 3×3 magic squares.

FIGURE 1.

Prior to the investigation of systems of complex numbers composed of higher roots of unity in conjunction with the search for higher reciprocity laws, the question of uniqueness had never arisen in applications. Hence, Euclid deserves enormous credit for even considering the property of uniqueness in the first place, especially since he did not possess a suitable vocabulary and notation to state the full theorem. Euler and Legendre had the notation but were too willing to accept without question or proof the uniqueness portion of the theorem.

Perhaps the geometric nature of Euclid's vocabulary did not make uniqueness seem so self-evident as it appeared to be in the algebraic atmosphere of the 17th and 18th centuries. Gauss' comments and the continued assumption of unique factorization (even when it did not apply) on the part of later mathematicians undermine Bochner's judgment concerning the lower level of mathematical abstraction in the 17th and 18th centuries as compared with the 19th century. Thus mathematics, like life itself, does not always proceed linearly.

References

[1] Salomon Bochner, Mathematical reflections, Amer. Math. Monthly, 81 (1974) 827–852.

[2] ______, The Role of Mathematics in the Rise of Science, Princeton Univ. Press, Princeton, 1966.

[3] H. Davenport, The Higher Arithmetic, 2nd ed., Hutchinson, London, 1962; 1st ed., 1952.

[4] Harold M. Edwards, The background of Kummer's proof of Fermat's last theorem for regular primes, Arch. History Exact Sci., 14 (1975) 219–236.

[5] Leonard Euler, Elements of Algebra, translator not identified, J. Johnson, St. Paul's Churchyard, London, 1797 (originally 1770).

[6] Carl F. Gauss, Disquisitiones Arithmeticae, Werke, Bd. I, Koniglichen Gesellschaft der Wissenschaften, Gottingen, 1863, pp. 1–466 (originally 1801). Also available in an English translation by Fr. Arthur Clarke, S.J., Yale Univ. Press, New York, 1966.

[7] G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Oxford Univ. Press, 1960; 1st ed., 1938.

[8] Thomas Heath, The Thirteen Books of Euclid's Elements, vol. 2, Cambridge Univ. Press, Cambridge, 1926.

[9] Morris Kline, Mathematical Thought from Ancient to Modern Times, Oxford Univ. Press, New York, 1972.

[10] Adrien-Marie Legendre, Théorie des Nombres, vol. 1, Librairie Scientifique et Technique A. Blanchard, Paris, 1955, 1st ed., 1830; the Essai originally appeared in 1798.

[11] William LeVeque, Fundamentals of Number Theory, Addison-Wesley, Reading, Mass., 1977.

A Family of Ninth Order Magic Squares

CHARLES W. TRIGG

2404 Loring Street

San Diego, CA 92109

A family of 134,217,728 ninth order magic squares can be generated from the standard nine-digit third order square displayed in part (1) of FIGURE 1. First add nine to each element of the square to form a second square labeled (2) in FIGURE 1; then repeat the operation until the eight derived squares of FIGURE 1 have been formed. Each of these squares is magic and

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
8 1 6	17 10 15	26 19 24	35 28 33	44 37 42	53 46 51	62 55 60	71 64 69	80 73 78
3 5 7	12 14 16	21 23 25	30 32 34	39 41 43	48 50 52	57 59 61	66 68 70	75 77 79
4 9 2	13 18 11	22 27 20	31 36 29	40 45 38	49 54 47	58 63 56	67 72 65	76 81 74

Nine 3×3 magic squares.

FIGURE 1.

remains magic under eight transformations: the square as shown, rotations through 90°, 180° and 270°, and the mirror images of these four.

To construct ninth order magic squares divide a 9-by-9 grid into nine 3-by-3 grids. Label this 3-by-3 grid by the corresponding entries of the original magic square. In each small 3-by-3 grid place the 3×3 square, in any of its eight orientations, that has the same identification number as the grid. Since each small grid can be filled in eight ways, 8^9 (or 134,217,728) ninth order magic squares (exclusive of rotations and reflections) can be constructed in this fashion from the first 81 integers, with each of the rows, columns and principal diagonals summing to the magic constant 369. In FIGURE 2 a different orientation is given to each of the eight outer 3-by-3 squares.

71	64	69	4	3	8	47	54	49
66	68	70	9	5	1	52	50	48
67	72	65	2	7	6	51	46	53
24	19	26	44	37	42	60	61	56
25	23	21	39	41	43	55	59	63
20	27	22	40	45	38	62	57	58
29	34	33	78	73	80	17	12	13
36	32	28	79	77	75	10	14	18
31	30	35	74	81	76	15	16	11

A 9×9 magic square.

FIGURE 2.

This technique can be applied to each of the ninth order squares to create a total of $(8^9)^2$ or 18 01439 85094 81984 twenty-seventh order squares from the first 729 integers and with a magic constant of 9855.

It is interesting to note that a related method of producing magic squares seems to have been used by the ancient Chinese [1].

Reference

[1] F. Swetz, Mysticism and magic in the number squares of old China, Math. Teacher (January 1978) 50–56.

Generalized Magic Cubes

DANA RICHARDS
University of Illinois at Urbana-Champaign
Urbana, IL 61801

Magic squares, $n\times n$ arrays of the integers 0 to n^2-1 (or 1 to n^2) where all the rows and columns have the same "magic" sum, have been known for well over a thousand years. As ways of producing such arrays became well known and understood, people inevitably started developing generalizations. Perhaps the most interesting generalization is the pandiagonal square, where all $2n$ diagonals (including the "broken" diagonals) also have the magic sum. This can be visualized nicely by identifying the top and bottom edges of the square, as well as the two sides, to form a torus. All the rows and columns and diagonals on this square (turned into a torus) have the magic sum. Pandiagonal squares are now also well understood, prompting yet another generalization: If we specify a collection of arbitrary "lines," where a line is a set of any n cells of the array, can we then place the integers so that each line has the magic sum?

remains magic under eight transformations: the square as shown, rotations through 90°, 180° and 270°, and the mirror images of these four.

To construct ninth order magic squares divide a 9-by-9 grid into nine 3-by-3 grids. Label this 3-by-3 grid by the corresponding entries of the original magic square. In each small 3-by-3 grid place the 3×3 square, in any of its eight orientations, that has the same identification number as the grid. Since each small grid can be filled in eight ways, 8^9 (or 134,217,728) ninth order magic squares (exclusive of rotations and reflections) can be constructed in this fashion from the first 81 integers, with each of the rows, columns and principal diagonals summing to the magic constant 369. In FIGURE 2 a different orientation is given to each of the eight outer 3-by-3 squares.

71	64	69	4	3	8	47	54	49
66	68	70	9	5	1	52	50	48
67	72	65	2	7	6	51	46	53
24	19	26	44	37	42	60	61	56
25	23	21	39	41	43	55	59	63
20	27	22	40	45	38	62	57	58
29	34	33	78	73	80	17	12	13
36	32	28	79	77	75	10	14	18
31	30	35	74	81	76	15	16	11

A 9×9 magic square.

FIGURE 2.

This technique can be applied to each of the ninth order squares to create a total of $(8^9)^2$ or 18 01439 85094 81984 twenty-seventh order squares from the first 729 integers and with a magic constant of 9855.

It is interesting to note that a related method of producing magic squares seems to have been used by the ancient Chinese [1].

Reference

[1] F. Swetz, Mysticism and magic in the number squares of old China, Math. Teacher (January 1978) 50–56.

Generalized Magic Cubes

DANA RICHARDS
University of Illinois at Urbana-Champaign
Urbana, IL 61801

Magic squares, $n\times n$ arrays of the integers 0 to n^2-1 (or 1 to n^2) where all the rows and columns have the same "magic" sum, have been known for well over a thousand years. As ways of producing such arrays became well known and understood, people inevitably started developing generalizations. Perhaps the most interesting generalization is the pandiagonal square, where all $2n$ diagonals (including the "broken" diagonals) also have the magic sum. This can be visualized nicely by identifying the top and bottom edges of the square, as well as the two sides, to form a torus. All the rows and columns and diagonals on this square (turned into a torus) have the magic sum. Pandiagonal squares are now also well understood, prompting yet another generalization: If we specify a collection of arbitrary "lines," where a line is a set of any n cells of the array, can we then place the integers so that each line has the magic sum?

This generalization proved to be too nebulous, so we restricted our attention to a clearly defined set of lines called $[x,y]$-lines. An **$[x,y]$-line** is a set of n cells such that the coordinates of any two cells (x_1,y_1) and (x_2,y_2), are related by $(x_1+ax,y_1+ay)\equiv(x_2,y_2) \bmod n$ for some integer a. In the torus interpretation, the cells of an $[x,y]$-line appear as "steps" of (x,y) (i.e., x "up" and y "over") around the torus. We call $[x,y]$ the **slope** of the line. The rows, columns, and right and left diagonals are $[0,1]$-, $[1,0]$-, $[1,1]$-, and $[1,-1]$-lines, respectively. In general, given a collection C of slopes, $C=\{[x_i,y_i]|i=1,\dots,k\}$, we say a square is **magic with respect to C** if it has the magic sum for each $[x,y]$-line whose slope $[x,y]$ is in C.

Although we will use this notation for collections of slopes in what follows, our principal interest is generalization in another direction. We wish to study magic cubes and magic hypercubes. In this note we will concentrate on algorithms for finding magic cubes and relationships among these algorithms. We will treat higher dimensional extensions only in the last section.

For a cube we define an **$[x,y,z]$-line** to be a set of n cells (of the $n\times n\times n$ cube) which are connected by (x,y,z) steps taken modulo n. As before, given a collection of slopes, $C=\{[x_i,y_i,z_i]|i=1,\dots,k\}$, we say the cube is **magic with respect to C** if each $[x,y,z]$-line whose slope $[x,y,z]$ is in C has the magic sum.

We say that a cube is **magic** if each of the $3n$ $n\times n$ square slices of the cube parallel to a coordinate plane are pandiagonal squares. If in addition all $4n^2$ diagonals parallel to the 4 main diagonals of the cube have the magic sum and the cube is magic, we say it is **pandiagonal**. A cube is pandiagonal if it is magic with respect to C, where $C=\{[1,0,0], [0,1,0], [0,0,1], [1,1,0], [1,-1,0], [1,0,1], [1,0,-1], [0,1,1], [0,1,-1], [1,1,1], [1,1,-1], [1,-1,1], [1,-1,-1]\}$.

Our construction of magic cubes (not usually pandiagonal) uses an algorithm which is an extension of the Graeco-Latin method which produces the so-called orthogonal Latin squares. In the past, this technique was virtually abandoned when a more tractable algorithm, called the Uniform Step algorithm, became popular. However, the Uniform Step method yielded only limited results **[1]** when extended to three and higher dimensions. Recent work using the Graeco-Latin method has been done by Wynne **[10]**, Howard **[4]**, and Brooke **[3]**. The last paper contains many special cases of our results.

We begin our study of algorithms by reviewing the methods of constructing magic squares from Latin squares. A Latin square is a square design that has a permutation of $\{0,1,\dots,n-1\}$ in each row and column. To generalize these designs to $[x,y]$-lines, we call a square **Latin with respect to a collection of slopes** C if each $[x,y]$-line whose slope $[x,y]$ is in C contains a permutation of $\{0,1,\dots,n-1\}$. We leave to the reader the obvious generalization for a cube and a set of slopes $C=\{[x_i,y_i,z_i]|i=1,\dots,k\}$.

There is an enormous literature devoted to the construction of Latin squares, but we will need only the simplest of Latin square algorithms. To construct an $n\times n$ Latin square L select integers a_1 and a_2 relatively prime to n and let $L_{ij}\equiv a_1 i+a_2 j$. A necessary and sufficient condition that L be Latin with respect to a set of slopes C is that a_1x+a_2y be relatively prime to n for every $[x,y]$ in C. Shapiro **[9]** proved both this and the remarkable result that if for a particular n and C there exists a Latin square with respect to C, then there exist suitable a_1 and a_2. Similarly, we can construct a Latin cube L by letting $L_{ijk}\equiv a_1 i+a_2 j+a_3 k$. We call the triple (a_1,a_2,a_3) the **parameter**. If $a_1x+a_2y+a_3z$ is relatively prime to n for every slope $[x,y,z]$ in C, then L is Latin with respect to C. The proof of this is a simple extension of the proof for squares.

Two Latin squares are orthogonal if when superimposed the n^2 ordered pairs of elements in corresponding positions are distinct. If the two Latin squares were constructed with parameters (a_1,a_2) and (a_1',a_2'), then they are orthogonal if $\det\begin{pmatrix} a_1 & a_2 \\ a_1' & a_2' \end{pmatrix} \bmod n$ is prime to n. Similarly three Latin cubes are orthogonal if the n^3 ordered triples of elements in corresponding positions are distinct. If the parameters are $a=(a_1,a_2,a_3)$, $a'=(a_1',a_2',a_3')$ and $a''=(a_1'',a_2'',a_3'')$, then the three cubes are orthogonal if $\det(P)$ mod n is prime to n, where P has rows a,a',a''.

An $n\times n$ magic square M can be constructed from an orthogonal pair of $n\times n$ Latin squares, L and L', by taking $M_{ij}=nL_{ij}'+L_{ij}$. Similarly, three orthogonal cubes L,L',L'', each Latin with respect to C, can be used to form a magic cube M, which is magic with respect to C, by taking $M_{ijk}=n^2L''_{ijk}+nL'_{ijk}+L_{ijk}$. To demonstrate how all this ties together, we exhibit in FIGURE 1 an order-5 cube which is magic with respect to $C=\{[1,0,0],[0,1,0],[0,0,1],[1,-1,0],[1,0,-1],[0,1,-1]\}$. Each of the parameters (1,2,4), (2,3,4) and (1,2,3) constructs cubes Latin with respect to C. Since $\det\begin{pmatrix}1&2&4\\2&3&4\\1&2&3\end{pmatrix}\equiv 1 \bmod 5$, the cubes are orthogonal. The three Latin cubes and the magic cube are shown in FIGURE 1. For example, a $[1,0,-1]$-line is 11, 78, 40, 107, 74 which has the magic sum 310.

We shall call this algorithm for constructing generalized magic cubes the **Graeco-Latin algorithm**, the historical name for orthogonal Latin squares. We now describe the more widely known **Uniform-Step algorithm**, and demonstrate the close relationship between the two.

The Uniform-Step algorithm is quite simple for squares. Choose parameters $(b_1,b_2),(b'_1,b'_2)$ so that $\det\begin{pmatrix}b_1&b'_1\\b_2&b'_2\end{pmatrix}$ is prime to n. Number the cells from 0 to n^2-1 by starting at some specified cell and stepping as on a torus. Take steps of (b_1,b_2) until you return to a numbered cell, then step further (b'_1,b'_2) to a vacant square, once, and continue with (b_1,b_2) steps and so on. For a cube, use parameters (b_1,b_2,b_3), (b'_1,b'_2,b'_3) and (b''_1,b''_2,b''_3). Start at some cell and step (b_1,b_2,b_3) until a numbered cell is hit, then use (b'_1,b'_2,b'_3) once. If the step (b'_1,b'_2,b'_3) hits a numbered cell, then use (b''_1,b''_2,b''_3) and, in either case, return to the (b_1,b_2,b_3) step. For further details see **[8]**. We shall call the matrix whose columns are the triples of parameters the matrix of parameters.

We now turn to the relation between the Uniform-Step and Graeco-Latin algorithms. Assume that both produced the same square M. Suppose that some cell (i,j) of M has the number with digits $k'k$, written in base n. Assume, without loss of generality since our operations are done mod n, that the starting cell for the Uniform-Step algorithm is $(0,0)$. A little reflection shows that the Uniform-Step algorithm places $k'k$ on (i,j) iff $b_1k+b'_1k'\equiv i$ and $b_2k+b'_2k'\equiv j$ mod n. Further, from the Graeco-Latin algorithm, we know that $a_1i+a_2j\equiv k$ and $a'_1i+a'_2j\equiv k'$ mod n. Since the two squares are the same, with a little algebra we find $QP\equiv I$ mod n, where I is the identity matrix, and Q and P are the matrices of parameters for the Uniform-Step and Graeco-Latin algorithms respectively.

Similarly for the case of magic cubes, we find $Q\equiv I$ mod n. So if the Graeco-Latin method constructs a cube magic with respect to C with parameter P, then the Uniform-Step method constructs one with $Q\equiv P^{-1}$ mod n. There is a one-to-one mapping between the two parameters and thus, between the cubes they construct. (Note that the requirement that det(P) be prime to n plays the same role as $\det(P)\neq 0$ normally does in assuring invertibility.) For example, the cube shown in FIGURE 1 can be constructed by the Uniform-Step algorithm with $Q=\begin{pmatrix}1&2&1\\3&4&4\\1&0&4\end{pmatrix}$.

Even though theoretically one algorithm "is as good as" the other, this is not operationally true: With the Graeco-Latin method it is easy to generate a parameter matrix P consistent with a given set of slopes C, while for cubes, there is no way to verify that the parameter matrix Q is consistent with C. The relationship between P and Q (but not the mapping) was discovered for the square by Lehmer **[5**, p. 537**]**. These results were largely anticipated in a forgotten paper by McDonald **[6]**. The germ of this approach appeared in **[2]** in 1888 and again in **[1]**.

By looking at a given set of slopes C we can easily determine the values of n for which parameters exist. For squares known results are $n\not\equiv 0$ mod 2 for a magic square; and $n\not\equiv 0$ mod 2, mod 3 or mod 5 for a magic cube and in addition $n\not\equiv 0$ mod 7 for a pandiagonal cube.

We now turn to the question of the main diagonals of magic cubes. A square is **associative** if the sum of any two pairs of cells symmetric about the center are equal. An associative cube is defined in an analogous way. Note that associativity assures that the main diagonals have the magic sum. These main diagonal sums have often been included in the definition of magic squares and cubes. However, we find that they do not fit our framework of $[x,y]$-lines since $[1,1]$-lines include too many "diagonals." In general we can obtain the main diagonal sums by associativity (for the permissible values of n).

Level	L''	L'	L	Magic Cube M $n^2L''+nL'+L$
	0 1 2 3 4	0 2 4 1 3	0 1 2 3 4	0 36 72 83 119
	2 3 4 0 1	3 0 2 4 1	2 3 4 0 1	67 78 114 20 31
Top	4 0 1 2 3	1 3 0 2 4	4 0 1 2 3	109 15 26 62 98
	1 2 3 4 0	4 1 3 0 2	1 2 3 4 0	46 57 93 104 10
	3 4 0 1 2	2 4 1 3 0	3 4 0 1 2	88 124 5 41 52
	3 4 0 1 2	4 1 3 0 2	4 0 1 2 3	99 105 16 27 63
	0 1 2 3 4	2 4 1 3 0	1 2 3 4 0	11 47 58 94 100
II	2 3 4 0 1	0 2 4 1 3	3 4 0 1 2	53 89 120 6 42
	4 0 1 2 3	3 0 2 4 1	0 1 2 3 4	115 1 37 73 84
	1 2 3 4 0	1 3 0 2 4	2 3 4 0 1	32 68 79 110 21
	1 2 3 4 0	3 0 2 4 1	3 4 0 1 2	43 54 85 121 7
	3 4 0 1 2	1 3 0 2 4	0 1 2 3 4	80 116 2 38 74
III	0 1 2 3 4	4 1 3 0 2	2 3 4 0 1	22 33 69 75 111
	2 3 4 0 1	2 4 1 3 0	4 0 1 2 3	64 95 106 17 28
	4 0 1 2 3	0 2 4 1 3	1 2 3 4 0	101 12 48 59 90
	4 0 1 2 3	2 4 1 3 0	2 3 4 0 1	112 23 34 65 76
	1 2 3 4 0	0 2 4 1 3	4 0 1 2 3	29 60 96 107 18
IV	3 4 0 1 2	3 0 2 4 1	1 2 3 4 0	91 102 13 49 55
	0 1 2 3 4	1 3 0 2 4	3 4 0 1 2	8 44 50 86 122
	2 3 4 0 1	4 1 3 0 2	0 1 2 3 4	70 81 117 3 39
	2 3 4 0 1	1 3 0 2 4	1 2 3 4 0	56 92 103 14 45
	4 0 1 2 3	4 1 3 0 2	3 4 0 1 2	123 9 40 51 87
Bottom	1 2 3 4 0	2 4 1 3 0	0 1 2 3 4	35 71 82 118 4
	3 4 0 1 2	0 2 4 1 3	2 3 4 0 1	77 113 24 30 66
	0 1 2 3 4	3 0 2 4 1	4 0 1 2 3	19 25 61 97 108

Construction of a 5×5×5 magic cube.

FIGURE 1.

When using the Uniform-Step algorithm, an appropriate choice of the starting cell makes the square or cube associative (see [**1**] for details). If each of the Latin squares or cubes used by the Graeco-Latin algorithm is associative with a pair-sum of $n-1$, then the main diagonals of M will have the magic sum. This is easy to achieve since each of the Latin cubes is already associative mod n. We then note both that adding (mod n) a fixed integer to each cell of a Latin square or cube will not affect the magic property of the resulting magic cube and that by adding an appropriate integer each square or cube will have pair-sums equal to $n-1$. As an example, consider the cube in FIGURE 1. By adding (mod 5) 0, 4 and 3 to L'', L' and L, respectively, we get the cube shown in FIGURE 2 which is associative and magic with respect to the C used in FIGURE 1.

All the results of this note extend to m-dimensional hypercubes in straightforward ways. We define $[x_1, x_2, \cdots, x_m]$-lines in the same spirit as $[x,y]$-lines. And a parameter $(a_1, a_2, \cdots, a_m)$ leads to an m-cube Latin with respect to a set of slopes C if and only if $a_1x_1 + a_2x_2 + \cdots + a_mx_m$ is prime to n for each line with slope in C. Using m of these parameters to construct an $m \times m$ matrix P such that $\det(P)$ is prime to n yields an M which is magic with respect to C. If the Uniform-Step algorithm is generalized to m dimensions, we still have $QP \equiv I \bmod n$.

Since no standard definitions have been accepted, we offer the following definitions for general use. Begin with the standard definition of a magic and pandiagonal square. An m-cube is magic iff all its $(m-1)$-cube sections are pandiagonal. It is pandiagonal if, in addition, all diagonals parallel to the 2^{m-1} main diagonals have the magic sum. As noted earlier this algorithm will not create magic m-cubes for even n, since a few of the requisite lines are not

	L''	L'	L	Magic Cube M $n^2L''+nL'+L$
	0 1 2 3 4	4 1 3 0 2	3 4 0 1 2	23 34 65 76 112
	2 3 4 0 1	2 4 1 3 0	0 1 2 3 4	60 96 107 18 29
Top	4 0 1 2 3	0 2 4 1 3	2 3 4 0 1	102 13 49 55 91
	1 2 3 4 0	3 0 2 4 1	4 0 1 2 3	44 50 86 122 8
	3 4 0 1 2	1 3 0 2 4	1 2 3 4 0	81 117 3 39 70
	3 4 0 1 2	3 0 2 4 1	2 3 4 0 1	92 103 14 45 56
	0 1 2 3 4	1 3 0 2 4	4 0 1 2 3	9 40 51 87 123
II	2 3 4 0 1	4 1 3 0 2	1 2 3 4 0	71 82 118 4 35
	4 0 1 2 3	2 4 1 3 0	3 4 0 1 2	113 24 30 66 77
	1 2 3 4 0	0 2 4 1 3	0 1 2 3 4	25 61 97 108 19
	1 2 3 4 0	2 4 1 3 0	1 2 3 4 0	36 72 83 119 0
	3 4 0 1 2	0 2 4 1 3	3 4 0 1 2	78 114 20 31 67
III	0 1 2 3 4	3 0 2 4 1	0 1 2 3 4	15 26 62 98 109
	2 3 4 0 1	1 3 0 2 4	2 3 4 0 1	57 93 104 10 46
	4 0 1 2 3	4 1 3 0 2	4 0 1 2 3	124 5 41 52 88
	4 0 1 2 3	1 3 0 2 4	0 1 2 3 4	105 16 27 63 99
	1 2 3 4 0	4 1 3 0 2	2 3 4 0 1	47 58 94 100 11
IV	3 4 0 1 2	2 4 1 3 0	4 0 1 2 3	89 120 6 42 53
	0 1 2 3 4	0 2 4 1 3	1 2 3 4 0	1 37 73 84 115
	2 3 4 0 1	3 0 2 4 1	3 4 0 1 2	68 79 110 21 32
	2 3 4 0 1	0 2 4 1 3	4 0 1 2 3	54 85 121 7 43
	4 0 1 2 3	3 0 2 4 1	1 2 3 4 0	116 2 38 74 80
Bottom	1 2 3 4 0	1 3 0 2 4	3 4 0 1 2	33 69 75 111 22
	3 4 0 1 2	4 1 3 0 2	0 1 2 3 4	95 106 17 28 64
	0 1 2 3 4	2 4 1 3 0	2 3 4 0 1	12 48 59 90 101

Construction of a $5\times5\times5$ associative magic cube.

FIGURE 2.

magic. However, in these cases, an adaptation of Margossian's method as seen in [1] can be used for some n. (Details can be obtained from the author.)

These procedures work well for generating complex magic m-cubes, where $m=3$, but as m increases significantly, more values of n become special cases. So there is room for a more powerful and general theory. There are lots of open questions. For example, Planck [7] has shown that the smallest m-cube that is pandiagonal is of order 2^m and, if associativity is required, of order 2^m+1. We conjecture that the smallest (associative) magic m-cube is of order 2^m-1.

References

[1] W. W. R. Ball and H. S. M. Coxeter, Mathematical Recreations and Essays, Univ. of Toronto Press, Cambridge, 1974.

[2] F. A. P. Barnard, Theory of magic squares and of magic cubes, Mem. Nat. Acad. Sci., 4 (1888) 209–270.

[3] P. Brooke, Perfect and pandiagonal magic hyper-cubes, Math. Spect., 9 (1976–77) 82–94.

[4] I. P. Howard, Pandiagonal, associative magic cubes and m-dimensional magic arrays, J. Recreational Math., 9 (1976–77) 276–278.

[5] D. N. Lehmer, On the congruences with certain magic squares, Trans. Amer. Math. Soc., 31 (1929) 529–551.

[6] K. McDonald, Magic cubes which are uniform step cubes, Univ. Calif. Publ. Math., 2 (1934) 197–215.

[7] C. Planck, Theory of Path Nasiks, privately published, 1905.

[8] D. S. Richards, Combinatorial Games and Puzzles, Master's thesis, University of Virginia, 1976.

[9] H. D. Shapiro, Generalized Latin squares on a torus, Discrete Math., 24 (1978) 63–77.

[10] B. E. Wynne, Perfect magic cubes of order seven, J. Recreational Math., 8 (1975–76) 285–293.

Magic Possibilities of the Weighted Average

RUMA FALK
MAYA BAR-HILLEL
Hebrew University
Jerusalem, Israel

Some years ago a well-known public official in California quit his job and moved to Alabama, thereby—in the words of a local editorial—raising the average IQ in both states. That this should indeed be possible is apparent. By the same token it is also possible, by a mere redistribution of the population of the United States, to raise the average IQ in all fifty states. Does this, then, imply that the average IQ of the entire country—being itself an average of the average IQs of its constituent states—can be thus raised?

Upon reflection one realizes that what saves us from a paradoxical "yes" is the fact that as average IQ's steadily increase in the different states, their respective weights shift, too, with the weights for the "California"s going down and those for the "Alabama"s going up.

Weighted averages depend, of course, not only on the values being averaged, but on their weights as well. Students readily perceive that, while holding values constant, a weighted average can be made to assume any value between the maximum value and the minimum value by a suitable choice of weights. They are, however, greatly surprised that weights can actually counteract the effects of changing values, i.e., that a set of values can *all* be increased (or decreased), yet its average might decrease (or increase) if the weights are changed appropriately.

A nice example of this effect emerges from a study of bias in graduate school admissions carried out at the University of California Berkeley campus (see [1]). In the fall of 1973 the University admitted 44% of all male applicants and only 35% of the female applicants. In an attempt to discover the departments "guilty" of discrimination, the admission rates for each sex were tabulated by departments. Surprisingly, it turned out that most of the departments had similar rates of admission for men and for women, and where there was a considerable difference, it tended to be in favor of the women!

When confronted with this seeming contradiction between global (i.e., university-wide) and local (i.e., department-by-department) statistics, those who aren't completely baffled usually hypothesize that there must exist at least one department which has higher admission rates for men, and which is very popular, thus outweighing the effects of other departments with higher admission rates for women. That this is not necessarily so is illustrated in the following hypothetical example:

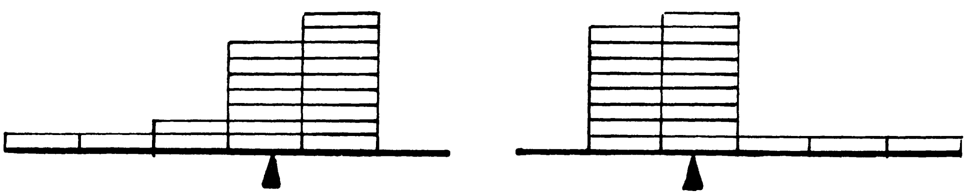
> A certain Art School has only two departments—Painting and Music. The school has been collecting statistics since it was founded, and these show that the Painting department admits about 80% of its female applicants versus about 40% of its male applicants, whereas the Music department, which has stricter requirements, admits about 12% of its female applicants, versus about 6% of its male applicants. The statistics show, furthermore, that women are more interested in Music, i.e., 90% of the female applicants to the school enroll for Music and 10% enroll for Painting, while amongst men the opposite is true: 90% of the male applicants enroll for Painting and only 10% enroll for Music.

A group of 66 applicants to the Hebrew University read this story in the context of their university entrance exams and were asked the following question: "On the basis of these statistics who, would you say, has had a better chance of being admitted to this school over the years, men or women?" Fifty-six percent of the respondents answered that women had had better chances of admission than men, clearly taking their cue from the admissions statistics. (We have independent evidence, not presented here, to the effect that the 56% figure is not merely the result of randomly choosing a sex.) In fact, although in each department a woman's chance of being admitted is double that of a man, in the school as a whole the reverse is true.

We can verify this by using the formula of total probability $P(A)=P(A/B)\cdot P(B)+P(A/\bar{B})\cdot P(\bar{B})$. The overall rate of admission for men is 37% $((.4)(.9)+(.06)(.1)=.366)$, and for women it is only 19% $((.8)(.1)+(.12)(.9)=.188)$. This reversal is due to the effect of the enrollment statistics: Women enroll heavily in the more selective department, whereas men favor the department easier to get into.

Once the enrollment statistics are taken into account, they lend a different interpretation to the global admission rates: these may be higher for men not because men are better qualified, but merely because they prefer the easy-to-get-into departments. Maybe some people interpret "chance of admission" as a measure of how qualified an individual is rather than as an indication of how selective an admissions policy is, thus overlooking the sex difference in rates of application to the two departments.

The "paradoxical" properties of the weighted average can be illustrated in a concrete representation. Suppose a set of uniform blocks arranged in stacks of varying heights is located on a weightless platform (as, for example, in Figure 1a), which is balanced on a pivot located at the center of gravity. There are two obvious ways to shift the center of gravity to the right: either shift the entire construction of blocks (or some part of it) rightward on the platform, or move some blocks from some stacks to other stacks on their right. One can, however, shift the entire construction to the *right*, while simultaneously moving individual blocks to other stacks on their *left*. If done appropriately, the net result could then be a new center of gravity which is to the left of the old one (see FIGURE 1b).



(a) (b)

FIGURE 1.

The intriguing effect which weights may have on a weighted average become acute with respect to several common measures in statistics and economics. For example, the most popular measure which demographers use for characterizing the mortality of a population is the Crude Death Rate (CDR)—defined as the ratio of deaths during a year to the mid-year population. However, they are aware that it is possible for the CDR to be higher in population X than in population Y, even though all the age-specific death rates are higher in population Y than in population X. This happens to be the actual case with respect to the Jewish (as population X) and Arabic (as Y) subpopulations in Israel. The global measure, CDR, though more compact than a list of specific death rates, is affected not only by the level of mortality which it is intended to measure, but also by the age distribution in the population, from which its weights derive. Indeed, because the Jewish population is "older," both its CDR and its Life Expectancy (a measure which is, in a way, the inverse of mortality) are higher than those of Israeli Arabs. Similar considerations pertain to so-called index numbers, such as the cost-of-living index, all of which are weighted averages.

As a consequence of the above considerations, many simple questions such as "Who has a better chance of admission?" or "Which population has a higher mortality?" may not be sufficiently well defined to have a uniquely determined answer.

We thank G. Schwarz for bringing the California-Alabama anecdote to our attention.

Reference

[1] D. Freeman, R. Pisani, and R. Purves, Statistics, Norton, 1978.

Vector Spaces of Magic Squares

JAMES E. WARD III
Bowdoin College
Brunswick, ME 04011

Exercise. An **$n \times n$ magic square** is an $n \times n$ matrix of real numbers in which the sum along each row, each column and each diagonal is a constant (called the **line-sum** of the magic square). For example, three 3×3 magic squares with line-sums 15, 3/2 and 0, respectively, are

$$\begin{matrix} 2 & 9 & 4 \\ 7 & 5 & 3 \\ 6 & 1 & 8 \end{matrix} \qquad \begin{matrix} -1 & 5/2 & 0 \\ 3/2 & 1/2 & -1/2 \\ 1 & -3/2 & 2 \end{matrix} \qquad \begin{matrix} 1 & 0 & -1 \\ -2 & 0 & 2 \\ 1 & 0 & -1 \end{matrix}$$

(i) Show that the (matrix) sum of two $n \times n$ magic squares is an $n \times n$ magic square. If the line-sums of the squares are m_1 and m_2, what is the line-sum of the sum?
(ii) Show that the (matrix) scalar multiple of an $n \times n$ magic square by a real number k is an $n \times n$ magic square. If the original square has line-sum m, what is the line-sum of the scalar multiple?
(iii) Is the set of all $n \times n$ magic squares (with all possible line-sums) a vector space? Why?
(iv) Is the set of all $n \times n$ magic squares with line-sum $m \neq 0$ a vector space? Why?
(v) Is the set of all $n \times n$ magic squares with line-sum zero a vector space? Why?

This exercise, suggested by Fletcher [3], encourages consideration of the algebraic structure of magic squares, as opposed to methods for generating them. In this article we follow Fletcher's suggestion, using familiar linear algebra techniques to determine the dimensions of the vector spaces of magic squares. Then we use these dimensions to establish an upper bound on the number of magic squares.

Magic squares have fascinated people for centuries. A Chinese emperor is supposed to have seen one—on the back of a divine turtle, no less—as early as 2200 B. C. From that time on, mystical properties have been ascribed to them. In the middle ages, a magic square engraved on a silver plate and worn about the neck was thought to ward off the plague [5]. Writing in 1844, Hutton [4] reported:

> These squares have been called magic squares because the ancients ascribed to them great virtues, and because this disposition of numbers formed the basis and principle of many of their talismans. According to this idea a square of one cell, filled up with unity, was the symbol of the Deity, on account of the unity and immutability of God; for they remarked that this square was, by its nature, unique and immutable, the product of unity by itself being always unity. The square of the root two was the symbol of imperfect matter, both on account of the four elements and of the impossibility of arranging this square magically. A square of nine cells was assigned or consecrated to Saturn, that of sixteen to Jupiter, that of twenty-five to Mars, that of thirty-six to the Sun, that of forty-nine to Venus, that of sixty-four to

Mercury, and that of eighty-one, or nine on each side, to the Moon. Those who can find any relation between the planets and such an arrangement of numbers must, no doubt, have minds strongly tinctured with superstition; but such was the tone of the mysterious philosophy of Jamblichus, Porphyry, and their disciples. Modern mathematicians, while they amuse themselves with these arrangements, which require a pretty extensive knowledge of combination, attach to them no more importance than they really deserve.

Nevertheless, mathematicians both before and after 1844 apparently attached enough importance to magic squares to write thousands of pages about them. In 1888, F. A. P. Barnard, then president of Columbia and after whom Barnard College is named, published a 61-page paper **[2]** at the end of which he included an "approximately complete" bibliography of 47 scholarly papers and books on the subject. Today a complete bibliography on magic squares would probably require all 61 pages!

The magic squares which are best known are those $n\times n$ squares which use only the first n^2 positive integers. These square arrays of the numbers 1 to n^2 will be referred to as **classical** magic squares. The first example given in the opening exercise is a 3×3 classical magic square.

Each magic square yields seven other magic squares, obtained from it by rotating it in the plane through angles of 90°, 180°, and 270° and by rotating it in space about its horizontal, vertical and two diagonal axes. These seven, together with the original, constitute the **symmetries** of the magic square. Symmetric magic squares are regarded as being identical. It is easy to show that the first example in the opening exercise is the only 3×3 classical magic square up to symmetry.

More generally, we shall call any $n\times n$ array of n^2 (integral, real, or complex) numbers in which each line-sum is constant a **magic square**. The second and third examples of the opening exercise are real 3×3 magic squares. Note that the same number may appear several times in a magic square, a statement which is not true of classical magic squares.

This paper will consider only real magic squares, although all of the results are true for complex magic squares as well. If the entries are restricted to the integers, all of the results hold if "vector space" is replaced judiciously by "$\mathbb{Z}$-module." From now on, all magic squares will be real magic squares unless it is specifically stated otherwise. Because 1×1 and 2×2 magic squares are not very interesting and because they bog down the proofs with special cases, it will be assumed that $n\geqslant 3$.

We shall denote by **MS(*n*)** the set of all $n\times n$ magic squares; by ***m*MS(*n*)** the set of all $n\times n$ magic squares with line-sum m; and by **0MS(*n*)** the set of all $n\times n$ magic squares with line-sum zero. The opening exercise reveals that MS(n) and 0MS(n) are vector spaces but that mMS(n) for $m\neq 0$ is not. The space 0MS(n) is a subspace of MS(n) which is, in turn, a subspace of the n^2-dimensional vector space of all $n\times n$ real matrices. Thus the dimension of MS(n) is at most n^2. (Note that mMS(n) is never empty: it always contains the square in which each entry is m/n.)

Let us call each magic square in 0MS(n), whose line-sums are all zero, a **zero magic square**. We will call two $n\times n$ magic squares **equivalent** if one can be obtained from the other by adding the same real number to each entry. It follows, trivially, that each magic square is equivalent to one and only one zero magic square: if an $n\times n$ magic square has line-sum m, it can " zeroed" by subtracting m/n from each entry. Thus there is a one-to-one correspondence between the set mMS(n) for a fixed m and the vector space 0MS(n).

This sets the stage for the main result of this paper.

THEOREM. *The dimension of* $0MS(n)$ *is* n^2-2n-1.

Proof. If an $n\times n$ matrix $A=(a_{ij})$ is in 0MS(n), its $2n+2$ line-sums are all zero. Thus there are $2n+2$ homogeneous linear equations in the n^2 variables $a_{ij}, 1\leqslant i,j\leqslant n$. Write these equations in the following order, called the **standard order**: first the n row sums in order, then the n column sums in order, then the NW-SE diagonal sum, and, last, the SW-NE diagonal sum. The resulting coefficient matrix will be a $(2n+2)\times n^2$ matrix of 0's and 1's. When $n=3$, it is the 8×9 matrix

$$\begin{matrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0, \end{matrix}$$

where the elements in the jth column are the coefficients of the jth variable in the list $a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}, a_{31}, a_{32}, a_{33}$.

In the coefficient matrix determined by A, the first $2n-1$ rows are clearly linearly independent. But the $2n$th row is a linear combination of the first $2n-1$, being the sum of the first n rows minus the sum of rows $n+1$ through $2n-1$. Moreover, the last two rows are both linearly independent of the first $2n-1$ rows: the nth column of the coefficient matrix has 1's in the first, $2n$th and $(2n+2)$nd rows, and zeroes everywhere else, and the n^2th column has 1's in the nth, $2n$th and $(2n+1)$st rows, and zeroes everywhere else, making it impossible to find a nontrivial zero linear combination of the first $2n-1$ rows with either of the last two rows. Finally, it is clear that neither of the last two rows is a scalar multiple of the other. Thus the matrix of coefficients has exactly $2n+1$ linearly independent rows and hence has rank $2n+1$. By the rank and nullity theorem [1], the dimension of 0MS(n), which is the nullity of the coefficient matrix, is $n^2-(2n+1)$.

COROLLARY. *The dimension of MS*(n) *is* $n^2-2n=n(n-2)$.

Proof. Let $q=n^2-2n-1$ and $S_1,\ldots,S_q$ be a basis for 0MS(n), a subspace of MS(n). Let I be the magic square in MS(n) with 1 in every position, and consider the set $B=\{S_1,\ldots,S_q,I\}$ consisting of n^2-2n vectors of MS(n).

The set B spans MS(n), for if M is any magic square in MS(n) and M has line-sum m, then M is equivalent to the zero magic square $M_0=M-(m/n)I$ of 0MS(n). As $S_1,\ldots,S_q$ is a basis for 0MS(n), $M_0=c_1S_1+\cdots+c_qS_q$ for some scalars $c_1,\ldots,c_q$, so $M=c_1S_1+\cdots+c_qS_q+(m/n)I$. Moreover, B is linearly independent, for the line-sum of the vector $c_1S_1+\cdots+c_qS_q+c_{q+1}I$, where the c_i's are scalars, is nc_{q+1}. [See parts (i) and (ii) of the opening exercise.] If this vector is to equal the zero vector, which has line-sum zero, we must have $nc_{q+1}=0$ or $c_{q+1}=0$. Then the linear independence of $S_1,\ldots,S_q$ implies that $c_1=\cdots=c_q=0$ as well. Thus B is a basis for MS(n).

It is easy to see that the central entry of any magic square in 0MS(3) is zero. This means that 0MS(3) magic squares are anti-symmetric with respect to the diagonals. By the Theorem, the dimension of 0MS(3) is 2; thus a magic square in 0MS(3) is uniquely determined by specifying any two entries not collinear with the central zero. If we choose the first two entries in the first row to be 1, 0 and 0, 1, we get the following basis for 0MS(3):

$$\begin{matrix} 1 & 0 & -1 \\ -2 & 0 & 2 \\ 1 & 0 & -1 \end{matrix} \qquad \begin{matrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{matrix}$$

According to the Theorem, the dimension of 0MS(4) is 7. Using an argument of the same nature as that which shows that there is a unique 3×3 classical magic square up to symmetry, it can be established that the sum of the four corner entries and the sum of the four central entries of a magic square in 0MS(4) are both zero. With these facts, it is easy to find seven entries which, when specified, completely determine a 4×4 zero magic square. Two examples are:

$$\begin{matrix} x & x & x & - \\ x & x & x & - \\ x & - & - & - \\ - & - & - & - \end{matrix} \qquad \begin{matrix} - & - & - & - \\ - & - & x & - \\ x & - & x & x \\ x & - & x & x \end{matrix}$$

The seven squares obtained by putting 1 in one of the designated positions of either pattern and 0's in the other six in all possible ways constitute a basis for 0MS(4). For instance, the seven magic squares which form a basis for 0MS(4) according to the first pattern are:

$$\begin{matrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 2 & -2 & 0 \\ -1 & -2 & 2 & 1 \end{matrix} \qquad \begin{matrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -2 & 1 & 1 \end{matrix} \qquad \begin{matrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 0 & 1 \end{matrix} \qquad \begin{matrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ -1 & -1 & 1 & 1 \end{matrix}$$

$$\begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & -1 & 1 \\ 0 & -1 & 1 & 0 \end{matrix} \qquad \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{matrix} \qquad \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \end{matrix}$$

Motivated by this example, it is natural to make the following definition. A selection of n^2-2n-1 positions in an $n\times n$ matrix is called a **skeleton** of 0MS(n) if the assignment of real numbers to those positions uniquely determines entries in all other positions using the zero magic square conditions. Given a skeleton of 0MS(n), the array of the $2n+1$ positions *not* specified is called the **frame** of that skeleton. In the 0MS(4) examples above, each array of x's is a skeleton and each array of dashes is a frame. When n is large, the number of positions in a skeleton is much greater than the number of positions in its frame, so it is convenient to think of skeletons in terms of the frames they determine.

Every skeleton of 0MS(n) leads to a basis of 0MS(n) in a natural way, by assigning 1 to one skeletal position and 0's to the rest in all n^2-2n-1 possible ways. We shall call this basis the **natural basis** associated with that skeleton. Thus we could determine a canonical basis for 0MS(n) if we could agree on a canonical skeleton of 0MS(n). Unfortunately, there does not appear to be any one skeleton of 0MS(n) which is superior to, or more natural than, all the others. Some skeletons possess certain kinds of symmetry or near-symmetry, while others guarantee the presence of a large number of zeroes in the magic squares of the natural bases they determine. Preference for one skeleton over another seems to be largely a matter of taste.

The preceding ideas can be used to determine a crude upper bound on the number of $n\times n$ classical magic squares. Since the sum of the first n^2 integers is $n^2(n^2+1)/2$, each line-sum of an $n\times n$ classical magic square must be $n(n^2+1)/2$. Letting $l=n(n^2+1)/2$, an $n\times n$ classical magic square can be zeroed by subtracting l/n from each entry. A skeleton of this $n\times n$ zero magic square consists of n^2-2n-1 positions and determines the zero magic square, and hence the classical magic square, completely. Thus the number of $n\times n$ classical magic squares is the number of ways the n^2-2n-1 positions of this skeleton can be chosen from the n^2 numbers $1-l/n, 2-l/n, \ldots, n^2-l/n$, choosing each number no more than once. Since the number of permutations of n^2-2n-1 elements which can be formed from a set of n^2 elements is $(n^2)!/(2n+1)!$, the maximum number of $n\times n$ classical magic squares, taking into account the 8 symmetries of a magic square, is $(n^2)!/8(2n+1)!$. The imprecision of this bound is revealed even when $n=3$: in that case, the bound says that there are at most nine 3×3 classical magic squares, while, as suggested earlier in this paper, it is easy to show that there is, in fact, only one.

References

[1] H. Anton, Elementary Linear Algebra, 2nd ed., Wiley, New York, 1977.

[2] F. A. P. Barnard, Theory of magic squares and of magic cubes, Memoirs of the National Academy of Sciences, vol. 4, Part 1, 1888, pp. 209–270.

[3] T. J. Fletcher, Linear Algebra Through Its Applications, Van Nostrand Reinhold, New York, 1972.

[4] Hutton, Mathematical Recreations, 1844.

[5] H. M. Stark, Introduction to Number Theory, Markham, Chicago, 1970.

PROBLEMS

DAN EUSTICE, Editor
LEROY F. MEYERS, Associate Editor
The Ohio State University

Proposals

To be considered for publication, solutions should be mailed before November 1, 1980.

1093. Prove that for complex numbers u, v, and w,

$$|u-v|+\big|u+v-2w+|u-v|\big|<|u+v| \tag{1}$$

if and only if

$$|w-v|+\big|w+v-2u+|w-v|\big|<|w+v|. \tag{2}$$

[*M. S. Klamkin, University of Alberta.*]

1094. For each positive integer n, show that there is a solution to the congruence

$$32k^2+21k+14=0 \pmod{2^n}.$$

[*Tom Moore, Bridgewater State College.*]

1095*. It is mathematical folklore that the fundamental period of a linear combination (with nonzero coefficients) of simple sine and cosine functions having commensurable fundamental periods is the least common multiple of the fundamental periods of the separate functions. Is there a noncalculus proof (or disproof) of this? (The *fundamental period* of a periodic function from R to R is the smallest positive period. The function f is a *simple* sine or cosine function if $f(x)=\sin\alpha x$ for all x, or $f(x)=\cos\alpha x$ for all x, for some positive constant α.) [*S. Yeshurun, Bar-Ilan University, Ramat-Gan, Israel.*]

ASSISTANT EDITORS: DON BONAR, *Denison University;* WILLIAM A. MCWORTER, JR., *The Ohio State University. We invite readers to submit problems believed to be new. Proposals should be accompanied by solutions, when available, and by any information that will assist the editors. Solutions to published problems should be submitted on separate, signed sheets. An asterisk (*) will be placed by a problem to indicate that the proposer did not supply a solution. A problem submitted as a Quickie should be one that has an unexpected succinct solution. Readers desiring acknowledgment of their communications should include a self-addressed stamped card. Send all communications to this department to Dan Eustice, The Ohio State University, 231 W. 18th Ave., Columbus, Ohio 43210.*

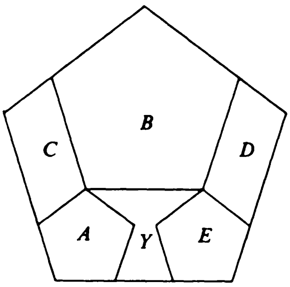
Solutions

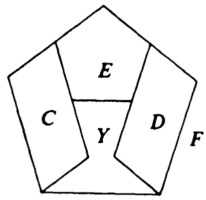
Heximating a Pentagon **November 1978**

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

Solutions:

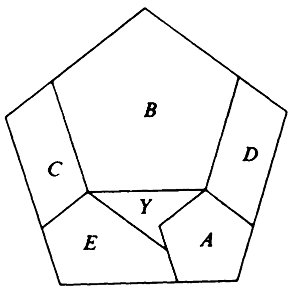
Solution is pentagons *A*, *B*, and *F*.

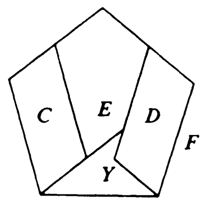




J. M. STARK
Lamar University

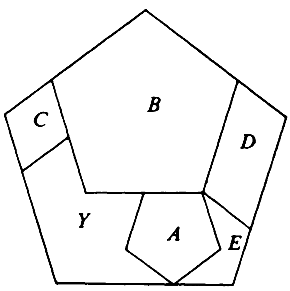
Solution is pentagons *A*, *B*, and *F*.

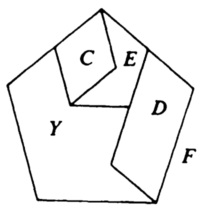




EDUARDO VALDEZ, student
Casa Grande Union High School
Casa Grande, Arizona

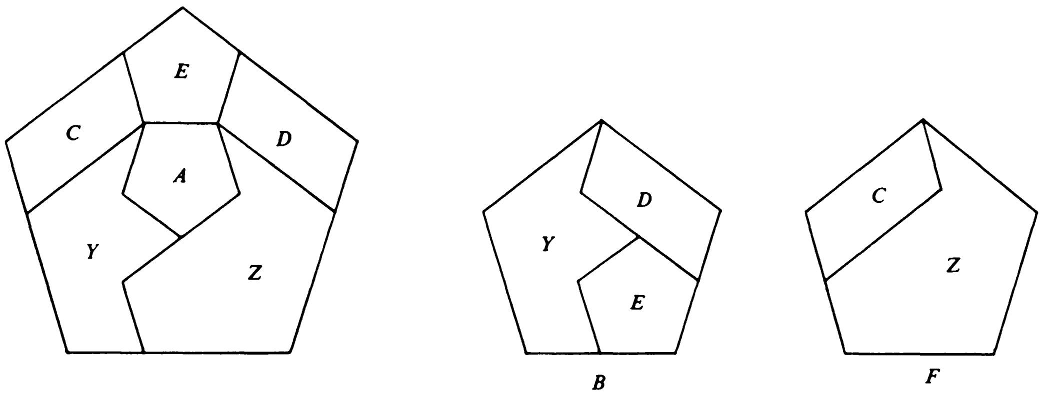
Solution is pentagons *A*, *B*, and *F*.





WALTER BLUGER
Ottawa, Canada

Solution is pentagons A, B, and F.



D. M. Collison
Anaheim, California

Also solved by Robert D. Baer and Lyn Ratcliff. Collison also found Stark's solution. Howard Eves and Michael Goldberg found seven-piece solutions.

A True Similarity **January 1979**

1058. Is it true that a square matrix that is not a scalar multiple of the identity is always similar to a matrix with all nonzero elements? [*H. Kestelman, University College, London.*]

Solution: True. Suppose that among the matrices similar to A, B is one with a maximal number of nonzeros. B cannot be diagonal since it would have by hypothesis to have two distinct diagonal elements, and evidently

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \text{ and } \begin{pmatrix} a & a-b \\ 0 & b \end{pmatrix}$$

are similar and the second has fewer zeros than the first if $a \neq b$.

Suppose then that b_{pq} is an off-diagonal nonzero element of B; it is enough to show that all elements in the qth column of B are nonzero. Suppose $r \neq p$ and E_{rp} is the matrix with 1 in the (r,p)th place and 0 elsewhere. Then

$$(I + tE_{rp})(I - tE_{rp}) = I,$$

and if we set

$$M_t = (I + tE_{rp})B(I - tE_{rp}),$$

M_t is similar to B for all numbers t. Its (r,q)th element differs from that of

$$B + t(E_{rp}B - BE_{rp})$$

by an amount that is $o(t)$ when t is small. The (r,q)th element of $E_{rp}B - BE_{rp}$ is easily seen to be b_{pq} (since $p \neq q$). So, if $b_{rq}=0$ for some r, the (r,q)th element of M_t is nonzero and at the same time to every nonzero element of B corresponds a nonzero of M_t provided t is small enough; but this would contradict the maximality of B.

H. KESTELMAN
University College, London

Also solved by J. M. Stark.

Cyclic Extrema **January 1979**

1059. How should n given non-negative real numbers be indexed to minimize (maximize) $a_1a_2 + a_2a_3 + \cdots + a_{n-1}a_n + a_na_1$? [*David E. Daykin, Reading University.*]

Solution: Firstly, there is a minimal value of $C = \sum_{i=1}^{n} a_ia_{i+1}$ since there is only a finite number of cyclic arrangements. Now consider any two nonoverlapping adjacent pairs in a cyclic arrangement, say

$$\ldots, a, b, \ldots, c, d, \ldots.$$

If we reverse the section between b and c inclusive, we preserve all but two of the adjacent products a_ia_{i+1}. Thus we have a net gain for C of $ac + bd - ab - cd = (a-d)(c-b)$. For a minimal arrangement, it is necessary that this gain must be $\geqslant 0$, which means that $a = d$ or $b = c$ or ($a > d$ iff $b < c$). Now let us arrange our given numbers in ascending order for convenience in referring to them: $a_1 \leqslant a_2 \leqslant a_3 \leqslant \cdots \leqslant a_n$. Now a_1 and a_n must occur somewhere in a minimal arrangement, say

$$\ldots, a_1, b, \ldots, a_n, d, \ldots.$$

If $a_1 \neq d$ and $b \neq a_n$, then $a_1 < d$ and $b < a_n$, contrary to our necessary condition. Thus the maximum and minimum values of the a_i's must be adjacent. Renumbering equal values if necessary and choosing one of two symmetric arrangements, we must have the following pattern:

$$\ldots, a_n, a_1, b, \ldots, a_{n-1}, d, \ldots.$$

If $a_1 = d$ or $b = a_{n-1}$, we can reverse the section from b to a_{n-1} inclusive without changing C. If $a_1 \neq d$ and $b \neq a_{n-1}$, then $a_1 < d$ and $b < a_{n-1}$, contrary to the necessary condition. So, by reversal of a section if necessary, we can assume a minimal arrangement has the following pattern:

$$\ldots, a, a_n, a_1, a_{n-1}, \ldots, c, a_2, \ldots.$$

By the same argument, we must have $a = a_2$ or $a_n = c$ and, by reversal of a section if necessary, we can assume a minimal arrangement must have the following pattern:

$$\ldots, a_2, a_n, a_1, a_{n-1}, b, \ldots, a_3, c, \ldots.$$

Repeating the argument shows that we can assume $b = a_3$, by reversing if necessary. Continuing, we see that, up to reversal between equal values and cyclic shifts and reversals, there is a unique minimal arrangement:

$$\ldots, a_4, a_{n-2}, a_2, a_n, a_1, a_{n-1}, a_3, a_{n-3}, a_5, \ldots.$$

If the a_i are distinct, then this pattern is unique up to cyclic shifts and reversals. It is also unique in some other cases, e.g., for $a_1 = a_2 < a_3 < \cdots < a_n$. The smallest case where nonuniqueness of the minimal pattern can occur seems to be when $n = 7$, e.g.,

```
     1             1
  6     5       6     5
   2   2   and   2   2
    4 3           3 4
```

The argument is readily adapted to the maximization problem and shows that the unique maximizing arrangement, up to the same processes, is:

$$\ldots, a_9, a_7, a_5, a_3, a_1, a_2, a_4, a_6, a_8, \ldots.$$

DAVID SINGMASTER
The Open University

Also solved by Walter Bluger (Canada), Duane Broline, Michael Ecker, Thomas E. Elsner, Michael Goldberg, Robert Patenaude, and the proposer.

No Such Function **January 1979**

1060. Prove or disprove: There exists a function f defined on $[-1,1]$ with f'' continuous such that $\sum_{n=1}^{\infty} f(1/n)$ converges but $\sum_{n=1}^{\infty} |f(1/n)|$ diverges. [*Peter Ørno, The Ohio State University.*]

Solution: Assume that such a function f does exist. By the extended mean value theorem, there is a c_n such that $0<c_n<1/n$ and

$$f\left(\frac{1}{n}\right)=f(0)+\frac{f'(0)}{n}+\frac{f''(c_n)}{2n^2}.$$

Since $\sum f(1/n)$ converges, $f(1/n) \to 0$ as $n \to \infty$ and by the continuity of f, $f(0)=0$. Since f'' is continuous, it is bounded, say $|f''(x)| \leqslant M$. Therefore $\sum |f''(c_n)/2n^2| \leqslant \sum M/2n^2$ so that both $\sum |f''(c_n)/2n^2|$ and $\sum f''(c_n)/2n^2$ converge. But if $f'(0)$ were nonzero, then $f'(0)\sum 1/n$ would diverge and $\sum f(1/n) = \sum((f'(0)/n)+(f''(c_n)/2n^2))$ would diverge. Hence, $f'(0)=0$ and $\sum |f(1/n)| = \sum |f''(c_n)/2n^2|$ converges, contradiction.

It is clear from the above proof that the condition "f'' is continuous on $[-1,1]$" could be replaced with "f'' exists and is bounded on some interval $[0,a]$."

ROBERT CLARK, student
Temple University

Also solved by Roger Cuculière (France), Michael J. Dixon, Michael W. Ecker, G. A. Heuer, Eli L. Isaacson, Michael Raship, Adam Riese, J. M. Stark, and the proposer.

Minmax Equals Maxmin **January 1979**

1061. In how many ways can n^2 distinct real numbers be arranged into an $n \times n$ array (a_{ij}) such that $\max_j \min_i a_{ij} = \min_i \max_j a_{ij}$? [*Edward T. H. Wang, Wilfrid Laurier University.*]

Solution: If $\max_j \min_i a_{ij} = \min_i \max_j a_{ij} = a_{\alpha\beta}$, then clearly $a_{\alpha\beta}$ is at once the largest number in the αth row and the smallest number in the βth column, and hence

$$a_{\alpha j} < a_{\alpha\beta} < a_{i\beta} \text{ for all } i \neq \alpha \text{ and for all } j \neq \beta. \qquad (*)$$

Conversely, if (*) holds for some $a_{\alpha\beta}$, then $\min_i a_{ij} \leqslant a_{\alpha j} < a_{\alpha\beta}$ for all $j \neq \beta$ and $\max_j a_{ij} \geqslant a_{i\beta} > a_{\alpha\beta}$ for all $i \neq \alpha$ would imply that $\max_j \min_i a_{ij} = \min_i \max_j a_{ij}$. To obtain a required configuration, it is therefore necessary and sufficient to choose any $2n-1$ of the given n^2 numbers, say $x_1 < x_2 < \cdots < x_{2n-1}$. Put x_n anywhere in the array. Then put $x_1, x_2, \ldots, x_{n-1}$ in the same row as x_n and put $x_{n+1}, x_{n+2}, \ldots, x_{2n-1}$ in the same column as x_n. The remaining n^2-2n+1 numbers can then be used to fill up the remaining n^2-2n+1 positions. Therefore, the total number of such configurations is

$$\binom{n^2}{2n-1} \cdot n^2 \cdot [(n-1)!]^2 \cdot (n^2-2n+1)! = \frac{(n^2)!(n!)^2}{(2n-1)!}.$$

EDWARD T. H. WANG
Wilfrid Laurier University

Also solved by Thomas E. Elsner and Jinku Lee.

1062. (a) Let (x_1,y_1), (x_2,y_2), and (x_3,y_3) be three points in the Cartesian plane. Assume the points and their negatives are all distinct. Show that there is an ellipse, centered at the origin, passing through the three points if and only if

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} \cdot \begin{vmatrix} x_1 & y_1 & -1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} \cdot \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & -1 \\ x_3 & y_3 & 1 \end{vmatrix} \cdot \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & -1 \end{vmatrix} > 0.$$

Interpret this condition geometrically.

(b*) Find a necessary and sufficient condition for the existence of an ellipsoid, centered at the origin, passing through for given points in 3-space. [*G. A. Edgar, The Ohio State University.*]

Solution: (a) The geometric interpretation is: the three points and their negatives are the vertices of a convex hexagon. The proof is organized as (i)⇔(ii)⇔(iii):

(i) There is an ellipse through the six points.

(ii) The six points are the vertices of a convex hexagon.

(iii) $$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} \begin{vmatrix} x_1 & y_1 & -1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & -1 \\ x_3 & y_3 & 1 \end{vmatrix} \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & -1 \end{vmatrix} > 0.$$

(i)⇒(ii) is trivial.

(ii)⇒(i). Assume the six points are the vertices of a convex hexagon. Since the truth or falsity of (i) or (ii) is unchanged by a linear change of coordinates, we may assume without loss of generality that the points $(x_1,y_1),(x_2,y_2),-(x_1,y_1),-(x_2,y_2)$ are $(1,1),(1,-1),(-1,-1),(-1,1)$ in some order and (x_3,y_3) satisfies $x_3>1$, $-1<y_3<1$. The expression $ax_3^2+(1-a)y_3^2$ is less than 1 when $a=0$ and greater than 1 when $a=1$, so it is exactly 1 for some a between 0 and 1. This means that the ellipse $ax^2+(1-a)y^2=1$ passes through all six points.

(ii)⇒(iii). Assume the six points are the vertices of a convex hexagon. Note that the truth or falsity of (ii) or (iii) is unchanged if two of the points $(x_1,y_1),(x_2,y_2),(x_3,y_3)$ are interchanged or if one of the points is replaced by its negative. Therefore we may assume without loss of generality that the vertices of the hexagon are, in counterclockwise order, (x_1,y_1), $-(x_3,y_3)$, (x_2,y_2), $-(x_1,y_1),(x_3,y_3)$, $-(x_2,y_2)$. In this case, all four of the determinants are positive.

(iii)⇒(ii). Conversely, again by interchanging the points, we may assume that all four of the determinants are positive. I claim that the other four points are all on the same side of the line l joining (x_1,y_1) and $-(x_2,y_2)$: certainly the points $-(x_1,y_1),(x_2,y_2)$ are on the same side of l as the origin, and since the second and third determinants are positive, (x_3,y_3) and $-(x_3,y_3)$ are also on the same side of l as the origin. Similar reasoning applies to the line through $-(x_2,y_2)$ and (x_3,y_3) and to the line through (x_3,y_3) and $-(x_1,y_1)$. Thus we have a convex hexagon when the vertices are taken in the order (x_1,y_1), $-(x_3,y_3),(x_2,y_2)$, $-(x_1,y_1),(x_3,y_3)$, $-(x_2,y_2)$.

Comment on (b): The analogous geometric criterion is: the four points and their negatives are the vertices of a convex polyhedron (with positive volume). I do not know whether any four points with this property lie on a centered ellipsoid.

G. A. Edgar
The Ohio State University

J. M. Stark provided an interpretation of Part (a) *in terms of oriented triangles and for Part* (b) *he formulated the problem in terms of the existence of a solution to a system of equations.*

REVIEWS

Assistant Editor: Eric S. Rosenthal, Princeton University. Articles and books are selected for this section to call attention to interesting mathematical exposition that occurs outside the mainstream of the mathematics literature. Some reviews of books are adapted from the Telegraphic Reviews in the American Mathematical Monthly.

Chow, William W., *Interlocking shapes in art and engineering*, Computer Aided Design 12 (January 1980) 29-34.

Discussion of a computer program to generate Escher-type tessellations, and application of such tessellations to sheet metal cutting and manufacturing.

Brams, Steven, Jr., et al., *The geometry of the arms race*, International Studies Quarterly 23 (1979) 567-588; comment and reply, 589-600.

The authors investigate the arms race as a sequence of moves in a Prisoner's Dilemma game. Each player is assumed to have a specified probability of detecting whether the other side is cooperating or not, and both players are presumed to cooperate only if they detect cooperative behavior by the other. One finding: making the arms race more costly (e.g., by heavy investment in research and development) provides greater incentive for cooperation than increasing the benefits of an arms control agreement.

Moore, Gregory H., *The origins of Zermelo's axiomatization of set theory*, J. Philosophical Logic 7 (1978) 307-329.

"What gave rise to Ernst Zermelo's axiomatization of set theory in 1908? According to the usual interpretation, Zermelo was motivated by the set-theoretic paradoxes. This paper argues that Zermelo was primarily motivated, not by the paradoxes, but by the controversy surrounding his 1904 proof that every set can be well-ordered, and especially by a desire to preserve his axiom of choice from its numerous critics."

Feldman, Jerome A., *Programming languages*, Scientific American 241 (December 1979) 94-116.

A survey of the various high level programming languages, illustrating how languages relate to algorithms via such important intermediate concepts as data structures, arrays, iteration, linked lists, and recursion. Feldman aims to show that programming languages are not just fixed tools, but an essential and evolving part of the general research effort to improve computer algorithms.

Wenninger, Magnus J., Spherical Models, Cambridge U Pr, 1979; xii + 147 pp, $19.95, $7.95 (P).

Similar in style and excellence to the author's *Polyhedron Models* (1971), this volume treats spherical polyhedra and tessellations, including geodesic domes.

Directions are included for building models; construction of one may take from 1 to 100 hours.

Priest, Graham, *The logic of paradox*, J. Philosophical Logic 8 (1979) 219-241.

"The purpose of the present paper is to suggest a new way of handling the logical paradoxes. Instead of trying to dissolve them or explain what has gone wrong, we should accept them and learn to come to live with them." The "concluding self-referential postscript" provides a surprise ending.

Baxandall, P.R., et al., Proof in Mathematics ("if," "then," and "perhaps"), Institute of Education, University of Keele (Staffordshire, England ST5 5BG), 1978; v + 130 pp, (P).

A post-calculus handbook on proofs, packed with examples, advice, counterexamples, and notes on strategy.

Mueller, Dennis C., Public Choice, Cambridge U Pr, 1979; xiii + 297 pp.

Excellent review of the major contributions to social choice literature. The author melds together mathematical theorems, empirical results, and expert opinions, to give an account from the point of view of an economist--e.g., he never lets the reader lose sight of the fact that any political decision rule is also capable of redistributing resources.

Maggs, Peter B., *Programming strategies in the game of Reversi*, Byte 4:11 (November 1979) 66-79.

Reversi, a game dating to the 1880's, is undergoing revival under the trademark Othello. This article describes all steps in programming a computer in Basic to play the game, including the use of alpha-beta pruning of the game tree.

Whitney, Craig R. and Browne, Malcom W., *Soviet mathematician is obscure no more*, The New York Times, 27 November 1979, C1,2.

Background color on Leonid Khachiyan, the 27-year-old recent discoverer of a polynomial-time algorithm for linear programming.

National Research Council, The Role of Applications in the Undergraduate Mathematics Curriculum, NRC (2101 Constitution Ave., Washington, D.C. 20418), 1979; ix + 25 pp.

Report of the Ad Hoc Committee on Applied Mathematics Training, chaired by Peter Hilton (Case-Western). Recommendations fall into five themes: change in attitudes of mathematicians, a new integration of the mathematical sciences, curricular change, establishment of program giving experience in applications, and relations with society.

Stoutemyer, David R., *LISP based symbolic math systems*, Byte 4:8 (August 1979) 176-192.

Computer symbolic math systems are often called computer algebra systems, because they accept and transform input consisting of algebraic expressions containing variables. The article describes four such systems, noting their ability to do indefinite precision arithmetic (e.g., find 99^{99} exactly), trigonometric simplification (e.g., express sin(17x) in terms of sin x and cos x), and calculus (e.g., differentiate ln(cos x)). The widespread availability of symbolic math systems will affect research and education in mathematics, perhaps even more profoundly than the pocket calculator has.

NEWS & LETTERS

ICME IV: THE INTERNATIONAL CONGRESS OF MATHEMATICAL EDUCATION

The Fourth International Congress on Mathematical Education (ICME IV) will meet in Berkeley, California, August 10-16, 1980. Approximately 3,000 mathematicians and mathematics educators from around the world are expected to attend.

ICME IV will feature four plenary addresses:

Hans Freudenthal, Instituut Ontwikkeling Wiskunde Onderwijs, The Netherlands. Topic: *Major Problems of Mathematics Education.*

Hua Loo-Keng, Academy of Sciences, People's Republic of China. Topic: *Applications of Mathematics and Mathematics Education.*

Seymour Pappert, The Massachusetts Institute of Technology, United States. Topic: *Technology and Mathematics Education.*

Hermine Sinclair, University of Geneva, Switzerland. Topic: *Language and Mathematics Learning.*

The program of the Congress will be organized around several broad themes, including universal education, language and mathematics, the role of applications in the curriculum, the impact of technology on mathematics education, assessment, and research in mathematics education. There will also be a variety of special workshops and social events throughout the Congress.

Conference registration will be $115, and the cost of meals and lodging in University of California dormitories will be $105 for the week. Registration forms and further information may be obtained by writing to:

ICME IV
Mathematics Department
University of California
Berkeley, CA 94720

A TRIMMER MAGAZINE

Observant readers may have noticed that the January issue of *Mathematics Magazine* is 1/4 inch narrower than usual. This is a consequence of the shift of these M.A.A. journals to a new printer, Capital City Press in Monteplier, VT, who uses a different size paper.

UNDERGRADUATE RESEARCH PROGRAMS

The National Science Foundation announced recently 126 grants for Undergraduate Research Participation (URP) for the summer of 1980; four of these programs are in mathematics. Since most of these programs accept applications from students at other institutions, we list below the project director and address for each of these four programs. Numbers in the left margin indicate the number of available stipends: students may earn up to $1000 for summer research in URP programs. Students interested in applying for these programs should communicate directly with the project directors.

7 Dr. K.L. de Bouvere
Department of Mathematics
University of Santa Clara
Santa Clara
California 95053

4 Dr. Joan P. Hutchinson
Department of Mathematics
Smith College
Northampton
Massachusetts 01063

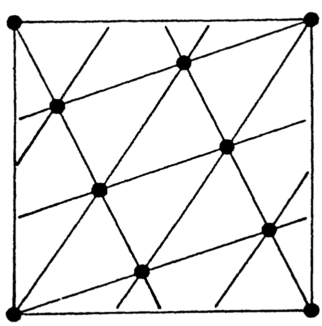
6 Dr. A. Wayne Roberts
Department of Mathematics
Macalester College
St. Paul
Minnesota 55105

3 Dr. Joseph A. Gallian
Department of Mathematical Sciences
University of Minnesota, Duluth
Duluth
Minnesota 55812

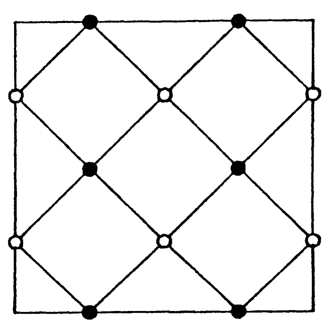
UTILITY COMMENTS

The equivalence claimed by David E. Kullman (in "The Utilities Problem" this *Magazine*, November 1979, pp. 299-302) of the utilities problem and the two urn problem is false. When the urn handles have been relabeled as directed, we find that instead of having three utilities to be joined to three houses, now the goal is to join the Electric utility to all three houses, but Water and Gas are to be joined to only two houses each, plus each other! Water is not to be joined to house A, and Gas is not to be joined to house B.

We may view the three handles of each urn to be joined as triangles lying within the urns. The problem is to add three more edges to form the triangular prism $K_3 \times K_2$. (This graph can also be viewed as the complement of the cycle C_6.) If no labeling is given or if a consistent labeling is given for the triangles, the task is trivial. Only when an inconsistent labeling is imposed on the triangles (that is, both are labeled clockwise or both counterclockwise) does the problem become insoluble.



K_7 on a Torus



$K_{4,4}$ on a Torus

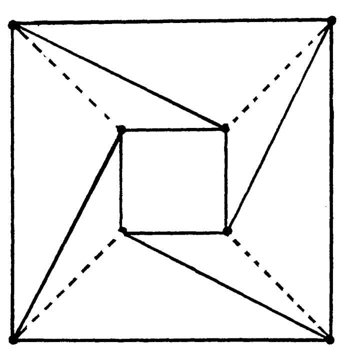
Figure 1

The common thread to all three problems (the utilities, the urns, and the attempt to embed K_5) is that impossibility follows easily from Euler's polyhedron formula $V-E+F=2$. The "proofs" presented for K_5 and $K_{3,3}$ are essentially correct, but are weak on the point of assuming all faces have the same number of sides. These properties could be developed for the two problems considered, but for similar questions such an assumption may be fatal. The proofs are better phrased as follows:

K_5 has $V=5$ and $E=10$, so if it can be embedded, $F=7$. Each face has at least three edges, and every edge bounds two faces, so adding the number of edges about each face yields $2Q=2E \geq 3F = 21$. The contradiction implies no embedding is possible.

Similarly, $K_{3,3}$ has $V=6$ and $E=9$, so if it can be embedded, $F=5$. Since $K_{3,3}$ has no triangles, each face must have at least four edges, and so $18=2E \geq 4F=20$ and the embedding is impossible.

For $K_3 \times K_2$, we have $V=6$ and $E=9$, so $F=5$. Two faces are triangles, and the other three have at least four edges but then $18 = 2E \geq 2\cdot3 + 3\cdot4 = 18$. This is valid, but requires each of the last three faces to be four-cycles. As soon as handle 1 is joined to handle 1, the four-cycle condition assures a consistent labeling. When an inconsistent labeling is imposed on the triangles,



As a graph in the plane this figure contains eight triangular and two quadrilateral faces, but as a square torus, dotted lines underneath, it has eight quadrilateral faces.

Figure 2

the remaining faces are forced to be five-cycles, hence the impossibility.

In closing, if we wish to marvel at the improved conditions on the torus, why not go all the way? Both K_7 and $K_{4,4}$ can be embedded on a torus as shown in Figure 1. Of course, each rectangle is to be made into a torus by identifying parallel sides.

Allen J. Schwenk
U.S. Naval Academy
Annapolis
Maryland 21402

A graph is an abstract object made of vertices and edges, not embedded anywhere, and having no faces. There is no way to see faces correctly in a figure of the graph. To demonstrate this fact I offer the graph shown in Figure 2. On one hand, it is drawn in the plane, with 8 triangular faces and 2 quadrilateral ones. On the other hand, it can also be drawn on a torus with 8 quadrilateral faces. A correct proof must begin by assuming that the graph has been embedded in the plane.

Frederic Cunningham, Jr.
Bryn Mawr College
Bryn Mawr
Pennsylvania 19010

SOLUTIONS TO THE 1979 PUTNAM EXAM

In January we printed in this column questions from the 1979 Putnam Examination. To assist those who have been puzzling over these problems, we provide here hints and answers. The official report on the results of the competition, including names of winners and complete sample solutions, will be published later this year in the American Mathematical Monthly.

These answers and hints are adapted for publication in Mathematics Magazine by Loren Larson from those supplied by the Putnam Committee. The solution to Problem A-4 was submitted by Philip Straffin of Beloit College.

A-1. Find positive integers n and $a_1, a_2, \ldots, a_n$ such that

$$a_1 + a_2 + \ldots + a_n = 1979$$

and the product $a_1a_2\ldots a_n$ is as large as possible.

Sol. The largest such product can have no a_i greater than 4 since we could increase the product by replacing 5 by $2\cdot3$, 6 by $3\cdot3$, 7 by $4\cdot3$, etc. Also, we could increase the product by replacing $2\cdot4$ by $3\cdot3$ and $2\cdot2\cdot2$ by $3\cdot3$. Combining these observations with the fact that $1979 = 3\cdot659 + 2$ it follows that $n = 660$, that all but one of the a_i equals 3, and the exceptional a_i equals 2.

A-2. Establish necessary and sufficient conditions on the constant k for the existence of a continuous real valued function $f(x)$ satisfying $f(f(x)) = kx^9$ for all real x.

Sol. The condition is $k \geq 0$. If $k \geq 0$, define $f(x) = k^{1/4}x^3$. Then $f(f(x)) = kx^9$. For the converse, note that $k \neq 0$ and $f(f(x)) = kx^9$ imply that f is one-to-one and onto the real numbers. Then, because f is continuous, f is strictly monotonic. It follows that $f(f(x))$ is increasing, and therefore $k \geq 0$.

A-3. Let $x_1, x_2, x_3, \ldots$ be a sequence of nonzero real numbers satisfying

$$x_n = \frac{x_{n-2}x_{n-1}}{2x_{n-2} - x_{n-1}} \text{ for } n = 3, 4, 5, \ldots .$$

Establish necessary and sufficient conditions on x_1 and x_2 for x_n to be an integer for infinitely many values of n.

Sol. By induction,

$$x_n = \frac{x_1x_2}{(n-1)x_1 - (n-2)x_2} .$$

Substituting $x_2 = ax_1$, we find that x_n is an integer for an infinite number of n if and only if $a = 1$ and x_1 is an integer, or equivalently, if and only if $x_1 = x_2 = m$ for some integer m.

A-4. Let A be a set of $2n$ points in the plane, no three of which are collinear. Suppose that n of them are colored red and the remaining n blue. Prove or disprove: there are n closed straight line segments, no two with a point in common, such that the endpoints of each segment are points of A having different colors.

Sol. If we disregard line crossings, there are a number of ways that the given n red points can be paired with the given blue points by n closed straight line segments. Assign to each such pairing the total length of all the line segments in the configuration. Because there are only a finite number of such pairings, one of these configurations will have minimal total length. This pairing will have no segment crossings. (If segments R_1B_1 and R_2B_2 intersect (R_1, R_2 red points, B_1, B_2 blue points), we could reduce the total length of the configuration by replacing these segments with R_1B_2 and R_2B_1.)

A-5. Denote by $[x]$ the greatest integer less than or equal to x and by $S(x)$ the sequence $[x], [2x], [3x], \ldots$. Prove that there are distinct real solutions α and β of the equation $x^3 - 10x^2 + 29x - 25 = 0$ such that infinitely many positive integers appear both in $S(\alpha)$ and in $S(\beta)$.

Sol. The equation $x^3 - 10x^2 + 29x - 25 = 0$ has three real solutions a, b, c with $1 < a < 2$, $2 < b < 3$, $5 < c < 6$. The number of integers that the set $\{1,2,\ldots,n\}$ has in common with $S(a)$, $S(b)$, and $S(c)$ is $[n/a]$, $[n/b]$, and $[n/c]$, respectively. Since

$$\frac{1}{a} + \frac{1}{b} + \frac{1}{c} > \frac{1}{2} + \frac{1}{3} + \frac{1}{6} = 1$$

we see that

$$\lim_{n\to\infty}\{[\tfrac{n}{a}] + [\tfrac{n}{b}] + [\tfrac{n}{c}] - n\} = \infty$$

and hence that an infinite number of positive integers appear in more than one of $S(a)$, $S(b)$, $S(c)$.

A-6. Let $0 \le p_i \le 1$ for $i = 1,2,\ldots,n$. Show that

$$\sum_{i=1}^{n} \left|\frac{1}{x-p_i}\right| \le 8n\left(1 + \frac{1}{3} + \frac{1}{5} + \ldots + \frac{1}{2n-1}\right)$$

for some x satisfying $0 \le x \le 1$.

Sol. For $k = 0,1,\ldots,2n-1$, let I_k be the open interval $(k/2n, (k+1)/2n)$. Among the $2n$ intervals I_k there exist n not containing any of the p_i; place an x_j at the center of each of these n intervals. Let $|x_j - p_i| = d_{ij}$ and

$$B = 8n\left(1 + \frac{1}{3} + \frac{1}{5} + \ldots + \frac{1}{2n-1}\right) .$$

For fixed i, the d_{ij} satisfy $d_{ij} \ge 1/4n$, at most two of them do not satisfy $d_{ij} \ge 3/4n$, at most four do not satisfy $d_{ij} \ge 5/4n$, etc. Hence

$$\sum_{j=1}^{n} \frac{1}{d_{ij}} \le 2\sum_{h=0}^{n-1} \frac{4n}{1+2h} = B .$$

Thus we have

$$\sum_{j=1}^{n} \sum_{i=1}^{n} \frac{1}{d_{ij}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{d_{ij}} \le nB .$$

So clearly there is a value of j for which $\sum_{i=1}^{n}(1/d_{ij}) \le B$ and the x_j for such a j can serve as the desired x.

B-1. Prove or disprove: there is at least one straight line normal to the graph of $y = \cosh x$ at a point $(a, \cosh a)$ and also normal to the graph of $y = \sinh x$ at a point $(c, \sinh c)$.

Sol. Assume there is a common normal at $(a, \cosh a)$ and $(c, \sinh c)$. This assumption implies that $\cosh c = \sinh a$ (examine the derivatives). At the same time, an easy argument shows that $0 < a < c$ and this implies that

$$\sinh a < \cosh a < \cosh c ,$$

a contradiction.

B-2. Let $0 < a < b$. Evaluate

$$\lim_{t\to 0}\left[\int_0^1 [bx + a(1-x)]^t \, dx\right]^{1/t}$$

[The final answer should not involve any operations other than addition, subtraction, multiplication, division, and exponentiation.]

Sol. Let $u = bx + a(1-x)$. Then the definite integral becomes

$$I = \frac{1}{b-a}\int_a^b u^t du = \frac{b^{t+1}-a^{t+1}}{(1+t)(b-a)} ,$$

so $\lim_{t\to 0} I = 1$. Hence we can apply l'Hopital's rule to $\frac{1}{t}(\ln I)$:

$$\lim_{t\to 0} I^{1/t} = \exp \lim_{t\to 0} \frac{1}{t}(\ln I)$$

$$= \exp \lim_{t \to 0} \frac{b^{t+1} \ln b - a^{t+1} \ln a}{b^{t+1} - a^{t+1}} - \frac{1}{t+1}$$

$$= \exp \frac{b \ln b - a \ln a}{b - c} - 1$$

$$= e^{-1} (b^b / a^a)^{1/(b-a)}$$

B-3. Let F be a finite field having an odd number m of elements. Let $p(x)$ be an irreducible (i.e., nonfactorable) polynomial over F of the form

$$x^2 + bx + c, \qquad b, c \in F .$$

For how many elements k in F is $p(x) + k$ irreducible over F?

Sol. The polynomial $x^2 + bx + d$ is reducible if and only if there are elements s and t in F such that

$$st = d$$

$$s + t = -b ,$$

or, equivalently, if and only if there is an element s in F such that $-s(s+b) = d$. For $s \in F$, define $f(s) = s(s+b)$. We have just seen that the number of reducible polynomials of the form $x^2 + bx + d$ is equal to the number of elements in the image of f. Notice that $f(s) = f(t)$ if and only if either $t = s$ or $t = -s-b$. Because F is not of characteristic 2, $s = -s-b$ for only one s in F. It follows that the image of F has $1 + [(m-1)/2]$ distinct elements, and therefore the number of irreducible polynomials of the form $x^2 + bx + c + k$ is $m-(1+(m-1)/2) = (m-1)/2$.

B-4. (a) Find a solution that is not identically zero, of the homogeneous linear differential equation

$$(3x^2+x-1)y'' - (9x^2+9x-2)y' + (18x+3)y = 0.$$

Intelligent guessing of the form of a solution may be helpful.
(b) Let $y = f(x)$ be the solution of the nonhomogeneous differential equation

$$(3x^2+x-1)y'' - (9x^2+9x-2)y' + (18x+3)y = 6(6x+1)$$

that has $f(0) = 1$ and $(f(-1) - 2)(f(1) - 6) = 1$. Find integers a,b,c such that $(f(-2) - a)(f(2) - b) = c$.

Sol. (a) Trial of e^{mx} shows that $y = e^{3x}$ satisfies the homogeneous equation. Trial of a polynomial $x^d + \ldots$ shows that d must be 2 and trial of $x^2 + px + q$ shows that $y = x^2 + x$ is a solution. (b) It is easy to see that $y = 2$ satisfies the nonhomogeneous equation and hence that $f(x)$ is of the form $2 + he^{3x} + k(x^2+x)$. Now $f(0) = 1$ gives us $h = 1$, and $[f(-1)-2][f(1)-6] = 1$ leads to $k = 2$. Hence $f(-2) = 6 - e^{-6}$, $f(2) = 14 - e^6$, and we can take $a = 6$, $b = 14$, and $c = 1$.

B-5. In the plane, let C be a closed convex set that contains (0,0) but no other point with integer coordinates. Suppose that $A(C)$, the area of C, is equally distributed among the four quadrants. Prove that $A(C) \leq 4$.

Sol. A support line for C is a straight line touching C such that one side of the line has no points of C. There is a support line containing (0,1); let its slope be m. If $m \geq 1/2$, the part of the area of C in the 4th quadrant is no more than 1 and we are done. Proceed similarly if $m \leq -1/2$.

So we assume that $-1/2 < m < 1/2$ and assume the analogous facts for support lines containing (1,0),(0,-1), and (-1,0). At least one of the angles of the quadrilateral formed by these four support lines is not acute; we may take this angle α to be at a vertex (h,k) in the 1st quadrant. Then $\alpha \geq \pi/2$ implies that $h + k \leq 2$ and this in turn implies that the area of C in the 1st quadrant does not exceed 1. Hence $A(C) \leq 4$.

B-6. For $k = 1,2,\ldots,n$ let $z_k = x_k + iy_k$, where the x_k and y_k are real and $i = \sqrt{-1}$. Let r be the absolute value of the real part of

$$\pm \sqrt{z_1^2 + z_2^2 + \ldots + z_n^2} ,$$

Prove that

$$r \leq |x_1| + |x_2| + \ldots + |x_n| .$$

Sol. Let $X = (x_1,\ldots,x_n)$ and $Y = (y_1,\ldots,y_n)$, and let $a + bi$ be either square root of $z_1^2 + \ldots + z_n^2$. Then $a^2 - b^2 = \|X\|^2 - \|Y\|^2$ and $|ab| = |X \cdot Y|$. Suppose $|a| > \|X\|$. Then, from the Cauchy-Schwarz inequality, $|a||b| = |X \cdot Y| < \|X\| \|Y\|$, it follows that $|b| < \|Y\|$. But then $a^2 = \|X\|^2 - \|Y\|^2 + b^2$ yields $|a| < \|X\|$, a contradiction. Therefore $|a| \leq \|X\|$ and we have $|a|^2 \leq \|X\|^2 \leq (|x_1| + \ldots + |x_n|)^2$, and this implies the desired result.

THE CARUS MATHEMATICAL MONOGRAPHS

The Monographs are a series of expository books intended to make topics in pure and applied mathematics accessible to teachers and students of mathematics and also to non-specialists and scientific workers in other fields.

These numbers are currently available:

1. *Calculus of Variations,* by G. A. Bliss.
2. *Analytic Functions of a Complex Variable,* by D. R. Curtiss.
3. *Mathematical Statistics,* by H. L. Rietz.
4. *Projective Geometry,* by J. W. Young.
6. *Fourier Series and Orthogonal Polynomials,* by Dunham Jackson.
8. *Rings and Ideals,* by N. H. McCoy.
9. *The Theory of Algebraic Numbers* (Second edition), by Harry Pollard and Harold G. Diamond.
10. *The Arithmetic Theory of Quadratic Forms,* by B. W. Jones.
11. *Irrational Numbers,* by Ivan Niven.
12. *Statistical Independence in Probability, Analysis and Number Theory,* by Mark Kac.
13. *A Primer of Real Functions* (Second edition), by Ralph P. Boas, Jr.
14. *Combinatorial Mathematics,* by H. J. Ryser.
15. *Noncommutative Rings,* by I. N. Herstein.
16. *Dedekind Sums,* by Hans Rademacher and Emil Grosswald.
17. *The Schwarz Function and its Applications,* by Philip J. Davis.
18. *Celestial Mechanics,* by Harry Pollard.
19. *Field Theory and Its Classical Problems,* by Charles Hadlock.

Prices for volumes 1-18: List $12.50, MAA member $9.00. Price for volume 19: List $16.00, MAA member $12.00. (Orders for under $10.00 must be accompanied by payment. Prepaid orders will be delivered postage and handling free.)

Orders should be sent to:

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

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

MATHEMATICS MAGAZINE VOL. 53, NO. 2, MARCH 1980