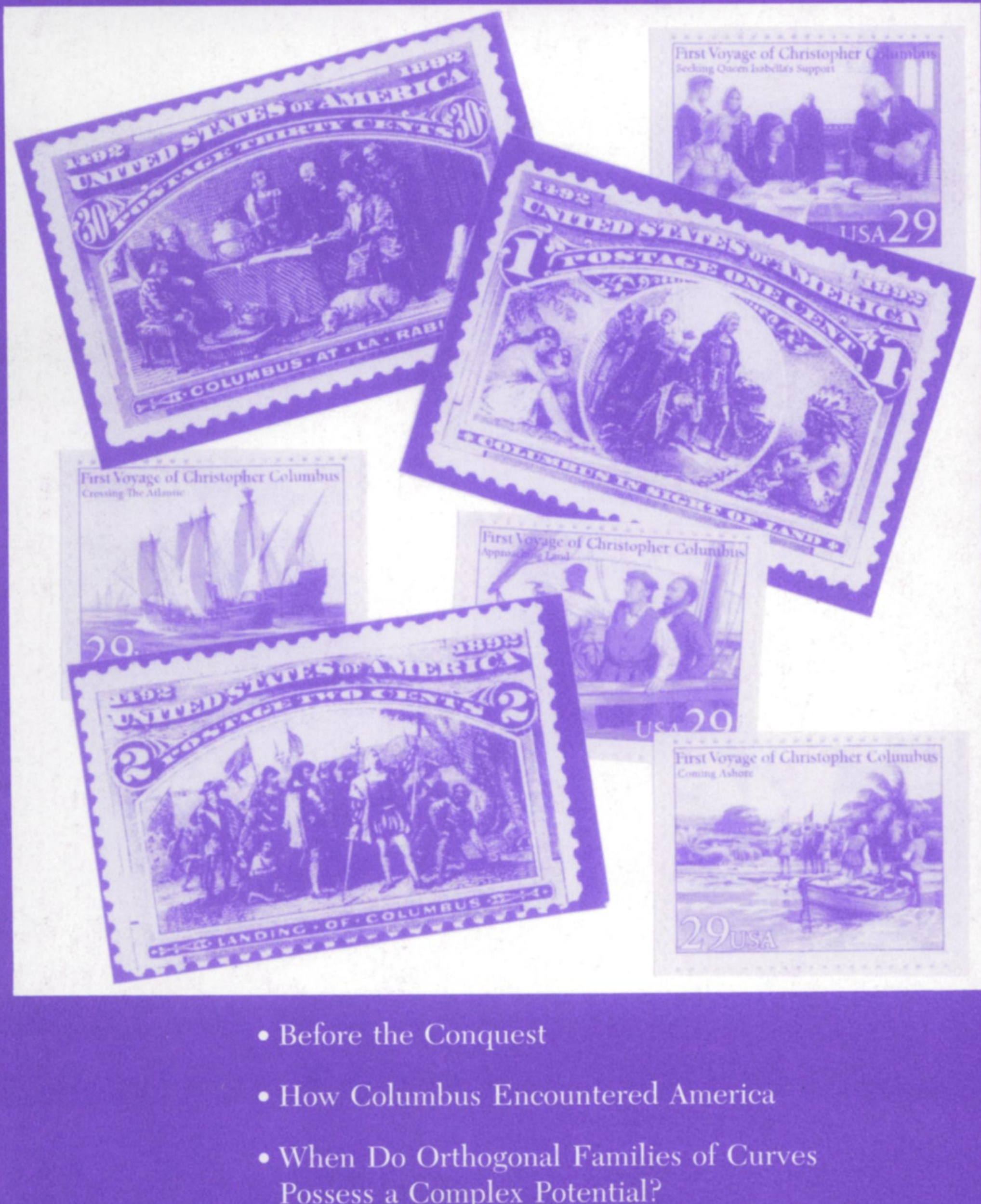
Vol. 65 No. 4, October 1992

MATHEMATICS MAGAZINE



- Before the Conquest
- How Columbus Encountered America
- When Do Orthogonal Families of Curves Possess a Complex Potential?

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

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

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

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

The 1892 stamps shown on the cover were graciously lent to the *Magazine* by Dr. Herbert H. Richtol, Dean of the Undergraduate College, Rensselaer Polytechnic Institute, Troy, N.Y.

AUTHORS

Marcia Ascher is professor of mathematics at Ithaca College. She was educated at Queens College (CUNY) and UCLA. Her original interest was in computational methods. Her more recent interest in the mathematical ideas of traditional peoples was sparked by collaboration with an anthropologist in a study of Inca quipus. She co-authored *Code of the Quipu: A Study in Media, Mathematics and Culture* (University of Michigan Press, 1981) and is the author of *Ethnomathematics: A Multicultural View of Mathematical Ideas* (Brooks/Cole, 1991).

V. Frederick Rickey, a logician turned historian, earned three degrees from Notre Dame (Ph.D. 1968). His research centers on the history of the calculus, and since he is especially interested in encouraging teachers to use history in their teaching, he is preparing an MAA Notes volume from material presented in his MAA Minicourses on 'Using History in Teaching Calculus.' In 1988 his paper "Isaac Newton: Man, Myth, and Mathematics" won a Pólya Prize. This spring he was the first recipient of the Ohio Section's Award for Distinguished Teaching of Mathematics. In the summer he sails on Lake Erie.

Irl C. Bivens, a professor of mathematics at Davidson College, received his A.B. degree from Pfeiffer College and his Ph.D. from the University of North Carolina-Chapel Hill, under the direction of Patrick Eberlein. He has long been interested in connections between geometry and analysis. The present article had its origin in what were initially rather technical computations involving the curl and divergence of vector fields in the plane. It evolved into its current form when he discovered that these computations had a natural reformulation in terms of curvature and complex analysis. When not working or sleeping, he can be found swimming, reading, at the movies, or trying to keep up with his 7-year-old son, Robert.

Vol. 65 No. 4, October 1992

MATHEMATICS MAGAZINE

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

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

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

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

PRINTED IN THE UNITED STATES OF AMERICA

ARTICLES

Before The Conquest

MARCIA ASCHER
Ithaca College
Ithaca, NY 14850

Introduction

In the late 15th century, through their explorers, Europeans "discovered" the New World. Although the discovery would cause drastic change, the New World was, of course, not new to its inhabitants. When the Europeans arrived, there were at least 9 million people in about 800 different cultures living in the Western Hemisphere. Because of the vast disruptions that eventually took place, what we know about them and their mathematical ideas is limited. Most of the cultures had no writing as we commonly use the term and so there are no writings by them in their own words. For the cultures that did not survive, we have primarily what can be learned from archeology and from the writings of the Europeans of the time, who had little understanding and little respect for these cultures so different from their own. For those that did survive, we also have their oral traditions.

We focus here on the mathematical ideas of two sizable groups, the Incas and the Maya. The regions the groups inhabited, their cultures, and their histories are quite distinct, as are their mathematical ideas. Fortunately, for both groups, there is sufficient information for us to gain some understanding of their rich and complex ideas. Here we present an abbreviated introduction to the content and context of the sophisticated data handling system of the Incas and the intricate calendric system of the Maya.

The Incas

The Incas comprised a complex state of about 5 million people that existed from about 1400 C.E. to 1560 C.E. in what is now Peru, and also parts of modern Ecuador, Bolivia, Chile, and Argentina. There were many different peoples in the region, but, starting about 1400, the Incas forcibly consolidated the others into a single bureaucratic entity. The consolidation was achieved by the overlay of a common state religion and a common language, relocation of groups of people, extensive systems of roads and irrigation, and a system of taxation involving, for example, agricultural products, labor, and cloth. The Incas also built a network of storehouses to hold and redistribute goods as well as to feed the army as it moved. The Inca bureaucracy can be characterized as methodical, highly organized, and intensive data users. Although the Incas did not have what we call writing, they did keep extensive records. These were encoded via a logical-numerical system on spatial arrays of colored, knotted cords called *quipus*.

A few selected people from each region that the Incas occupied were trained to serve in the Inca administration and, in particular, to be responsible for gathering, and then encoding and decoding a wide variety of information on the quipus. Believing the quipus to be works of the Devil, the Spanish destroyed thousands of them. Only about 500 remain. These were recovered from graves, probably buried with those who made them. Only rarely can we read the quipus in the sense that specific *meaning* can be assigned. However, we can reconstruct something of their logical-numerical system and, as a result, see the interrelationships of some of the data they contain.

A photograph of a quipu is in FIGURE 1. FIGURE 2 is a schematic. In general, a quipu has a *main cord* from which other cords are suspended. Most of the suspended cords are attached such that they fall in one direction (*pendant cords*); some few fall in the opposite direction (*top cords*). *Subsidiary cords* are often suspended from the pendant or top cords. And there can be subsidiaries of subsidiaries, and so on. (Notice that in FIGURE 2 the first pendant has two subsidiaries on the same level while the fourth pendant has two levels of subsidiaries.) Some pendant cords have as many as 18 subsidiaries on one level, and some have as many as 10 levels of subsidiaries. Sometimes a single cord (*dangle end cord*) is attached to the end of the main cord in a way that sets it apart from the pendant and top cords. All cord attachments are tight so that the spacing between the cords is fixed and serves to group or separate the cords. Overall, a quipu can be made up of as few as three cords or as many as 2000.

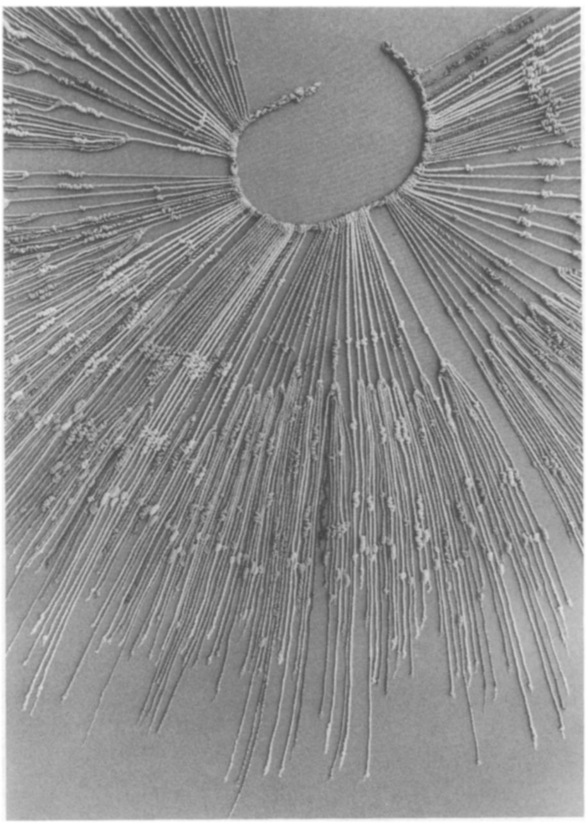


FIGURE 1

A quipu in the collection of the Museo Nacional de Anthropologia y Arquelogía, Lima, Peru. (Photo by Marcia and Robert Ascher.)

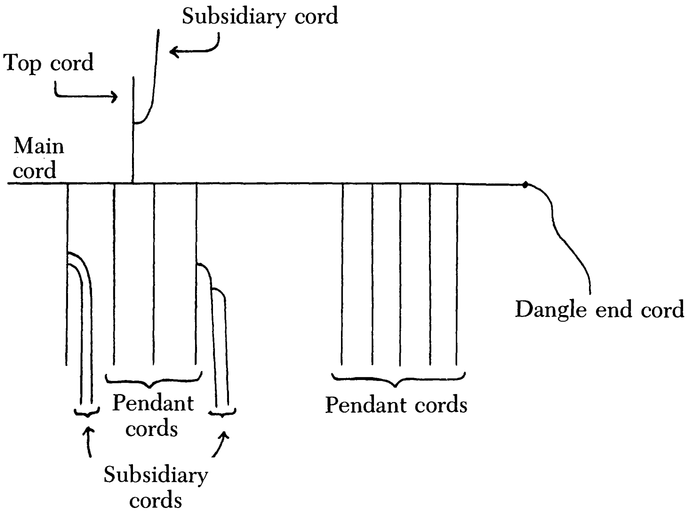


FIGURE 2
A schematic of a quipu.

Color is also a feature of the logical system. It is used primarily to associate or differentiate cords within a single quipu. Thus, color as well as space can create cord groupings. For example, eight pendants can be formed into two groups by having four white pendants followed by four green pendants, or by a four-color sequence repeated twice. In the latter case, each cord is not only associated with its group, but also with the like-colored cord in the other group. Similarly, subsidiaries are associated or differentiated by color as well as by level and relative position on the given level.

Spaced clusters of knots on the cords represent numbers. No matter what the cord placement, only three types of knots appear (single knots, long knots, and figure-8 knots). Depending on the knot types and relative cluster positions, each cord can be interpreted as one number or as multiple numbers. If it is one number, it is an integer in the base 10 positional system. Each knot cluster is read as a digit and each consecutive cluster, starting from the free end of the cord, is valued at one higher power of 10. The units position is always a long-knot cluster or a figure-8 knot, while all other positions are clusters of single knots. When, instead, the cord carries multiple numbers, long-knot clusters or figure-8 knots are interspersed with single-knot clusters thereby signaling the start of a new number. The color coding of the cords also helps in the interpretation of values by enabling the distinction between a numerical value of zero and an intentional omission or blank.

Knot types and knot positions, cord directions, cord levels, color, and spacing are all structural indicators that were combined together in sufficiently standardized ways to be read and interpreted by the community of quipumakers. That is, the quipus served for communication, not as ad hoc personal mnemonic devices. Top cords, for example, generally carry the sum of the pendant cords with which they are grouped on the main cord. Another aspect of the system that is crucial to its general applicability is that numbers were used as labels as well as magnitudes. Particularly with the advent of computers, this usage is now very prevalent in our own culture. For example, the composite number 202-387-5200 is a label identifying a geographic region, a locale within that region, and a specific telephone within that locale.

The quipus, then, are logically structured arrays of magnitudes and labels. Let us translate three of them into notation that is familiar but preserves their logical structure. Then we can delve into some of their internal data relationships.

The quipu shown in FIGURE 1 contains solely quantitative data. Analysis of the pendant cord colors and spacing shows that there are six ordered sets of 18 values each. We will call the jth value in the ith set a_{ij} where $i=1,\dots,6$; $j=1,\dots 18$. When the knots are interpreted as magnitudes, we find that, for all j,

$$a_{1j}=a_{2j}+a_{3j}$$

and, in turn,

$$a_{2j}=\sum_{i=4}^{6} a_{ij}.$$

Hence, the relationship

$$a_{1j}=\sum_{i=3}^{6} a_{ij}$$

also holds. Additionally, there are subsidiary cords on the pendants in five of the six cord groups. Thus, for each a_{ij}, for $i=2,\dots,6$; $j=1,\dots,18$, there are as many as 11 ordered subsidiary values. Call them

$$a_{ijk} \quad \text{with } k=1,\dots,11.$$

Here, too, consistent summation relationships exist:

$$a_{2jk}=\sum_{i=3}^{6} a_{ijk} \quad \text{for } k=1,\dots,11;\ j=1,\dots 18.$$

A modern analogy of data with sums of sums and sets of sums, as is seen in this example, is an accounting scheme for a company, broken down to reflect that it is made up of several departments and producing a variety of products.

In our second example, the arrangement of values and their sums is analogous to a matrix that has, as a subset, the transpose of the sum of two other matrices. Specifically, this quipu's data can be thought of as two 3×3 matrices, each preceded by a single value, and a 3×5 matrix. Calling the elements of the matrices

$$a_{ijk} \quad i=1,2,3;\ j=1,2,3;\ k=1,2,$$

and

$$b_{ij} \quad i=1,2,3;\ j=1,2,3,4,5,$$

the relationship is

$$b_{i,2j-1}=\sum_{k=1}^{2} a_{jik} \quad \text{for } i=1,2,3;\ j=1,2,3.$$

And, continuing the analogy to matrices, the single value preceding each of the 3×3's would be the sum of its first row; that is,

$$c_k=\sum_{j=1}^{3} a_{1jk} \quad \text{for } k=1,2.$$

Some of the data structures remind us of spreadsheets, matrices, and tree diagrams. Other quipus have other layouts, nonquantitative as well as quantitative data, other kinds of internal data relationships, or even relationships with data on other quipus. Many remain fascinating puzzles. One of these, our final example, is from a pair of quipus that were found together.

The specific numbers on these two quipus are different, but the quipus share several internal data relationships, including what we commonly call a difference table. While both appear to be expressions of the same algorithm, a concise unifying description escapes me. Translated into tabular form, they are compared in FIGURE 3. I have superimposed arrows and heavy lines on the tables to indicate the similarities that I see. Perhaps you can find additional similarities or, perhaps, you can find a generalization that unites the data sets.

1	7	4	54	4	4
1	11	1	41	4	4
1	12	2	35	4	5
1	14	2	37	5	6
1	16	2	41	6	2

6	4	4	82	10	10
6	8	6	30	10	12
6	14	8	10	12	17
6	22	12	13	17	14
6	34	15	15	14	15

FIGURE 3

Data excerpted from a pair of quipus found together. Arrows and heavy lines highlight some of their similarities. In both, for example, all values are the same in the first column and in row 2, column 3. Also, in both, the third column contains the differences of consecutive values in the second column. (The quipus are described by C. Radicati di Primeglio in La "seriación" como posible clave para descifrar los quipus extranumerales, *Documenta: Revista de La Sociedad Peruana de Historia* 4 (1965), 112–215.)

Reference [1] contains more details, examples, and discussion of the context, contents, and interpretation of quipus. Although we lack the cultural associations needed to know what a specific quipu means, the quipus surely are records of human and material resources and calendric information. But they contain much more—possibly information as diverse as construction plans, dance patterns, and even aspects of Inca history. Overall, the logical-numerical system embedded in these spatial arrays of colored, knotted cords was sufficiently *general* to serve the needs of the Inca bureaucracy. Their use was terminated soon after the destruction of the Inca state in 1560 C.E.

The Maya

The Mayan peoples have a complex cultural tradition extending over a long period of time and encompassing different groups speaking about 25 different languages. They shared much in the way of culture but, spread through time and space, they had different centers and political organizations, some different ideas, and some different practices. Discussions of their history usually begin sometime before 1000 B.C.E. The period 200–1000 C.E. is referred to as the Classic period and is marked by ceremonial centers with monumental architecture, a system of writing, an elaborate astrological science, and numerous centers of social, religious, economic, and political activities interrelated by marriage and trade networks. During the Classic period, the

Maya inhabited what are now the eastern Mexican states of Chiapas, Tabasco, Campeche, Quintano Roo, and Yucatan; Belize; Guatemala; and the western portions of Honduras and El Salvador. On the eve of the Spanish conquest, there were spread out in this area, many independent yet culturally interrelated states, none as grandiose as earlier. Because they were dispersed and independent, they did not succumb to the Spanish as quickly and easily as did the Incas. Today, primarily in Chiapas and the highlands of Guatemala, some Maya traditions continue.

Christopher Columbus, in 1502, is said to have been the first European to encounter the Maya, and his brother, Bartholomew, was the first to record the name of the group. By that time, however, remains of the Classic Maya period were already covered over, and so another "discovery"—this time archeological—took place beginning in the mid-1800s. In addition to some ongoing traditions, what we know of the Maya, and in particular of their mathematical ideas, comes from archeological materials, including thousands of inscribed stone monuments (*stelae*) and four post-Classic books (codices), the only ones remaining of the *thousands* that were burned by the Spanish.

We will concentrate on the idea of *time* as it permeates the Mayan culture. Time is considered to be cyclic. Supernatural forces and beings are associated with and influence units of time. Events of the past, present, and future are related through the recurrence of named time units. There are, however, not just one, but several, overlapping cycles that all must be taken into consideration to give meaning to any particular time unit. Although their calendric concerns extend to the incorporation of astronomical phenomena, the Maya were preoccupied with the interrelationship of the arbitrary cycles they created and imposed on time. For this reason, the Maya are said to have "mathematized" time and, through it, their religion and cosmology.

There is, first of all, a 260-day *ritual almanac*. Each day within it is identified by a number in a cycle of 13 and a named deity in a cycle of 20. (Each of the 13 numbers also has an associated deity.) There is a *vague year* of 365 days (called "vague" because it does not keep in alignment with the seasons). It results from a cycle of 20 numbers *within* a cycle of 18 named deities plus five unnamed days. (The cycle of 20 is now referred to as a *month* but does not have a lunar correspondence.) One *calendar round* is 18,980 days (52 vague years, 78 almanac cycles) since that is the least common multiple of 365 and 260. A date within this, made up of an almanac date and a vague year date, reads, for example, 4 Ahau 8 Cumku where Ahau and Cumku are names of deities.

In the ceremonial centers of the Classic period, there were temples atop large, stepped pyramid frusta as high as 213 feet. Hundreds of stelae, some as tall as 32 feet, were erected around them to commemorate different events. To mark an event, what was needed was to accurately and *fully* identify it in time and, sometimes, to state. how many days it was from another event. In addition to the calendar round date, another significant identifier was a *Long Count*: the number of days from the beginning of the then current *Great Cycle*. A Great Cycle is based on a 360-day period (a *tun*) consisting of 18 *uinals* of 20 days each; 20 tuns are a *katun*; 20 katuns are a *baktun*; and 13 baktuns are a Great Cycle. An example of a Long Count transcribed into our numerals is 9.0.19.2.4. From left to right this reads "9 baktuns, 0 katuns, 19 tuns, 2 uinals, and 4 days." To convert to our system, starting at the right, each position—with the exception of the third—is multiplied by one higher power of 20. In the third position, an 18 is used instead. Hence, the Long Count date of 9.0.19.2.4 is interpreted as:

$$9 \cdot 18 \cdot 20^3 + 0 \cdot 18 \cdot 20^2 + 19 \cdot 18 \cdot 20 + 2 \cdot 20 + 4 = 1{,}302{,}884 \text{ days}$$

from the beginning of the Great Cycle that started on the calendar round date of 4 Ahau 8 Cumku. The exact correlation of this date with the Gregorian calendar is not known. However, by one commonly accepted correlation, the beginning of the Great Cycle was in 3114 B.C.E. and the date given by the Long Count above is, thus, in 454 C.E.

This Long Count date appeared on a stela that also was dated in the calendar round as 2 Kan 2 Yax. But, as with most stelae, it had dates placing it within still more cycles. There was a 9-day cycle of *Lords of the Night*, each associated with one of the nine levels of the underworld. Hence, a specific Lord of the Night also dates the day being marked. And, in addition, the day is placed within a *lunar cycle*. Lunar years and half-years are made up of 29- and 30-day lunar months. The stela contains the moon number within the lunar half-year, the age of the moon, and whether it is a 29- or 30-day month.

Just as there are nine levels below the earth, there are 13 levels in the heavens above. There are four cardinal directions and each of the quadrants they define is associated with a different color. Uniting time and space, the days of the 260-day ritual almanac move in a counterclockwise direction through the four quadrants. Hence, not only are time and space related, but the ritual almanac has within it a four-color cycle. In some cases, where dates also identify days within a 819-day cycle associated with the rain god, the use of four colors effectively makes that cycle $819 \cdot 4 = 3276$ days.

Many of the Maya computations are projections into the past or into the future that require dovetailing the cycles. For instance, one inscription, commemorating the enthronement of a ruler, gives the calendar round dates of his birth and his enthronement, as well as of the enthronement of an earlier, somehow related ruler or deity. The number of days between these events is also included in Long Count form. For example, the time elapsed since the enthronement of the deity is 7.18.2.9.2.12.1 days. Hence, given one calendar round date, a calendar round date some $1\frac{1}{4}$ million years earlier was calculated or, given two calendar round dates, their Long Count difference (number of days between them) was calculated.

To more fully savor the calculation, you might try to do such a problem. First, recall that each of the 260 days in the ritual almanac is identified by a number in a cycle of 1 to 13 and a named deity in a cycle of 20. For simplicity, let us call the deities $D_1, \ldots, D_{20}$. In the 365-day vague year, five days are unnamed while, for the rest, each day is identified by a number in a cycle of 1 to 20 within each of 18 months named for deities. Call these deities $d_1, \ldots, d_{18}$ and assume that the five unnamed days follow $20d_{18}$. The calendar round date is the almanac date followed by the vague year date. What, then, is the calendar round date that is 2.3.5.10 days after $8D_{10}13d_{10}$? And, what is the Long Count Difference between $12D_8 6d_2$ and the next $5D_4 12d_{17}$? (Answers are on page 235.)

The Dresden Codex, attributed to the eleventh century in Yucatan is the most mathematical of the codices. It is constructed as a long strip of paper made from tree bark, folded into pages, coated with white plaster, and painted. Perhaps as aids to computation, the codex includes several tables of multiples; for example, there are tables of multiples of 5.1.0 (that is, of 1820 that equals $7 \cdot 260$ and $5 \cdot 364$) up to 1.0.4.8.0. But, even more important, other tables in the codex combine backward and forward calendar projections with evidence of keen astronomical knowledge. One set of tables correlates lunar cycles with ritual almanac dates. These tables cover 405 lunations and are interpreted as prediction tables for possible eclipses. Another set of tables in this codex correlates Venus visibility events with ritual almanac and vague year dates. Covering 65 Venus cycles, which is 146 almanac cycles and 104 vague years, it includes corrections reflecting the fact that the mean synodic year of Venus is not an integral number of days. (The mean synodic year of Venus is 583.92 days.) The corrections are such that the error between real and tabulated times of the positions of Venus would be off by just two hours in about 500 years!

We know that much is unknown about the knowledge and mathematical ideas of the Maya. Dates and numbers, written in a variety of symbolic forms, have been recognized and deciphered. But the Maya writings contain much more. The writing system is complex because it contains about 1000 symbols and both phonetic and nonphonetic elements. A great deal of recent activity in decipherment raises the hope that more will become known about Maya ideas and the Maya culture in general. (For more details on number representation, the tables in the Dresden Codex, and possible algorithms for the date difference calculations, see [2] and [5]. Reference [3] discusses the importance of the Maya scribes including evidence that they were both women and men. Also, [4] is an excellent comprehensive overview of the Maya.)

Conclusion

The Inca and Maya are two substantial examples of cultures whose mathematical ideas were both sophisticated and independent of those of Western culture. We can never know about all of the mathematical ideas they had and, what is more, we cannot know what they might have developed had they continued to thrive.

REFERENCES

1. M. Ascher and R. Ascher, *Code of the Quipu: A Study in Media, Mathematics and Culture*, University of Michigan Press, Ann Arbor, MI, 1981.
2. M. Closs, The mathematical notation of the ancient Maya, *Native American Mathematics*, ed. by M. Closs, University of Texas Press, Austin, TX, 1986, pp. 291–369.
3. ______, My mother and my father were both scribes, *Research Reports on Ancient Maya Writing*, Center for Maya Research, Washington, DC, in press.
4. J. S. Henderson, *The World of the Ancient Maya*, Cornell University Press, Ithaca, NY, 1981.
5. F. G. Lounsbury, Maya numeration, computation, and calendrical astronomy, *Dictionary of Scientific Biography*, Vol. 15, suppl. 1, Scribners, New York, 1978, pp. 759–818.

How Columbus Encountered America

V. FREDERICK RICKEY
Bowling Green State University
Bowling Green, OH 43403-0221

Christopher Columbus (1451–1506) went to sea as a young lad, but probably only for a short voyage not far from his native Genoa. We don't know how or when he learned to sail, but he did so at every opportunity and became a serious student of the art. In 1476, after the ship he was on came under attack and was sunk, Columbus swam ashore near Cape St. Vincent, the southwestern tip of Portugal and Europe. The next year he was aboard a ship that probably sailed to Iceland and possibly even on to the Bay of Fundy [**9**]. Back in Lisbon after his experiences at sea, he was among people who could teach him all he wanted to learn: Portuguese and Castilian, the languages of seamen; Latin, to read geographical works; and mathematics and astronomy, for navigation [**5**, p. 27].

In 1479 Columbus married Dona Felipa Perestrello e Moniz, the daughter of a captain in the service of Prince Henry the Navigator (1394–1460). This provided the opportunity to pore over the charts and logs of his deceased father-in-law. With his brother Bartholomew, he worked for a time as a mapmaker. There were many occasions to talk with old seamen about their voyages, especially about the mysteries of the western seas.

He took every opportunity to sail out into the Atlantic: to the Azores (39° North), to Madeira (33° N), and to the Canary Islands (28° N), as well as down the coast of Africa. He heard stories of strange trees being blown in from the west of the Azores during storms, and observed for himself that the prevailing winds in the winter were from the west. Since he wanted to remember these stories, he recorded them in his notebooks [**10**, p. 76]. When he sailed around the Canaries he realized that in the fall the winds blew to the west and that the seas were generally calm. This was a key piece of information.

Whether it was from what he heard, what he read, or what he observed, Columbus conceived the idea of sailing west across the Atlantic to reach the Indies. Like a mathematician, once he had his conjecture and believed it was true, he set out to prove it. With great energy and tenacity, Columbus read and annotated the *Historia rerum ubique gestarum* (1477) of Aeneas Sylvius Piccolomini (Pope Pius II), which told of the work of Eratosthenes on the size of the Earth; the *Imagio mundi* (printed 1480) of Pierre d'Ailly (1350–1420), the first Latin edition of the *Geographiae* (1478 edition) of Ptolemy (A.D. 90–168), which contained maps; and the first Latin edition of the travels, *De consuetudinibus et condicionibus orientalium regionum*, of Marco Polo (1254–1324), which related his travels to Cathay (China) and Cipangu (Japan) [**7**, p. 57].

Earth is not flat

Of all the errors associated with Columbus, the most persistent is that he had to convince people that Earth was not flat and that by sailing west to America he proved it was round. There are two difficulties with this. Had he reached Asia, Columbus would have proved the world round—but reaching America is perfectly consistent with the flat-earth theory. In his day, and in fact since the time of the ancient Greeks, every educated individual knew that the world was round. Although Russell [**11**] has identified five people between the time of the ancient Greeks and Columbus' day who believed the Earth was flat, everyone else knew it was spherical. Russell has shown that the flat-earth myth was created by Washington Irving (1783–1859) in his romantic biography *History of the Life and Voyages of Christopher Columbus* (1828).

Pythagoras (560–480 B.C.) was the first to discover that the Earth is spherical, not flat or disk-shaped. How he discovered this is unknown. He may have argued that the sphere is the most perfect of solid figures, and so the Earth should be spherical. This argument would have been compelling to his brotherhood, even though we find it mystical. More likely, Pythagoras had seen ships "hull-down" as they sailed out into the Mediterranean and, like seamen throughout the ages, he had seen mountains "raising" as they were approached [**5**, p. 33].

In *De coelo*, Aristotle (384–322 B.C.) argued, from observing the Earth's shadow on the moon during an eclipse, that the Earth must be spherical. Another proof he mentioned is provided by the fact that different stars are visible in Greece and in Egypt. Since only a short trip is necessary to observe the difference, Aristotle concluded the Earth is not very large [**10**, p. 5]. In his copy of *Imago mundi*, Columbus wrote that "Aristotle says between the end of Spain and the beginning of India is a small sea navigable in a few days" [**6**, p. 22].

Eratosthenes on the size of the Earth

Once Earth's sphericity was accepted it was natural to ask its size. There are ancient records of a number of estimates. Archimedes (287–212 B.C.) in his *Sand Reckoner* used 3,000,000 stadia for the circumference. But that is obviously an overestimate, for he wanted an upper bound on the number of grains of sand it will take to fill the universe. All that we know about the methods used for measuring the Earth was reported by Cleomedes (1st c. A.D.?), who reported the techniques and results of Eratosthenes (276–195 B.C.) and Posidonius (135–51 B.C.). These two methods are the heart of the mathematical matter, so we examine them in some detail.

Eratosthenes, the librarian at the Museum in Alexandria who is best remembered for his sieve for finding primes, based his measurement of the Earth on several observations.

0. The Earth is spherical and the rays of the sun are parallel.
1. In Syene (Aswan) at the time of the summer solstice a gnomon, or vertical pole, cast no shadow, so the sun at noon would shine to the bottom of the deepest wells. Thus Syene was on the Tropic of Cancer.
2. At Alexandria, again at noon on the summer solstice, the angle cast by a shadow was $1/50$th of a great circle.
3. Syene is directly south of Alexandria and so they are on the same meridian.
4. The distance between Syene and Alexandria is 5,000 stadia, a distance that had been measured by bematistes, surveyors trained to walk with equal steps and to count them.

All of this can be summarized by FIGURE 1.

By elementary geometry Eratosthenes determined that the circumference of the Earth was $5{,}000 \cdot 50 = 250{,}000$ stadia. Later, when the circle was divided into 360 degrees, this was rounded up to 252,000 stadia, so that each degree was 700 stadia.

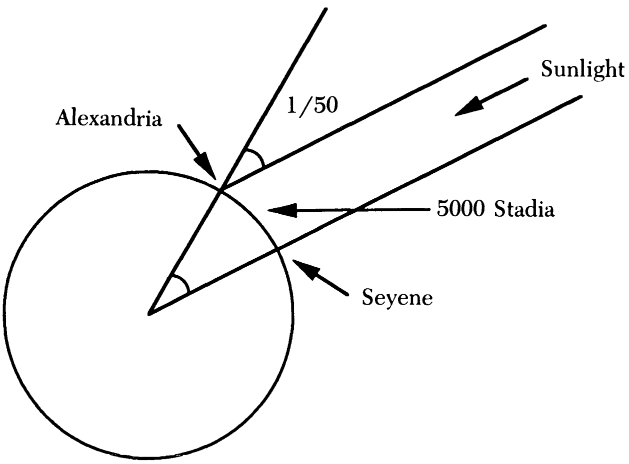


FIGURE 1

Posidonius on the size of the Earth

Posidonius used a different method to measure the circumference of the Earth. At Rhodes, which is 5,000 stadia north of Alexandria, the star Canopus rose just high enough to graze the horizon and then set again immediately. At Alexandria however, Canopus rose to a meridian height of 1/48th of a great circle. Thus the Earth's circumference was $5{,}000 \cdot 48 = 240{,}000$ stadia. Again a diagram is useful.

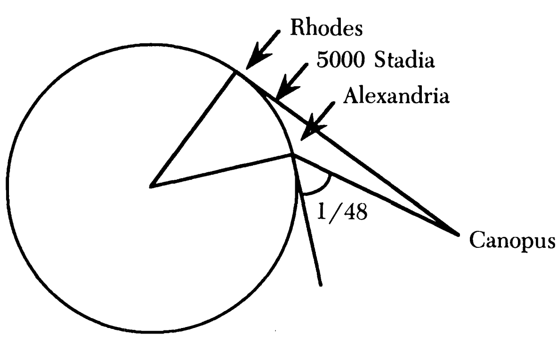


FIGURE 2

The Greek geographer Strabo (64 B.C.–A.D. 25), whose scientific skills were limited, criticized the method of Eratosthenes, and then recorded that Posidonius estimated the circumference of the Earth to be 180,000 [sic!] stadia. With this stroke of the pen, the size of the Earth was reduced by one-fourth. This value was accepted by Ptolemy in his *Geographiae*, a work that Columbus would read. All of this work has been much debated by scholars; see, besides the articles in the *Dictionary of Scientific Biography* [2], the papers of Diller [1] and Taibak [12].

An aside on modeling

A careful comparison of these two methods is a good exercise for budding students of mathematical modeling. From a geometrical point of view, both methods are unexceptional. Both would give the correct and identical result if the angles and distances had been measured correctly. From a physical point of view, the models are *not* equivalent. The method of Posidonius is subject to error introduced by atmospheric refraction. When observing celestial objects near the horizon, their apparent position is above their true position. A better diagram would be this:

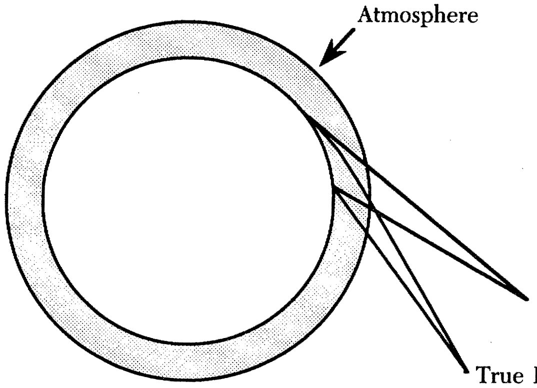


FIGURE 3

The moral of this story is that modeling requires a knowledge of the physical situation as well as a knowledge of mathematics.

It is also worthwhile to discuss the effect of errors of measurement. All of Eratosthenes' observations are incorrect. Yet, by a fortuitous cancellation of errors his result is amazingly accurate (taking 10 stadia to the mile—and there is much debate about this equivalence—gives a value quite close to the circumference of the Earth). Send your students to the library to check if Aswan is on the Tropic of Cancer and if it is directly south of Alexandria. Ask them how the angle measurements could be done by the naked eye before the creation of trigonometry (Goldstein [3] argues the values were computed, not measured). And discuss those 5,000 stadia. How could they be measured so accurately, so roundly? Surely these are the earliest examples of fudged data. All of this will generate a profitable classroom discussion.

We are all aware that mathematics is—and always has been—a vital force in our culture. Likewise, our students are always after relevant applications—and rightly so. The best applications for the classroom involve transparent mathematics, a problem that is readily understood, and that seems to be significant. The voyage of Columbus has clear significance; the fact that it has mathematical underpinnings should help to convince the doubting student that mathematics permeates our world. Consequently, we teachers should look at a map and find a school on our meridian to cooperate with us in recreating the measurement of Eratosthenes [13].

The argument of Columbus

The essential question that Columbus had to answer was the distance he had to sail between Europe and Asia. Moreover, he had to convince his contemporaries that it was possible to make the trip. Of necessity, he broke this into two tractable problems:

(1) the circumference of the Earth or, as he preferred, the length of a degree, and (2) an estimation of the width of Asia.

The earliest surviving book printed in Western Europe is dated 1447. By 1500 at least 30,000 individual editions had been published (totaling some 20 million copies in Europe), with perhaps 10 percent of them dealing with scientific subjects. Of interest to us is the fact that Jacopo Angelo returned from Constantinople with his Latin translation of Ptolemy's *Geographiae*, and after 1410 there were numerous manuscript copies. It was first printed in 1475, and Columbus read the 1478 edition, the first containing maps. Similarly, Strabo's *Geographiae* was published in Latin translation in Rome in 1469, and three more times by 1480. Thus there were four editions for Columbus to study.

The journal of Marco Polo was dispersed in numerous manuscript copies before the invention of printing. This work sparked the imagination of Columbus, among others. Those who were still under the spell of the ancients could not believe the extent of the land mass of Asia that Polo described. One of the few men of science to believe Polo was Paolo Toscanelli (1397–1482), Paul the Physician. In his native Florence, Toscanelli talked with many travelers about geography and navigation, and then decided to construct a nautical map of the Atlantic (for a reconstruction see [4], p. 85). On 25 June 1474, Toscanelli wrote in a letter that Asia could be reached by sailing west [5, p. 34]. Later, Columbus, learning of Toscanelli's ideas, wrote to obtain a copy of this letter and chart. Toscanelli inferred from Polo that Asia was 30° wider than scholars of the day generally believed, and even put Cipangu 1500 miles off the coast of Asia. Toscanelli thought that by sailing due west from Lisbon, one would reach Cathay in about 5,000 nautical miles (see [10], pp. 82–88 for the correspondence). This support for his belief and the encouragement to attempt the trip from a scholar was very provocative to Columbus. But he felt he knew better than Toscanelli and so made the ocean even narrower.

Certainly Columbus would have heard of the trip that Fernão Dulmo made in 1487. Dulmo sailed west, for 40 days, from the Azores to look for the mythical Antillia, the Isle of the Seven Cities. He found nothing, and Columbus knew why. The prevailing winds at that latitude are from the west, and it is difficult to sail into the wind, for the rigging must be repeatedly adjusted to sail a zig-zag course [5, p. 74]. This encouraged Columbus in his plan to sail from the Canary Islands where he could run down-wind, i.e., sail with the wind at his back and with little need to adjust the sails. Columbus knew that Eratosthenes had written of the climate south of the equator and that Pierre d'Ailly claimed that the Torrid Zone was uninhabitable, but he had sailed there himself, and so he knew better than these authorities.

With this information gathered, we are now ready for some arithmetic using "truly imaginary" numbers. Columbus knew that Eratosthenes estimated the length of a degree to be 59.5 nautical miles (the correct value is 60, because, by definition, a nautical mile equals one minute of latitude). But he preferred the value of Al-Farghānī (d. after A.D. 861), which is $56\frac{2}{3}$ Arabic miles. Columbus assumed that the shorter Roman or Italian mile was used, so equated the degree to 45 nautical miles. Nunn [7] has argued that Columbus actually checked this estimate on a trip down the coast of Africa. The error in Columbus's estimate of the length of a degree was caused by incorrect measurement of the latitude, which was always difficult to measure when aboard ship.

Ptolemy claimed that the known world covered half the globe, 180° from Cape St. Vincent to the coast of Asia. That was 50% too large, but Columbus insisted it was too small, preferring the estimate of 225° of Marinus of Tyre (2nd C. A.D.). To this

Columbus added 28° for the travels of Marco Polo, and 30° for the reputed distance between Cathay and Cipangu. He intended to start from the Canaries, which are 9° west of Portugal. This left only 68° of ocean to cross between the Canaries and Cathay. But Columbus was still not happy. He thought that Marinus' degree was too large, and so reduced the open water to 60°. Finally, since he intended to sail due west from the Canaries, which are at latitude 28°, the length of the degree along his course was only 40 nautical miles. Thus he only had to sail $60 \cdot 40 = 2400$ nautical miles. In fact, the distance along latitude 28° from the Canaries to Japan is 10,600 nautical miles [**5**, pp. 67–68].

In 1484 Dom João II, King of Portugal, appointed a Maritime Advisory Committee, the *Junta dos Mathemáticos*, to advise him on matters of navigation. When Columbus presented his ideas to the *Junta* they knew his estimates disagreed with the best estimates of the day, and so dismissed his proposal. Thus he headed for Spain. After several years of trying, he finally obtained support to sail west from King Ferdinand and Queen Isabella.

The first voyage west

On Friday the third of August 1492 (old-style Julian calendar), Columbus left Palos, on the southern coast of Spain, with three ships and about 90 men. A few days out, the rudder on the *Pinta* was damaged and several weeks were lost making repairs in the Canaries. On 6 September they left the Canaries, heading straight west. This course was maintained save for two deviations. The first was a little search for the mythical island of Antillia. The second occurred on 7 October after they had gone over 2400 miles. There was concern that they had already sailed past Cipangu, so a decision was made to sail WSW to follow the flocks of birds that they had seen. In the early hours of 12 October, land was sighted. Columbus christened it San Salvador, in honor of the Holy Savior who had guided their voyage.

The most amazing thing about this famous voyage is that Columbus found land just where he expected to find land. In his diary, which makes most interesting reading, Columbus recorded on 3 September 1492 that he estimated the voyage would take 21 days, but if the winds were light it might take 28 [**6**]. The voyage took 33 days. Because of this confirmation of his beliefs, he was convinced that he had reached the Indies. He never realized that what he had happened across was, for the Europeans, a new and unknown world.

After three months of exploring, Columbus and his crew headed for home on a north-west course. This course was continued until they reached 35° N when they headed east toward the Azores. Columbus chose this path because he did not want to fight the wind back to the Canaries and because he conjectured that the wind would again be at his back if he took the northerly route. We often lose sight of the fact that Columbus did more than encounter America. He discovered a way to sail there—and back.

Conclusion

Columbus was a lucky man, unbelievably lucky. Had America not been where he thought Cipangu was, we would not be commemorating this quincentennial. The story of this voyage is almost a comedy of errors. There is some good mathematics here, but Columbus did not do it. In geographical exploration, as in mathematical exploration, it is not the serpentine path to discovery that matters, but the fact that our initial naïve ideas lead to solid results. Although Columbus was wrong about the size of the Earth and the extent of Asia, his idea of sailing west brought solid results. Mathematics, however, took time to blossom in America. The 400th anniversary of the voyage of Columbus was celebrated in Chicago with the World's Columbian Exposition in 1893. This had a tremendous effect on the development of mathematics in America because Felix Klein came from Germany and spoke about some of the recent advances in mathematics and put us on the right course [8].

REFERENCES

Note: The literature on the topics briefly touched on in this paper would fill a stadium. I have read only 1/48th of that. Only items actually cited are listed here.

1. Aubrey Diller, The ancient measurements of the Earth, *Isis* 40 (1949), 6–9.
2. Charles Coulston Gillispie, ed., *The Dictionary of Scientific Biography*, Charles Scribner's Sons, New York, 16 volumes, 1970–1980. (This is the best source of information on deceased scientists. There are articles on many of the individuals mentioned in this paper.)
3. Bernard R. Goldstein, Eratosthenes on the 'measurement' of the Earth, *Historia Mathematica*, 11 (1984), 411–416.
4. Jay A. Levenson, ed., *Circa 1492. Art in the Age of Exploration*, National Gallery of Art, Washington, DC, 1991.
5. Samuel Eliot Morison, *Admiral of the Ocean Sea: A Life of Christopher Columbus*, Boston, 1942.
6. Samuel Eliot Morison, *Journals and Other Documents on the Life and Voyages of Christopher Columbus*, Heritage Press, New York, 1963.
7. George E. Nunn, *The Geographical Conceptions of Columbus*, American Geographical Society of New York, 1924. Reprinted by Octagon Books, New York, 1977.
8. Karen Hunger Parshall and David E. Rowe, American mathematics comes of age: 1875–1900, pp. 3–28 in volume 3 of *A Century of Mathematics in America*, edited by Peter Duren et al., AMS, Providence, 1989.
9. Kåre Prytz, *Inventio Fortunata*. The book that guided Columbus to America, *The Norseman* (January 1988), 9–11.
10. David B. Quinn, *New American World. A Documentary History of North America to 1612*, 5 vols., Arno Press, New York, 1979. Only volume 1 is cited.
11. Jeffrey Burton Russell, *Inventing the Flat Earth*, Praeger, New York, 1991.
12. C. M. Taibak, Posidonius vindicated at all costs? Modern scholarship versus the Stoic earth measurer, *Centaurus* 18 (1974), 253–369.
13. Jerry Webb and Roberta Bustin, Eratosthenes revisited, *The Physics Teacher* 26 (1988), 154–155.

When Do Orthogonal Families of Curves Possess a Complex Potential?

IRL C. BIVENS
Davidson College
Davidson, NC 28036

Introduction

Let $\Omega = \Omega(z) = \Omega(x, y) = u(x, y) + iv(x, y)$ denote an analytic function with nonzero derivative defined in a region U of the complex plane. It is well known that the respective level curves of u and v are everywhere orthogonal and thus form an orthogonal "net" in U. For example, the function $\Omega(z) = z = x + iy$ produces the families of vertical and horizontal lines in the plane, while the function $\Omega(z) = z^2 = (x + iy)^2 = x^2 - y^2 + i(2xy)$ generates the net comprised of two orthogonal families of hyperbolas (together with their asymptotes) in the plane minus the origin (FIGURE 1). Such nets are important both for the interpretation of Ω as a conformal transformation and in the study of two-dimensional, irrotational, incompressible fluid flow where they comprise the streamlines and equipotential lines of the flow. Likewise, these nets have interpretations as the equipotential lines and lines of force in planar electrostatics and as the isothermal and flux lines in two-dimensional heat flow. (These physical applications of complex analysis may be found, for example, in [4, Chapt. 5] or [5, Chapt. 9].)

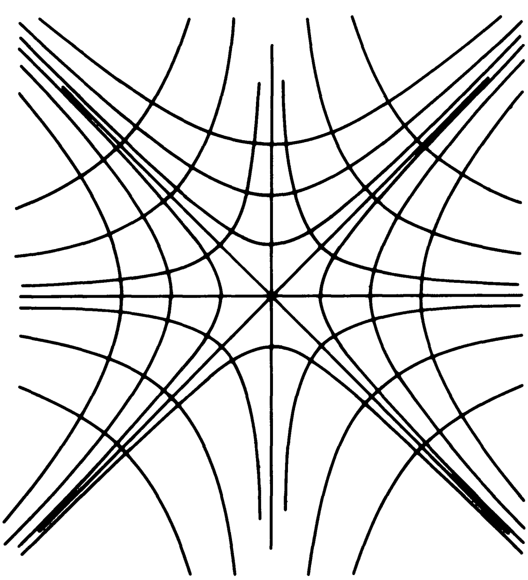


FIGURE 1

Whenever an orthogonal net is derived from an analytic function in this way, we may (borrowing some terminology from fluid mechanics) refer to the analytic function as a complex potential for the net. On the other hand, *given* an orthogonal net defined in a region, one can ask whether or not there exists (at least locally) a complex potential for the net. This paper presents a simple criterion for the existence of such a potential in terms of the arclength derivatives of the plane (signed) curvature functions for the curves in the net. Furthermore, when a potential exists we will be able to compute these curvature functions directly in terms of the (complex) derivatives of the potential. (While these results are more or less straightforward consequences of the classical theory of "isothermal parameters" on a surface, perhaps they deserve to be seen in a less esoteric context.) Finally, we will characterize those orthogonal nets for which a complex potential exists and for which the curvature functions corresponding to orthogonal curves of the net have a constant ratio.

Orthogonal nets and associated curvature functions

Let U denote an open subset of the complex plane and suppose that A and B are two orthogonal families of curves in U such that exactly one curve of each family passes through any given point of U. In this case we will say that the two families of curves comprise an **orthogonal net** $N = N(A, B)$ in U. We will assume that N is smooth in the sense that for any point P in U there exists a smooth, real-valued function $\Theta = \Theta(x, y)$ defined in a disk V containing P such that the orthonormal vector fields $\mathbf{T} = \cos\Theta\mathbf{i} + \sin\Theta\mathbf{j}$ and $\mathbf{N} = -\sin\Theta\mathbf{i} + \cos\Theta\mathbf{j}$ are everywhere tangent to the curves of the respective families A and B. (Note that the net uniquely determines the function Θ up to additive integral multiples of π.)

For example, suppose $\Omega = \Omega(z) = \Omega(x, y) = u(x, y) + iv(x, y)$ is an analytic function with nonvanishing derivative defined on an open subset U of the complex plane. Let A and B denote the collections of level curves of the imaginary and real parts of Ω respectively. Then the vector field $\operatorname{grad} u = u_x\mathbf{i} + u_y\mathbf{j} = u_x\mathbf{i} - v_x\mathbf{j} = (\operatorname{Re}\Omega')\mathbf{i} - (\operatorname{Im}\Omega')\mathbf{j} = (\operatorname{Re}\overline{\Omega'})\mathbf{i} + (\operatorname{Im}\overline{\Omega'})\mathbf{j}$ is everywhere normal to each curve in B and the vector field $\operatorname{grad} v = v_x\mathbf{i} + v_y\mathbf{j} = (\operatorname{Im}\Omega')\mathbf{i} + (\operatorname{Re}\Omega')\mathbf{j}$ is everywhere normal to each curve in A. But this implies that the families A and B comprise an orthogonal net in U with the vector field $\operatorname{grad} u = (\operatorname{Re}\overline{\Omega'})\mathbf{i} + (\operatorname{Im}\overline{\Omega'})\mathbf{j}$ tangent to each curve in A and with $\operatorname{grad} v = (\operatorname{Im}\Omega')\mathbf{i} + (\operatorname{Re}\Omega')\mathbf{j}$ tangent to each curve in B. If the curves in A and B are oriented in the directions of $\operatorname{grad} u$ and $\operatorname{grad} v$ respectively then we will say that N has been given the orientation "induced" by the complex potential Ω. Note that in the directions of the induced orientation, u increases on each level curve of v and v increases on each level curve of u. We will now show that there is a natural choice of a *harmonic* local function Θ for which the vector fields $\mathbf{T} = \cos\Theta\mathbf{i} + \sin\Theta\mathbf{j}$ and $\mathbf{N} = -\sin\Theta\mathbf{i} + \cos\Theta\mathbf{j}$ lie in the directions of the induced orientation. Given any point P in U choose a branch of $\log(\Omega')$ defined on a disk V containing P and let Θ denote the harmonic function in V defined by $\Theta = -\operatorname{Im}\log(\Omega') = \operatorname{Im}\log(1/\Omega')$. If $h = 1/\Omega'$ then $h(z)$ is a positive real multiple of $\overline{\Omega'(z)}$ and $h/|h| = e^{i\Theta} = \cos\Theta + i\sin\Theta$. Therefore, the vector fields $\mathbf{T} = \cos\Theta\mathbf{i} + \sin\Theta\mathbf{j}$ and $\mathbf{N} = -\sin\Theta\mathbf{i} + \cos\Theta\mathbf{j}$ are positive real multiples of $\operatorname{grad} u = (\operatorname{Re}\overline{\Omega'})\mathbf{i} + (\operatorname{Im}\overline{\Omega'})\mathbf{j}$ and $\operatorname{grad} v = (\operatorname{Im}\Omega')\mathbf{i} + (\operatorname{Re}\Omega')\mathbf{j}$ respectively, so that $\mathbf{T}$ and $\mathbf{N}$ lie in the directions of the induced orientation.

Not every orthogonal net can be realized as level curves of the real and imaginary parts of an analytic function since in general the local function Θ will not be harmonic. Therefore, it is natural to ask what *geometrically* distinguishes those nets with complex potentials from those without. We will see shortly that the existence of a complex potential affects the rates at which the curvature changes along the curves of the net. As a consequence, orthogonal nets with complex potentials have a distinctive "appearance". However, before describing this result in detail we will need to introduce the (local) curvature functions associated to any orthogonal net.

If $\mathbf{r}(s)$ is a parametrization by arclength of a smooth-oriented curve then it can be shown that there exists a smooth "angle of inclination" function $\alpha = \alpha(s)$ such that $\mathbf{r}'(s) = \cos\alpha(s)\mathbf{i} + \sin\alpha(s)\mathbf{j}$ (FIGURE 2a). In terms of α, the plane (signed) curvature function for the curve is defined to be $k = k(s) = \alpha'(s)$. Notice that k is positive where the curve is "turning" counterclockwise and is negative where it is turning clockwise. Although the actual definitions may vary, in most calculus textbooks "curvature" is identified with the absolute value of k. Geometrically, $|k(s)|$ is the reciprocal of the (possibly infinite) radius of the osculating or best approximating circle to the curve at $\mathbf{r}(s)$ (FIGURE 2b). In a very real sense, all the geometrical information about a curve in the plane is contained within the curvature function.

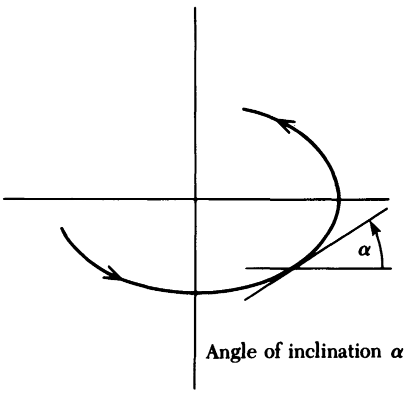


FIGURE 2a

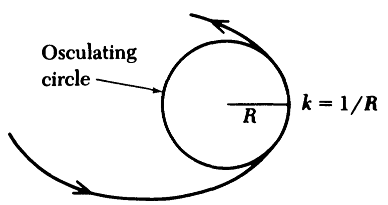


FIGURE 2b

Given a net $N = N(A, B)$, the locally defined vector fields $\mathbf{T}$ and $\mathbf{N}$ induce orientations on the connected components (hereafter, simply "components") of N in V. We wish to compute the curvature functions for each of these oriented curves. Suppose then that $\mathbf{r}(s)$ is an oriented parametrization of a component of A in V. Then $\mathbf{r}'(s) = \mathbf{T}(\mathbf{r}(s))$ so that $\alpha(s) = \Theta(\mathbf{r}(s))$ and $k_A(s) = k(s) = \alpha'(s) = d[\Theta(\mathbf{r}(s))]/ds = D_T\Theta$, where D_T denotes the directional derivative in the direction of $\mathbf{T}$. Likewise, if $\mathbf{r}(s)$ is an oriented parametrization of a component of B in V then $k_B = D_{\mathrm{N}}(\Theta + \pi/2) = D_{\mathrm{N}}\Theta$. Because a unique curve from each family passes through any given point, the curvature functions of the components of A and B in V extend respectively to real-valued functions k_A and k_B on V. Similarly, the derivatives with respect to arclength of the individual curvature functions of the curves in these two families extend to functions k'_A and k'_B defined on V. Note that changing the orientation of a curve changes both the sign of the curvature function and the direction of the arclength derivative. It follows that while the functions k_A and k_B are local functions, defined only up to sign, the functions k'_A and k'_B are actually well-defined functions on all of U.

When does a net possess a (local) complex potential?

We will say that an orthogonal net $N = N(A, B)$ in U is *isothermal* provided that the net possesses a complex potential in a neighborhood of every point. More precisely, for any point P in U, there must exist a complex analytic function Ω defined on a disk V containing P, such that Ω' is never zero in V and such that the components of A and B in V are respectively level curves of the imaginary and real parts of Ω.

PROPOSITION 1. *An orthogonal net $N = N(A, B)$ in U is isothermal if and only if $k'_A + k'_B = 0$.*

Proof. Let P denote any fixed point in U. By assumption there exists a smooth, real-valued function $\Theta = \Theta(x, y)$ defined in a disk V containing P such that the vector fields $\mathbf{T} = \cos\Theta \mathbf{i} + \sin\Theta \mathbf{j}$ and $\mathbf{N} = -\sin\Theta \mathbf{i} + \cos\Theta \mathbf{j}$ are everywhere tangent to the curves of A and B respectively. Then $k_A = D_T\Theta = \mathbf{T}\cdot \operatorname{grad}\Theta = \Theta_x \cos\Theta + \Theta_y \sin\Theta$ and $k_B = D_{\mathbf{N}}\Theta = \mathbf{N}\cdot\operatorname{grad}\Theta = -\Theta_x \sin\Theta + \Theta_y \cos\Theta$. The arclength derivatives are given by $k'_A = D_T k_A = \mathbf{T}\cdot\operatorname{grad} k_A$ and $k'_B = D_{\mathbf{N}} k_B = \mathbf{N}\cdot\operatorname{grad} k_B$ and a straightforward computation reveals that $k'_A + k'_B = \Theta_{xx} + \Theta_{yy}$. If N is isothermal then Θ is a harmonic function and therefore $k'_A + k'_B = 0$.

Conversely, if $k'_A + k'_B = 0$ then Θ is a harmonic function and (because V is a disk) has a harmonic conjugate $-\Gamma$ in V. It follows that the function $g = \exp(-\Gamma - i\Theta)$ is an analytic function in V and possesses an antiderivative $\Omega = \Omega(z) = \Omega(x, y) = u(x, y) + iv(x, y)$. Note that $\operatorname{grad} u = u_x\mathbf{i} + u_y\mathbf{j} = u_x\mathbf{i} - v_x\mathbf{j}$ by virtue of the Cauchy-Riemann equations while $\Omega' = u_x + iv_x = g = \exp(-\Gamma)(\cos\Theta - i\sin\Theta)$. Thus, $\operatorname{grad} u$ is a positive real multiple of $\mathbf{T}$ and u is constant on each of the components of the curves of B in V. Likewise, $\operatorname{grad} v$ is a real multiple of $\mathbf{N}$ and v is consequently constant on each of the components of the curves of A in V. Therefore, Ω is a complex potential for the components of N in V and N is isothermal.

Example 1. Fix an angle $\beta \neq \pi/2$ such that $0 < \beta < \pi$ and let A denote the family of logarithmic spirals with polar equation $r = C_A e^{(\cot\beta)\Theta}$, where C_A can be any nonzero real number. Orient the spirals in A counterclockwise and note that for any curve in A, the angle between the direction of the curve at any point and the outward radial direction at that point is β (FIGURE 3). (Conversely, suppose a curve has the property that the angle counterclockwise from the outward radial direction through any point on the curve to the tangent line to the curve at that point is a constant $\beta \neq 0, \pi/2$. Then it can be shown that the curve is a logarithmic spiral with pole the origin.) By construction A is invariant under both rotations about the origin and dilations with center at the origin. The collection of orthogonal trajectories to A is given by the family, B, of logarithmic spirals with polar equation $r = C_B e^{(\cot(\beta+\pi/2))\Theta} = C_B e^{(-\tan(\beta))\Theta}$ where again C_B can be any nonzero real number.

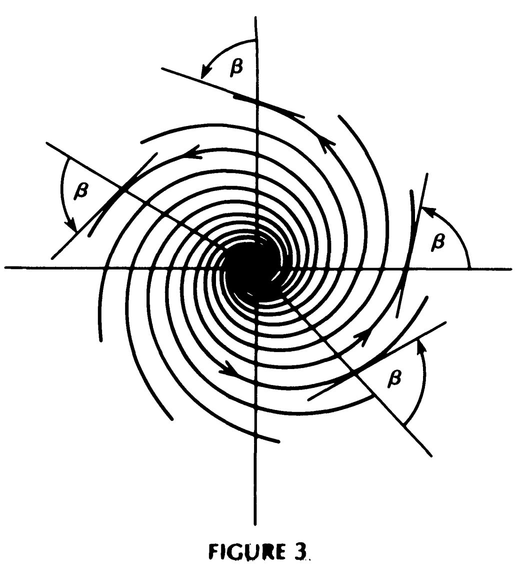
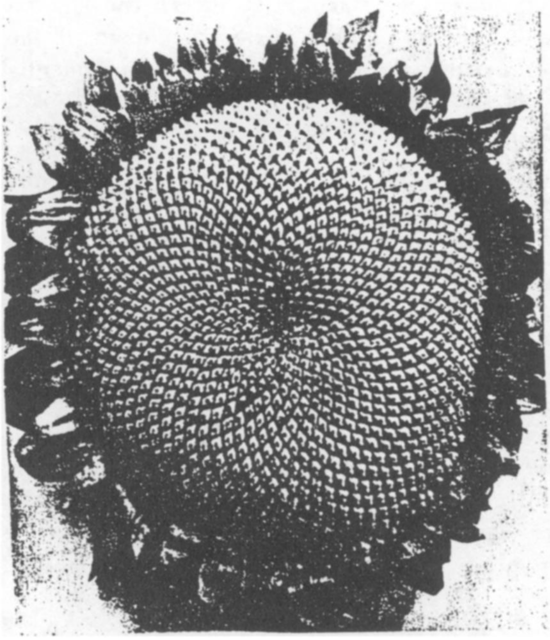


FIGURE 3

Likewise, B is invariant under rotations about the origin and dilations with center the origin. Orient the spirals in B counterclockwise if $0<\beta<\pi/2$ and clockwise otherwise. The invariance of both families under rotations and dilations has the analytical consequence that k_A and k_B are both inversely proportional to r and that k'_A and k'_B are both inversely proportional to r^2. In fact, it is straightforward to verify that $k_A=(\sin\beta)/r$ and $k_B=(\sin(\beta+\pi/2))/r=(\cos\beta)/r=(\cot\beta)k_A$ and that $k'_A=(-\sin 2\beta)/r^2$ and $k'_B=(-\sin 2(\beta+\pi/2))/r^2=(\sin 2\beta)/r^2=-k'_A$. Since $k'_A+k'_B=0$, the net $N=N(A,B)$ is isothermal. (This net recalls Hermann Weyl's description of the head of a sunflower: "... the florets will naturally arrange themselves into logarithmic spirals, two sets of spirals of opposite sense of coiling" [7, p. 70]. Note the striking similarity between Weyl's sunflower (FIGURE 4a) and the net $N=N(A,B)$ obtained by taking $\beta=\pi/4$ (FIGURE 4b).)



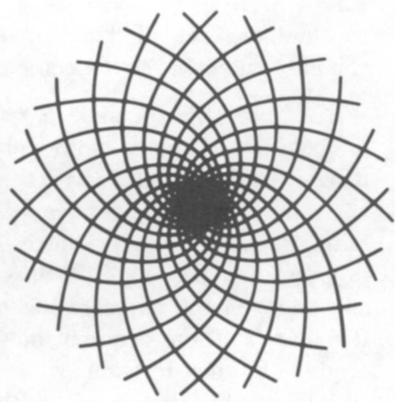


FIGURE 4b

FIGURE 4a
(From Hermann Weyl, *Symmetry*, Princeton Univ. Press, Princeton, NJ, 1952.)

Part of the value of Proposition 1 is that it directly connects the geometry of the net to the requirements for the net to be isothermal. In principle, one should be able to make an informed decision as to whether or not a net is isothermal based solely upon an accurate *picture* of the net.

For example, consider the orthogonal net of ellipses and punctured "parabolas" in FIGURE 5. Let P and Q denote the points indicated in the figure and assume that the ellipse through P and Q is oriented counterclockwise. Since the curvature of the ellipse is greater at P than at Q, the mean-value theorem implies that there exists a point, S, on the portion of the ellipse in the first quadrant, at which the arclength derivative of the curvature is negative. Assume that the arc of the parabola through S has also been oriented counterclockwise and observe that the curvature of the parabola decreases as one moves along this arc away from the origin. Therefore, the arclength derivative of the curvature of the parabola is nonpositive at S. But then, the sum of the respective arclength derivatives is *negative* at S and the net cannot be isothermal.

Other necessary and sufficient conditions for a net to be isothermal are known, but they tend to be formulated in terms of functions whose level curves comprise the net rather than in terms of the net itself. For example, suppose an orthogonal net is given by the respective level curves of two smooth functions f and g. Then the net is isothermal if and only if the ratio of the Laplacian of f to $\|\text{grad } f\|^2$ is a function of f alone [3, Prob. 2.2.16]. Equivalently, the net is isothermal if and only if the ratio of $\|\text{grad } g\|$ to $\|\text{grad } f\|$ is equal to the ratio of a function of f to a function of g [6, pp. 227–228]. In both instances the necessary and sufficient conditions for a net to be isothermal are stated in terms of functions whose level curves comprise the net. Consequently, these conditions may be difficult to verify if it is inconvenient to describe the net by means of level curves or if an unfortunate choice of f or g is made. While the curvature condition of Proposition 1 can be difficult to apply as well, in some examples it is quite easy to check.

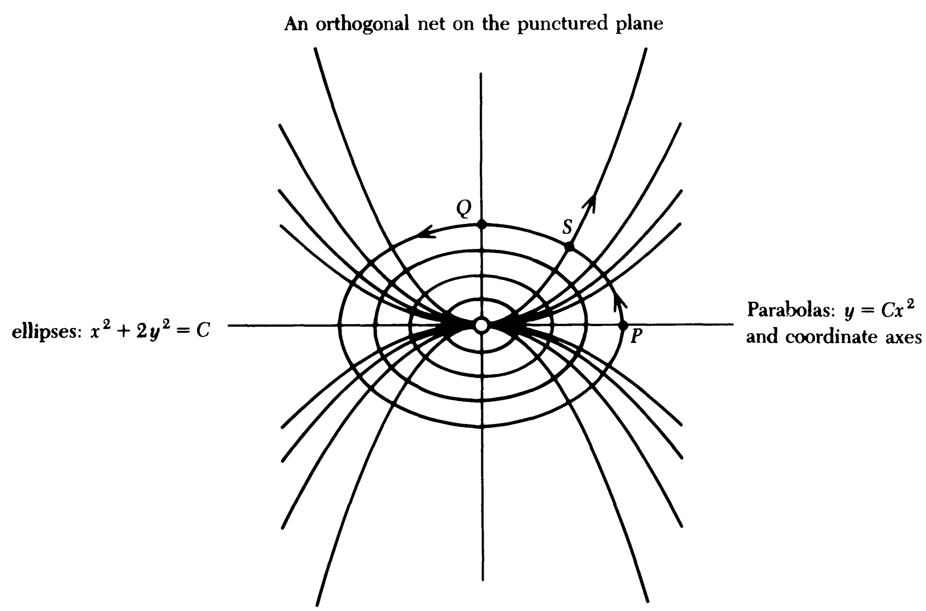


FIGURE 5

As a case in point, recall that a curve in the plane has constant curvature if and only if it is contained in a line or in a circle. It then follows immediately from Proposition 1 that if each curve in family A is either a line (segment) or an arc of a circle then N is isothermal if and only if the same is true for the curves in B. Translated into the context of irrotational, incompressible, planar fluid flow, this means that if the streamlines of the flow are all (subsets of) straight lines or circles then the same is true for the equipotential lines.

These observations may help to explain the ubiquitous appearance of orthogonal families of circles in the treatments of two-dimensional fluid flow and planar electrostatics. It follows from the discussion above that nets comprised of orthogonal families of circles are isothermal and therefore may be given physical interpretations. For example, let U denote the complement of $\{(-1,0),(1,0)\}$ in the plane and let A consist of the restriction to U of the x-axis together with the restriction to U of the family of all circles through $(-1,0)$ and $(1,0)$. The set, B, of orthogonal trajectories to A consists of the y-axis and a collection of nonintersecting circles centered on the x-axis (FIGURE 6) [2, Exercise P-14, p. 231]. It is straightforward to verify that the net $N = N(A, B)$ is smooth and since each curve of the net is a subset of either a line or a circle, the net is also isothermal. Interpreted in terms of fluid flow in the plane, this net represents the streamlines and equipotential lines of a "source" at $(-1,0)$ and a "sink" at $(1,0)$ of equal strengths [5, p. 248]. This net may also be regarded as the lines of force and equipotential lines for a pair of equally and oppositely charged point charges located at $(-1,0)$ and $(1,0)$ [4, pp. 383–384].

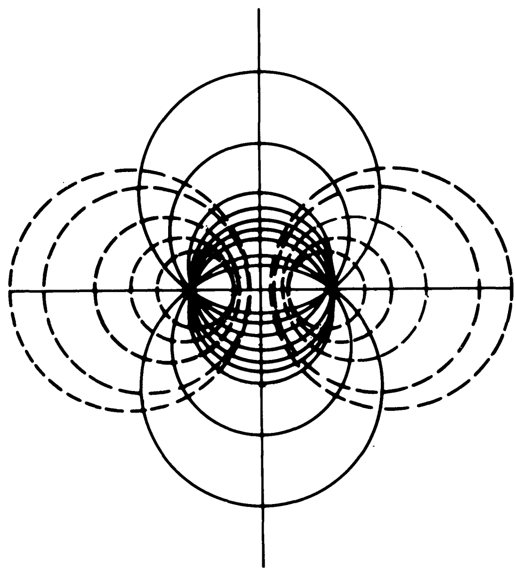


FIGURE 6

Before exploring the geometry of isothermal nets at greater length, we would like to briefly discuss a connection between some of the results of this section and some classical results concerning the geometry of orthogonal families of curves on a *surface*. However, because a detailed description of this connection would take us too far afield, readers interested in the particulars are referred to [1].

Suppose that A and B denote two orthogonal families of curves on a *surface* S. The resulting orthogonal "net" is said to be "isothermal" if there exists a (local) complex-valued function Ω on S such that A and B can be realized as level curves of the imaginary and real parts of Ω respectively and such that Ω is an *angle-preserving* transformation from S to the complex plane. (To say that Ω is angle preserving means that if C_1 and C_2 are any two curves on S intersecting at an angle α, then their images under Ω intersect at angle α as well.) The requirement that Ω be an angle-preserving transformation is the analogue for surfaces of our requirement that Ω be complex analytic. If C is a curve on S then the *geodesic curvature* of C is a function that at each point of the curve measures the rate at which the direction of C is changing "relative to S". (Geodesic curvature reduces to ordinary curvature if S is actually the complex plane. On the other hand, if S is a sphere and if C is a great circle on S then the geodesic curvature of C is identically zero.) The functions k_A, k_B, k'_A and k'_B may be defined almost as before, except that now these functions are defined in terms of the geodesic curvature. In Chapter IV, Section 58 of [1] we find the following theorem:

"When the curves of an orthogonal system have constant geodesic curvature, the system is isothermal."

This theorem is the analogue for surfaces of our observation that if each curve in an orthogonal net is either a line (segment) or the arc of a circle, then the net is isothermal. In fact, a slight modification of the proof of this theorem shows that an orthogonal net on a surface is isothermal if and only if $k'_A + k'_B = 0$. In other words, Proposition 1 (properly interpreted) is valid on a surface as well. On the other hand, it is well known that the only angle-preserving transformations from the complex plane to itself are either complex analytic or the conjugates of complex analytic functions. As a consequence, results from classical surface theory provide us with an additional (although somewhat tortuous) proof of Proposition 1.

Computing the curvature functions for an isothermal net

In the previous section we saw that the behavior of the curvature functions for an orthogonal net determines whether or not the net is isothermal. In the present section we will derive a formula for *computing* the curvature functions of an isothermal net. More precisely, our formula will simultaneously express both curvature functions in terms of the (complex) derivatives of a potential function.

PROPOSITION 2. *Suppose $N = N(A, B)$ is an isothermal net and that $\Omega = \Omega(z) = \Omega(x, y) = u(x, y) + iv(x, y)$ is a complex potential for N. Assume that N has been given the induced orientation. Then,*

$$-|\Omega'|\Omega''/(\Omega')^2 = k_B + ik_A.$$

Proof. Given any point P in U choose a branch of $\log(\Omega')$ defined on a neighborhood of P and (as before) let $\Theta = \operatorname{Im}\log(1/\Omega')$ and let $h = 1/\Omega' = |h|e^{i\Theta}$. We have seen that the vector fields $\mathbf{T} = \cos\Theta\mathbf{i} + \sin\Theta\mathbf{j}$ and $\mathbf{N} = -\sin\Theta\mathbf{i} + \cos\Theta\mathbf{j}$ lie in the directions of the induced orientation so that $k_A = D_T\Theta = \mathbf{T}\cdot\operatorname{grad}\Theta = \Theta_x\cos\Theta + \Theta_x\sin\Theta$ and $k_B = D_N\Theta = \mathbf{N}\cdot\operatorname{grad}\Theta = -\Theta_x\sin\Theta + \Theta_y\cos\Theta$. Let $D_z = (D_x - iD_y)/2$ denote the operation of "complex differentiation", where D_x and D_y denote the operations of partial differentiation with respect to the variables x and y respectively. Notice that $D_z f$ is defined for any smooth complex-valued function $f(z)$ and that $D_z f = f'$ and $D_z\bar{f} = 0$ when f is analytic. Then

$$\begin{aligned} D_z(\cos\Theta + i\sin\Theta) &= D_z\cos\Theta + iD_z\sin\Theta \\ &= (1/2)\big(-\Theta_x\sin\Theta + \Theta_y\cos\Theta + i(\Theta_x\cos\Theta + \Theta_y\sin\Theta)\big) \\ &= (1/2)(k_B + ik_A). \end{aligned}$$

On the other hand, $D_z(\cos\Theta + i\sin\Theta) = D_z(h/|h|) = (D_z h|h| - hD_z|h|)/|h|^2 = (h'|h| - hD_z|h|)/|h|^2$ (since h is analytic).

Now $2|h|D_z|h| = D_z|h|^2 = D_z(h\bar{h}) = D_z h\bar{h} + hD_z\bar{h} = h'\bar{h}$. Consequently, $D_z|h| = (h'\bar{h})/2|h|$ so that $(h'|h| - hD_z|h|)/|h|^2 = (h'|h| - h(h'\bar{h}/(2|h|)))/|h|^2 = h'/(2|h|)$. Therefore,

$$(1/2)(k_B + ik_A) = D_z(\cos\Theta + i\sin\Theta) = h'/(2|h|)$$

or equivalently, $h' = |h|(k_B + ik_A)$. Making the substitutions $|h| = 1/|\Omega'|$ and $h' = -\Omega''/(\Omega')^2$ then yields the formula

$$-|\Omega'|\Omega''/(\Omega')^2 = k_B + ik_A.$$

Proposition 2 provides a convenient means of analyzing the curvature functions of an isothermal net.

Example 2. Suppose that the (nondegenerate) hyperbolas $2xy = a$ and $x^2 - y^2 = b$ intersect at a point $P = (x_0, y_0)$ and let $k_A(P)$ and $k_B(P)$ denote their respective curvatures at P with respect to any choice of orientation. Show that $|k_A(P)/k_B(P)| = |a/b|$.

Solution. We have already observed that the function $\Omega(z) = z^2 = (x + iy)^2 = x^2 - y^2 + i(2xy)$ is a complex potential for an isothermal net containing the two hyperbolas. Assume, without loss of generality, that this net has been given the induced orientation. Then $-|\Omega'|\Omega''/(\Omega')^2 = -|z|/z^2 = -\overline{z}^2/|z|^3 = (y^2 - x^2 + i(2xy))/(x^2 + y^2)^{3/2} = k_B + ik_A$. Therefore, $|k_A(P)/k_B(P)| = |2x_0y_0/(y_0^2 - x_0^2)| = |a/b|$.

As an (almost) immediate corollary of Proposition 2, we have a formula for the curvature functions of the x and y parameter curves of an analytic mapping.

Let φ denote an analytic mapping with nonvanishing derivative and suppose that k_1 and k_2 denote the respective curvature functions of the x and y parameter curves oriented in the direction of increasing parameter. As an exercise the reader may show that

$$\varphi''/(\varphi'|\varphi'|) = k_2 + ik_1.$$

Example 3. The x and y parameter curves of the exponential mapping $\varphi(z) = e^z = e^{x+iy} = e^x e^{iy}$ are respectively rays to the origin and circles centered at the origin of radius e^x. Since $\varphi''/(\varphi'|\varphi'|) = 1/e^x$, we see that $k_1 = 0$ and $k_2 = 1/e^x$ as expected.

Isothermal nets with curvature functions in constant ratio

Isothermal nets $N = N(A, B)$ for which each curve in A is a line (segment) are particularly easy to characterize. Let us suppose that N is isothermal, that U is connected, and that every curve in A is either a line or a line segment. Because N is isothermal, each curve in B is then contained in either a line or a circle. A little thought shows that since U is connected, there are only two ways for this to occur. Either A and B are the restrictions to U of a pair of orthogonal collections of parallel straight lines, or there exists a choice of origin such that A and B consist of the restrictions to U of the family of radial lines in the plane and the family of concentric circles centered at the origin respectively. (This result will also follow from Proposition 3 below.) In either case, k_A is identically zero and is thus a constant multiple of k_B.

More generally, one can ask what isothermal nets $N = N(A, B)$ have the property that for any pair of the locally defined functions k_A and k_B, (at least) one is a constant multiple of the other. We have already seen that the net of two orthogonal families of logarithmic spirals with a common pole provides a further example of such an isothermal net. Our next result shows that these examples essentially exhaust the class of isothermal nets for which one of the two functions k_A and k_B is a constant multiple of the other.

PROPOSITION 3. *Suppose $N = N(A, B)$ is an isothermal net on a connected open set U. Assume that for any local orientation of the curves in A and B that one of the two functions k_A and k_B is a constant multiple of the other. Then A and B are the restrictions to U of one of the following orthogonal families*:

(a) *A family of parallel straight lines together with the corresponding orthogonal family of parallel straight lines.*

(b) *A family of radial lines (for some choice of origin) and the corresponding orthogonal family of concentric circles centered at the origin.*

(c) *Two orthogonal families of logarithmic spirals having a common pole.*

Proof. Assume that Ω is a complex potential for the components of N restricted to some connected neighborhood and let us retain the orientations, terminology, and notation used in the proof of Proposition 2. Since k_A and k_B are proportional, it follows from the equation $h' = |h|(k_B + ik_A)$ that the image of h' is contained in a line through the origin. However, h' is analytic and thus must be constant. Therefore, $h(z) = z_0 z + w_0$ for some fixed complex numbers z_0 and w_0. Furthermore, if z_0 is nonzero, we may scale h by the factor $\pm 1/|z_0|$ and still retain the property that the vector fields $\mathbf{T}$ and $(\operatorname{Re} h)\mathbf{i} + (\operatorname{Im} h)\mathbf{j}$ are parallel. Consequently, our analysis has shown that in a neighborhood of any point, either

(a) there exists a complex number w_0 such that $(\operatorname{Re} w_0)\mathbf{i} + (\operatorname{Im} w_0)\mathbf{j}$ is tangent to each of the curves of A in the neighborhood; or

(b) there exists a unique linear function $h(z) = e^{\beta i}z + w_0$ with $0 \leqslant \beta < \pi$ such that the vector field $(\operatorname{Re} h)\mathbf{i} + (\operatorname{Im} h)\mathbf{j}$ is tangent to the curves in A.

Since U is connected, a straightforward argument shows that either (a) or (b) must hold in the *entire* set U. If the former, then A consists of a collection of subsets of a family of straight lines parallel to $(\operatorname{Re} w_0)\mathbf{i} + (\operatorname{Im} w_0)\mathbf{j}$ with B a collection of subsets of the corresponding orthogonal family of lines. If the latter, then by translating the origin of the complex plane to the point $-w_0 e^{-\beta i}$ we may assume that $h(z) = e^{\beta i}z$. At any point in this translated coordinate system the angle counterclockwise between the outward radial direction at a point and the vector $(\operatorname{Re} h)\mathbf{i} + (\operatorname{Im} h)\mathbf{j}$ at that point is the constant β. Therefore, if $\beta \neq 0, \pi/2$ then A and B are the restrictions to U of two orthogonal families of logarithmic spirals with pole the origin. If $\beta = 0$ then A and B are the restrictions to U of the family of radial lines through the origin and the family of concentric circles centered at the origin respectively. The reverse is the case if $\beta = \pi/2$.

REFERENCES

1. L. P. Eisenhart, *A Treatise on the Differential Geometry of Curves and Surfaces*, Dover Publications, Inc., Mineola, NY, 1960.
2. M. J. Greenberg, *Euclidean and Non-Euclidean Geometries, Development and History*, 2nd edition, W. H. Freeman and Co., New York, 1980.
3. Jan G. Krzyż, *Problems in Complex Variable Theory*, Elsevier Science Publishing Co., New York, 1971.
4. J. E. Marsden, M. J. Hoffman, *Basic Complex Analysis*, 2nd edition, W. H. Freeman and Co., New York, 1987.
5. M. R. Spiegel, *Theory and Problems of Complex Variables*, (Schaum's Outline Series) McGraw-Hill Book Co., New York, 1964.
6. C. E. Weatherburn, *Differential Geometry of Three Dimensions*, Vol. 1, Cambridge Univ. Press, New York, 1961.
7. Hermann Weyl, *Symmetry*, Princeton Univ. Press, Princeton, NJ, 1952.

Answers to Before the Conquest: $11D_{20}8d_5$; 1.19.6.16

NOTES

A Problem of Pólya

ROSS HONSBERGER
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

(a) *The problem.* The sequence

$$\{a_1, a_2, \dots\} = \{1, -\tfrac{1}{2}, -\tfrac{1}{2}, 2^{-1/3}, -\tfrac{1}{2}\cdot 2^{-1/3}, -\tfrac{1}{2}\cdot 2^{-1/3}, \dots\},$$

whose terms are described nicely in groups of three, and whose nth triple of terms is $n^{-1/3}, -\frac{1}{2}\cdot n^{-1/3}, -\frac{1}{2}\cdot n^{-1/3}$, has the curious property that

> the series of its first powers *converges* (i.e., $a_1 + a_2 + a_3 + \cdots$),
> while the series of its 3rd powers *diverges* ($a_1^3 + a_2^3 + a_3^3 + \cdots$),
> whereas the series of its 5th powers *converges* ($a_1^5 + a_2^5 + a_3^5 + \cdots$),
> as does the series of its 7th powers *converge* ($a_1^7 + a_2^7 + \cdots$);
>
> in fact, the series of kth powers *converges*
> for all *odd* positive integers $k > 3$.

Again, the sequence $\{a_1, a_2, \dots\} = \{-1, -1, -1, -1, -1, \frac{1}{2}, \frac{1}{2}, \dots\}$, whose terms are conveniently lumped together in groups of 150, and whose nth group contains the 150 terms

$$-n^{-1/5}(5 \text{ times}),\ \tfrac{1}{2}\cdot n^{-1/5}\ (64 \text{ times}), \qquad \text{and} \qquad -\tfrac{1}{3}\cdot n^{-1/5}\ (81 \text{ times}),$$

yields a convergent series of kth powers, $a_1^k + a_2^k + a_3^k + \cdots$, for all odd positive integers k *except* 5, when it diverges.

In 1944, Problem 4142 in the *American Mathematical Monthly*, posed by George Pólya, asked for a general demonstration of the

> existence of a sequence of real numbers $\{a_1, a_2, \dots\}$, the series of whose kth powers, $a_1^k + a_2^k + \cdots$, converges for all *odd* positive integers k except for a single stipulated value w.

Beyond this, a proof of the following generalization was also requested:

> if M is *any* subset whatever of $\{1, 3, 5, 7, \dots\}$ there exists a corresponding *sequence of functions of* k,
>
> $$\{a_i(k)\} = \{a_1(k), a_2(k), \dots\},$$
>
> the series of whose kth powers, $a_1(k)^k + a_2(k)^k + \cdots$, yields a convergent sequence for all odd positive integers k except when $k \in M$, in which case the resulting series diverges.

The following brilliant solution by Nathan Fine was published in May 1946 (pages 283–284).

(b) *Fine's solution.*

1. The First Property Our first task, then, is to construct a sequence $\{a_1, a_2, \dots\}$ for which the series $a_1^k + a_2^k + \cdots$ converges for all odd positive integers k except for a prescribed value w, and which diverges for $k = w$.

(i) *The Construction.* Professor Fine proposes to construct the sequence in blocks of m terms at a time, where the constant m arises naturally within the construction. The nth block of terms is built around the number $n^{-1/w}$; in fact, each term in this block is of the form $A \cdot n^{-1/w}$. Denoting $(w+1)/2$ by r, a set of r coefficients $A_1, A_2, \dots, A_r$ is determined, and the nth block is assembled by gathering a prescribed number k_i of the respective terms $A_i n^{-1/w}$. The same frequencies $k_1, k_2, \dots, k_r$, and the same coefficients $A_1, A_2, \dots, A_r$ are used throughout; only n changes as we go from block to block. Thus, for all $n = 1, 2, \dots$, the nth block is given by

$$\begin{array}{ll} A_1 n^{-1/w}, A_1 n^{-1/w}, \dots, A_1 n^{-1/w}, & (k_1 \text{ times}) \\ A_2 n^{-1/w}, A_2 n^{-1/w}, \dots, A_2 n^{-1/w}, & (k_2 \text{ times}) \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ A_r n^{-1/w}, A_r n^{-1/w}, \dots, A_r n^{-1/w}, & (k_r \text{ times}). \end{array}$$

The number of terms in each block, then, is simply the sum

$$m = k_1 + k_2 + \cdots + k_r.$$

Of course, the key to everything lies in choosing suitable values for the coefficients A_i and the frequencies k_i. In order to meet our requirements, these numbers must satisfy the following system $(*)$ of r equations:

$$\begin{cases} \displaystyle\sum_{i=1}^{r} k_i A_i^k = 0 & \text{for each of the } r-1 \text{ values } k = 1, 3, 5, \dots, w-2, \\ \displaystyle\sum_{i=1}^{r} k_i A_i^w = \Delta, & \text{for } k = w, \text{ where } \Delta \text{ is a nonzero quantity to be specified in due course.} \end{cases} \qquad (*)$$

This system of r equations contains $2r$ unknowns. Since k_i denotes the number of times the term $A_i n^{-1/w}$ is to occur in a block, each k_i must be a *positive integer*. Beyond this, there are no conditions on the solutions except that they must be real numbers; thus the A_i can almost be specified arbitrarily. A set of randomly chosen A_i, however, throws the burden of satisfying the equations squarely on the k_i, and it is unlikely that their values will turn out to be positive integers. Accordingly, let us select values for the A_i and Δ that promote positive integral solutions for the k_i. Later on we shall see the need to have the A_i bounded, and so let us confine our choices to real numbers with magnitudes not exceeding unity: $|A_i| \leqslant 1$.

Since it is a simple matter to adjust *rational* values of the k_i to the desired integral solutions, let the A_i also be restricted to rational values. Recapitulating, we are faced with the problem of assigning suitable values to the A_i and Δ so as to obtain positive integral solutions in the resulting system $(*)$ of r equations in the r unknowns k_i.

$$\begin{cases} k_1 A_1 + k_2 A_2 + \cdots + k_r A_r = 0 \\ k_1 A_1^3 + k_2 A_2^3 + \cdots + k_r A_r^3 = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ k_1 A_1^{w-2} + k_2 A_2^{w-2} + \cdots + k_r A_r^{w-2} = 0 \\ k_1 A_1^w + k_2 A_2^w + \cdots + k_r A_r^w = \Delta \quad (\neq 0). \end{cases} \qquad (*)$$

Clearly, rational values of the A_i and Δ will generate rational solutions for the k_i in this *linear* system. Therefore let Δ be taken as the determinant of coefficients of the system itself:

$$\Delta = \begin{vmatrix} A_1 & A_2 & \cdots & A_r \\ A_1^3 & A_2^3 & \cdots & A_r^3 \\ \cdots & \cdots & \cdots & \cdots \\ A_1^w & A_2^w & \cdots & A_r^w \end{vmatrix}.$$

Rational A_i will certainly lead to a rational value of Δ, and if we are careful to avoid choosing two A_i to have the same magnitude, the resulting Δ will also be nonzero. Let us suppose, then, that we have selected appropriate values for A_i and have a system $(*)$ of equations at hand.

Since the system is linear and $\Delta \neq 0$, the solutions k_i will at least be rational numbers. If D is the common denominator of all the fractions among the values of the k_i, then, clearing of fractions in the k_i by multiplying each equation by D, *leaving the factors* A_i^k *unaffected*, we obtain a new system

$$\sum_{i=1}^{r} \bar{k}_i A_i^k = \sum_{i=1}^{r} (Dk_i) A_i^k = 0 \quad \text{for } k = 1,3,\ldots,w-2,$$

$$\sum_{i=1}^{r} \bar{k}_i A_i^w = \sum_{i=1}^{r} (Dk_i) A_i^w = D\Delta \quad (\neq 0),$$

which has the same *form* as the original system $(*)$ but which has *integral* solutions for the $\bar{k}_i$.

Of course, it is possible that $\bar{k}_i$ might be negative. However, since $\bar{k}_i A_i^k = (-\bar{k}_i)(-A_i)^k$, because k is always *odd*, we can reverse the sign of any negative $\bar{k}_i$ provided we also change the sign of the corresponding A_i. Denoting the final results again by just k_i and A_i, we have thus arrived at two sets of real numbers that satisfy the system of equations

$$\sum_{i=1}^{r} k_i A_i^k = 0 \quad \text{for } k = 1,3,5,\ldots,w-2,$$

$$\sum_{i=1}^{r} k_i A_i^w = D\Delta \quad (\neq 0),$$

where the k_i are positive integers and the A_i are rational numbers such that $|A_i| \leqslant 1$, with $|A_i| \neq |A_j|$ for $i \neq j$. With such sets of numbers the desired sequence $\{a_1, a_2, \ldots\}$ can now be constructed in blocks of size $m = k_1 + k_2 + \cdots + k_r$ as prescribed above.

(ii) *An Example.* In order to set these ideas, let's work through an example. Suppose $w = 5$. Then $r = (5+1)/2 = 3$, and if we take $A_1 = 1$, $A_2 = \frac{1}{2}$, and $A_3 = \frac{1}{3}$, we initially obtain the system

$$k_1 + \frac{1}{2}k_2 + \frac{1}{3}k_3 = 0$$

$$k_1 + \left(\frac{1}{2}\right)^3 k_2 + \left(\frac{1}{3}\right)^3 k_3 = 0$$

$$k_1 + \left(\frac{1}{2}\right)^5 k_2 + \left(\frac{1}{3}\right)^5 k_3 = \begin{vmatrix} 1 & 1/2 & 1/3 \\ 1 & 1/8 & 1/27 \\ 1 & 1/32 & 1/243 \end{vmatrix} = -\frac{5}{324}.$$

Solving we get $k_1 = -5/216$, $k_2 = 8/27$, $k_3 = -3/8$, and we have

$$-\frac{5}{216}(1)+\frac{8}{27}\left(\frac{1}{2}\right)-\frac{3}{8}\left(\frac{1}{3}\right)=0$$

$$-\frac{5}{216}(1)^3+\frac{8}{27}\left(\frac{1}{2}\right)^3-\frac{3}{8}\left(\frac{1}{3}\right)^3=0$$

$$-\frac{5}{216}(1)^5+\frac{8}{27}\left(\frac{1}{2}\right)^5-\frac{3}{8}\left(\frac{1}{3}\right)^5=-\frac{5}{324}.$$

The common denominator of the fractions among the k_i is $D = 216$, and, clearing of fractions and reversing the necessary signs all in one step, we obtain

$$5(-1)+64\left(\frac{1}{2}\right)+81\left(-\frac{1}{3}\right)=0$$

$$5(-1)^3+64\left(\frac{1}{2}\right)^3+81\left(-\frac{1}{3}\right)^3=0$$

$$5(-1)^5+64\left(\frac{1}{2}\right)^5+81\left(-\frac{1}{3}\right)^5=-\frac{10}{9}\quad(\neq 0).$$

Thus $k_1 = 5$, $k_2 = 64$, $k_3 = 81$, $A_1 = -1$, $A_2 = \frac{1}{2}$, $A_3 = -\frac{1}{3}$, and the nth block, consisting of $5 + 64 + 81 = 150$ terms, starts with five terms equal to $-n^{-1/5}$, continues with 64 terms equal to $\frac{1}{2}n^{-1/5}$, and finishes with 81 terms equal to $-\frac{1}{3}n^{-1/5}$.

It remains, then, in this section, to prove in general that the series of kth powers of a sequence thus constructed does indeed converge for all odd positive integers k except $k = w$.

(iii) *The Proof.* The kth power of a typical term $A_i n^{-1/w}$ is $A_i^k n^{-k/w}$, and the sum of the kth powers of all the terms in the nth group is

$$\begin{aligned}
&A_1^k n^{-k/w} + A_1^k n^{-k/w} + \cdots \quad (k_1 \text{ times})\\
&+ A_2^k n^{-k/w} + A_2^k n^{-k/w} + \cdots \quad (k_2 \text{ times})\\
&\cdots\\
&+ A_r^k n^{-k/w} + A_r^k n^{-k/w} + \cdots \quad (k_r \text{ times})\\
&= n^{-k/w}\left(k_1 A_1^k + k_2 A_2^k + \cdots + k_r A_r^k\right)\\
&= n^{-k/w} \sum_{i=1}^{r} k_i A_i^k.
\end{aligned}$$

(a) *The Case of $k < w$.* From the system $(*)$ of equations, we know that this sum is zero for all $k = 1, 3, \ldots, w-2$, and so for these values of k, a partial sum s_t of t terms of the series of kth powers reduces to the value of its final group of m or fewer terms:

$$\begin{aligned}
s_t &= 0 + 0 + \cdots + 0 + \left(A_1^k n^{-k/w} + \cdots\right)\\
&= n^{-k/w}\left(A_1^k + A_1^k + \cdots \text{ as far as it goes}\right).
\end{aligned}$$

Since $|A_i| \leqslant 1$ and the fact that there are not more than m terms in this final group, we have

$$|s_t| \leqslant n^{-k/w}(m), \quad \text{where } n \text{ is the only variable on the right side.}$$

Therefore, as t and n grow beyond all bounds, it follows that

$$|s_t| \to 0, \quad \text{forcing } \lim_{t\to\infty} s_t = 0,$$

and we conclude that the series $a_1^k + a_2^k + \cdots$ does indeed converge, and we have as a little bonus that its sum is 0 in each of these cases.

(b) *The Case of* $k = w$. When $k = w$, we have $\Sigma_{i=1}^r k_i A_i^k$ equal to the nonzero quantity $D\Delta$. In this case, then, the sum of the kth powers in the nth group is

$$n^{-w/w}(D\Delta) = D\Delta\left(\frac{1}{n}\right),$$

and a partial sum

$$s_t = D\Delta\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1}\right) + \frac{1}{n}\left(A_1^w + \cdots\right).$$

Again, since $|A_i| \leqslant 1$, the remnant in the final group cannot alter the earlier part by a magnitude exceeding m/n, and since $D\Delta \neq 0$, we have, as t and n increase, that s_t diverges with the harmonic series.

(c) *The Case of* $k > w$. For $k > w$ we have no knowledge of the value of $\Sigma_{i=1}^r k_i A_i^k$. However, with $k/w = p > 1$, we can still show that the series of kth powers is absolutely convergent as follows.

The sum of the *absolute values* of the kth powers of the terms in the nth group is

$$n^{-k/w}\sum_{i=1}^r k_i|A_i^k| = n^{-p}\sum_{i=1}^r k_i|A_i|^k \leqslant \frac{1}{n^p}(m), \quad \text{since } |A_i| \leqslant 1.$$

Thus a partial sum of the series of absolute values of the kth powers is

$$s_t \leqslant \frac{m}{1^p} + \frac{m}{2^p} + \cdots + \frac{m}{n^p} \quad \text{for some positive integer } n,$$
$$= m\left(1 + \frac{1}{2^p} + \cdots + \frac{1}{n^p}\right).$$

Now it is well known that the so-called p-series $\Sigma_{n\geqslant 1}(1/n^p)$ converges to a positive limit L whenever $p > 1$. Accordingly, s_t is bounded by mL, and since $\{s_t\}$ is a sequence of positive numbers (the absolute values), we conclude that $\lim_{t\to\infty} s_t$ exists, making $a_1^k + a_2^k + \cdots$ absolutely convergent.

Also, because k and w are both odd, the condition $k > w$ implies that $k \geqslant w + 2$. In this case, we can say with certainty that $k > w + 1$, and therefore that

$$\frac{k}{w} > 1 + \frac{1}{w}.$$

Setting $1 + 1/w = q$, we have $p > q > 1$.

Accordingly, the series $\Sigma_{n\geqslant 1}(1/n^q)$ converges to some positive number Q. Since $p > q$, then the convergent series

$$\sum_{n\geqslant 1}\frac{1}{n^p} \leqslant \sum_{n\geqslant 1}\frac{1}{n^q} = Q,$$

implying that the series $\Sigma_{n\geqslant 1}(1/n^p)$ must converge to some positive number $L \leqslant Q$. Therefore we see that, for *all* values of $k > w$, the partial sum

$$s_t \leqslant m\left(1 + \frac{1}{2^p} + \cdots + \frac{1}{n^p}\right) \leqslant mL \leqslant mQ.$$

That is to say, s_t is always bounded by the fixed constant mQ no matter what the value of $k > w$. For all $k > w$, then, we again have that

$$\lim_{t \to \infty} s_t \quad \text{exists (and is} \leqslant mQ),$$

and that $a_1^k + a_2^k + \cdots$ must converge to a sum S of magnitude not exceeding mQ. In more sophisticated language, we have shown that the series of kth powers converges absolutely and *uniformly* for $k > w$.

2. Arbitrary Subsets M We are now in a position to deal with arbitrary subsets $M = \{w_1, w_2, \dots\}$ of $\{1, 3, 5, \dots\}$.

From our previous considerations, we have that, corresponding to the single number w_1, there is a sequence

$$S_{w_1} = \{a_{11}, a_{12}, a_{13}, \dots\}$$

such that the series of its kth powers

$$S_{w_1}(k) = a_{11}^k + a_{12}^k + a_{13}^k + \cdots$$

converges for every odd positive integer except $k = w_1$. As k runs through the set $\{1, 3, 5, \dots\}$, an infinite collection of convergent series is obtained:

$$S_{w_1}(1), S_{w_1}(3), \dots S_{w_1}(w-2), S_{w_1}(w+2), \dots,$$

which, of course, does not contain the divergent series $S_{w_1}(w_1)$. Each of these series has its own infinity of partial sums s_t. However, the partial sums of a *convergent* series cannot accrue to infinite proportions, but must always remain within some finite limits, however large they may be. Thus for each $S_{w_1}(k)$, $k \neq w_1$, there is some bound b on the magnitudes of its partial sums, that is, such that $|s_t| \leqslant b$ for all values of t.

Now, we would like to determine a *single* value C_1 that can serve as a bound on the magnitudes of the partial sums of *all* the series $S_{w_1}(k)$, $k \neq w_1$. Such a single value may not exist for an arbitrary collection of series, but we are in luck because the sequences $S_{w_1}(k)$ are absolutely and uniformly convergent for the infinite number of cases in which $k > w_1$. Thus there is a single value mQ that bounds the magnitudes of all the partial sums s_t generated by any series $S_{w_1}(k)$ with $k > w_1$. Since there are only a finite number of other bounds to consider, namely those that govern the partial sums of $S_{w_1}(1), S_{w_1}(3), \dots, S_{w_1}(w-2)$, the greatest of this finite collection of bounds will be a universal bound C_1 on the magnitudes of all the partial sums in question. In summary, then, no matter what odd positive integer is substituted for k in the series $S_{w_1}(k)$, other than w_1 itself, no partial sum of the resulting series will have a magnitude exceeding C_1. Consequently, if each term of $S_{w_1}(k)$, $k \neq w_1$, were to be divided by $2C_1$, the magnitudes of the partial sums of the resulting series $(1/2C_1)S_{w_1}(k)$ would never exceed $\frac{1}{2}$.

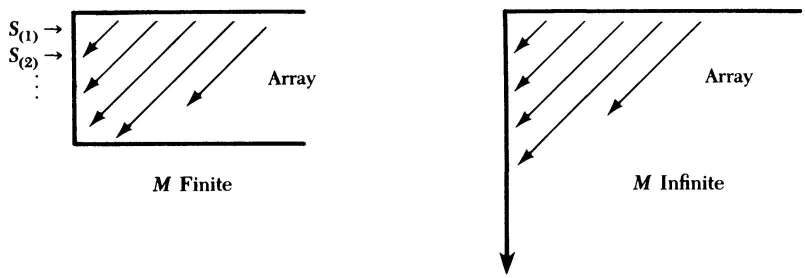
Now suppose that $C_1, C_2, \dots, C_i, \dots$ are universal bounds on the magnitudes of the partial sums of the respective collections of series represented by $S_{w_1}(k), S_{w_2}(k), \dots, S_{w_i}(k), \dots$, $k \neq w_i$ in each case. If each term of the series $S_{w_i}(k)$ were to be divided by $2^i C_i$, then the magnitudes of the partial sums of the resulting series $(1/2^i C_i)S_{w_i}(k)$, $k \neq w_i$, would never amount to more than $1/2^i$. Let each $S_{w_i}(k)$ be treated in this manner and the resulting series arranged one below the other, as shown. Of course, the ith series still converges for $k \neq w_i$ and diverges for $k = w_i$. Finally, the long-sought series of kth powers $S_M(k)$ is constructed from this array by gathering the terms along the diagonals, in order from the left, moving down each diagonal from the upper right.

$$
\begin{array}{ll}
\left(\frac{1}{2C_1}S_{w_1}(k)\right): & \frac{1}{2C_1}a_{11}^k + \frac{1}{2C_1}a_{12}^k + \frac{1}{2C_1}a_{13}^k + \cdots \\
\left(\frac{1}{4C_2}S_{w_2}(k)\right): & \frac{1}{4C_2}a_{21}^k + \frac{1}{4C_2}a_{22}^k + \cdots \\
\left(\frac{1}{8C_3}S_{w_3}(k)\right): & \frac{1}{8C_3}a_{31}^k + \frac{1}{8C_3}a_{32}^k + \cdots \\
\quad\cdots & \quad\cdots \\
\left(\frac{1}{2^iC_i}S_{w_i}(k)\right): & \frac{1}{2^iC_i}a_{i1}^k + \frac{1}{2^iC_i}a_{i2}^k + \cdots \\
 & \quad\cdots
\end{array}
$$

$$
S_M(k) = \left(\frac{1}{2C_1}a_{11}^k\right) + \left(\frac{1}{2C_1}a_{12}^k + \frac{1}{4C_2}a_{21}^k\right) + \left(\frac{1}{2C_1}a_{13}^k + \frac{1}{4C_2}a_{22}^k + \frac{1}{8C_3}a_{31}^k\right) + \cdots .
$$

A partial sum s_t of this series $S_M(k)$ clearly is composed of partial sums $s_{(i)}$, of diminishing lengths, of the series in the array:

$$s_t = s_{(1)} + s_{(2)} + \cdots .$$



Now, each series in the array converges for all odd positive integers k with a single exception, namely w_i. Therefore a particular integer k can be the exceptional value for at most one of these series. For each value of k, then, *at most one* of the series in the array can fail to be convergent.

(a) *For finite M.* For finite subsets M, then, all but at most one of the series in the array converge for a given k, and the partial sums of a convergent series approach arbitrarily close to the sum of that series. Altogether, as t gets indefinitely large, these $|M| - 1$ convergent series contribute toward the partial sum s_t of $S_M(k)$ an amount which is only negligibly different from the total sum S' of these $|M| - 1$ series. Clearly, then, if the $|M|$th series also converges, which will be the case when k does *not* belong to M, the values of s_t will approach as limit the total sum of all $|M|$ of the series:

$$s_t = s_{(1)} + s_{(2)} + \cdots$$

implies

$$\lim_{t \to \infty} s_t = \lim s_{(1)} + \lim s_{(2)} + \cdots .$$

On the other hand, if k belongs to M; then the final series will diverge, and this would be reflected in the values of s_t (which would either go to $\pm\infty$ or oscillate, as the case may be, since the total contributions of the other $|M|-1$ series still approach the value S'). Hence we conclude that, when M is finite, $S_M(k)$ diverges or converges as k belongs or does not belong to M.

(b) *For infinite M.* As observed above, as t grows beyond all bounds, a partial sum s_t of $S_M(k)$ is simply a sum of partial sums $S_{(i)}$ of the series in the array:

$$s_t = S_{(1)} + S_{(2)} + \cdots + S_{(i)} + \cdots;$$

and so we have

$$\begin{aligned}\lim_{t\to\infty} s_t &= \lim S_{(1)} + \lim S_{(2)} + \cdots \\ &= \sum_{i\geqslant 1} \lim S_{(i)}.\end{aligned}$$

In the case of an *infinite* subset M, our array would contain an infinity of series and the series on the right side here would also be infinite. In order to yield an acceptable value for $\lim s_t$, then, this series must converge for the value of k in question.

As we have noted, all but at most one of the series in the array will converge for a given k. In the event that k belongs to M, the array will contain a divergent series $S_{w_i}(w_i)$, whose partial sums will not converge, thus causing $\lim S_{(i)}$ to be undefined for some value of i and preventing the convergence of $\Sigma \lim S_{(i)}$. In this case, then, $\lim s_t$ does not exist and $S_M(k)$ diverges.

If k does not belong to M, however, the array contains only convergent series and the series $\Sigma \lim S_{(i)}$ is at least defined; whether it converges or not is what needs to be settled. We have set things up, however, so that it is a simple matter to show that it is, in fact, absolutely convergent.

If the ith series in the array converges, then the magnitudes of its partial sums are bounded by $1/2^i$. In the case at hand, all the series in the array converge, and so every $|S_{(i)}|$ is bounded by $1/2^i$. This means that, for all i, $S_{(i)}$ always lies between $\pm 1/2^i$, implying that $\lim S_{(i)}$, which is known to exist, must also lie between $\pm 1/2^i$. Thus, for all i, we have $|\lim S_{(i)}| \leqslant 1/2^i$, and it follows easily that $\Sigma \lim S_{(i)}$ is absolutely convergent:

$$\begin{aligned}\sum |\lim S_{(i)}| &\leqslant \sum 1/2^i \\ &= \tfrac{1}{2} + \tfrac{1}{4} + \cdots \\ &= 1.\end{aligned}$$

Thus $\lim s_t$ exists and we conclude that $S_M(k)$ indeed converges for k not in M.

Finally, by taking the kth roots of the terms of $S_M(k)$, we obtain the desired sequence of functions of k

$$\{a_i(k)\} = \left\{(2C_1)^{-1/k} a_{11}, (2C_1)^{-1/k} a_{12}, (4C_2)^{-1/k} a_{21}, \cdots\right\}.$$

Nice Cubic Polynomials, Pythagorean Triples, and the Law of Cosines

JIM BUDDENHAGEN
Southwestern Bell Telephone Co.
Saint Louis, MO 63101

CHARLES FORD
MIKE MAY, S.J.
Saint Louis University
Saint Louis, MO 63103

Problems This paper arose from an attempt to find examples suitable for a beginning calculus class that illustrate the techniques of sketching graphs of polynomials. An excellent example for this purpose is the cubic polynomial

$$y = 2x^3 + 5x^2 - 4x - 3$$

because both $y = (x+3)(2x+1)(x-1)$ and $y' = 2(x+2)(3x-1)$ factor over the integers. Our attempt to give a systematic description of all such 'nice' cubics led to Diophantine equation (2) below, whose solution is well known. Our techniques give systematic methods for producing nice numbers for other problems as well. Although these problems have been investigated by several authors, [2], [6], [7], [8], [10], [13], the solutions have not been given a systematic and unified presentation. We attempt to do so here, in a way that allows the reader to construct additional examples at will.

The first problem we are interested in is:

Given a cubic polynomial, find the three roots, the two critical points, and sketch the graph.

We want the roots and critical points to have rational x-coordinates (with small denominators). If a cubic with rational roots has a double or triple root its derivative will necessarily have rational roots. Thus we assume that the cubic has distinct roots. By multiplication we can assume that the roots are integers, and by translation that the middle root is zero. Thus we are investigating polynomials of the form

$$y = (x+a)x(x-b). \tag{1}$$

We are looking for nonnegative integers a and b so that the derivative $y' = 3x^2 + 2(a-b)x - ab$ has rational roots. This happens when the discriminant is a perfect square. This leads to the Diophantine equation

$$a^2 + b^2 + ab = c^2. \tag{2}$$

If, instead, we make zero the leftmost root, we proceed from

$$y = x(x-a)(x-b)$$

to the equation

$$a^2 + b^2 - ab = c^2. \tag{3}$$

The second problem is the box problem:

An open box is constructed from a rectangular piece of metal by cutting four equal squares from the corners and bending up the resulting tabs. Find the dimensions which maximize the volume of the box.

We want the given dimensions, a and b, of the rectangle to be positive integers and the maximum volume to be attained at a rational value for the edge length, x, of the squares. In this case we proceed from the equation

$$y = x(a-2x)(b-2x)$$

to equation (3) above.

The third problem reaches back to trigonometry:

Using the law of cosines, find the length of the third side of a triangle given the lengths of two sides and the cosine of the included angle γ.

We want the lengths of the sides of the triangle, a, b, and c, to be integers. This requires that $\cos\gamma$ be rational. If we let $g = -2\cos\gamma$, this reduces to the equation

$$a^2 + b^2 + gab = c^2. \tag{4}$$

Notice that equations (2) and (3) are the special cases of (4) corresponding to $\gamma = 120°$ and $\gamma = 60°$ respectively, while $g = 0$ gives the familiar Pythagorean equation.

Solutions All the solutions are a special case of a general theorem in Dickson [**5**, p. 44], which he gives as an exercise [**5**, p. 48, exercise 6].

If g is rational, $-2 < g < 2$, then all the rational solutions of

$$a^2 + b^2 + gab = c^2$$

are

$$(a,b,c) = k(u^2 - v^2, 2uv + gv^2, u^2 + guv + v^2)$$

where u and v are relatively prime integers, and k is rational.

Let $d = 2 + g$. The assumption $-2 < g < 2$ insures that $d > 0$. We reparameterize by replacing u by $i + j$ and v by j to get

$$(a,b,c) = k(i^2 + 2ij, 2ij + dj^2, i^2 + dij + di^2).$$

This form will be used to give all our solutions.

All cubic polynomials with rational roots and critical points can be obtained by a rational translation along the x-axis of a polynomial of the form (1) with

$$(a,b) = k(i^2 + 2ij, 2ij + 3j^2). \tag{5}$$

We refer to such cubics with $k = 1$ as canonical nice cubics.

For the box problem all rational dimensions are given by

$$(a,b) = k(i^2 + 2ij, 2ij + j^2).$$

The maximum occurs at $x = kij/2$ and the dimensions of the box are

$$(ki^2 + kij) \times (kij + kj^2) \times (kij/2).$$

All triangles with three rational sides and a given angle γ lying opposite side c are of the form

$$(a,b,c) = k(i^2 + 2ij, 2ij + dj^2, i^2 + dij + dj^2), \tag{6}$$

where $d = 2(1 - \cos\gamma)$. In particular

$$\begin{aligned}
\gamma = 60^\circ: &\quad (a,b,c) = k(i^2 + 2ij, 2ij + j^2, i^2 + ij + j^2) \\
\gamma = 90^\circ: &\quad (a,b,c) = k(i^2 + 2ij, 2ij + 2j^2, i^2 + 2ij + 2j^2) \\
\gamma = 120^\circ: &\quad (a,b,c) = k(i^2 + 2ij, 2ij + 3j^2, i^2 + 3ij + 3j^2).
\end{aligned}$$

Examples We first discuss nice cubics. The two simplest canonical cubics are $(x+3)x(x-5)$, corresponding to $i = 1$ and $j = 1$ in equation (5), and $(x+8)x(x-7)$, corresponding to $i = 2$ and $j = 1$. By negating and translating the roots, we can obtain from each canonical cubic six cubes that have zero as a root. From $(x+3)x(x-5)$, we obtain

$$\begin{aligned}
(x+3)x(x-5) &= x^3 - 2x^2 - 15x \\
(x+5)x(x-3) &= x^3 + 2x^2 - 15x \\
x(x-3)(x-8) &= x^3 - 11x^2 + 24x \\
x(x-5)(x-8) &= x^3 - 13x^2 + 40x \\
(x+8)(x+3)x &= x^3 + 11x^2 + 24x \\
(x+8)(x+5)x &= x^3 + 13x^2 + 40x.
\end{aligned}$$

From each of these we can produce nice cubics by translating and multiplying the roots by any rational values. If we use only multiplication, zero is preserved as a root. We multiply by the reciprocals of numbers that divide the product of the nonzero roots. From $x(x-5)(x-8)$, for example, we get eight cubics, four of which are

$$\begin{aligned}
x(x-5)(x-8) &= x^3 - 13x^2 + 40x \\
x(2x-5)(x-4) &= 2x^3 - 13x^2 + 20x \\
x(4x-5)(x-2) &= 4x^3 - 13x^2 + 10x \\
x(x-1)(5x-8) &= 5x^3 - 13x^2 + 8x.
\end{aligned}$$

The remaining four come from these by interchanging the coefficients of the cubic and linear terms. These with larger leading coefficients have roots bunched around zero and are less suitable for graphing.

The first procedure in this section always yields six cubics from any canonical cubic. The number produced by this second procedure, however, depends on the factorization of the nonzero roots. Only four are produced from $(x+3)x(x-5)$ while 16 are produced from $x(x-8)(x-15)$, one of the six cubics derived from $(x+8)x(x-7)$.

The cubic in the introductory paragraph comes from $(x+5)x(x-3)$ by adding -1 and then multiplying by $1/2$. Here are four other nice cubics:

$$(x+2)(x-1)(x-6) = x^3 - 5x^2 - 8x + 12$$
$$(x+1)(x-2)(x-7) = x^3 - 8x^2 + 5x + 14$$
$$(x+1)(2x-3)(x-3) = 2x^3 - 7x^2 + 9$$
$$(x+2)(3x-1)(x-3) = 3x^3 - 4x^2 - 17x + 6.$$

For the box problem we insist that a and b be relatively prime integers (but allow fractional box dimensions). This is achieved by letting $k = 1/3$ when 3 divides $i - j$ and $k = 1$ otherwise. It is not difficult to see that if a and b yield a nice box, so do $b - a$ and b. Thus rectangle dimensions for boxes come in pairs. The table below gives the five smallest such pairs. They have been ordered by the value of b.

(i,j)	k	(a,b)	x	(i,j)	k	(a,b)	x
(4,1)	1/3	(3,8)	2/3	(2,1)	1	(5,8)	1
(3,1)	1	(7,15)	3/2	(5,2)	1/3	(8,15)	5/3
(7,1)	1/3	(5,21)	7/6	(3,2)	1	(16,21)	3
(5,1)	1	(11,35)	5/2	(7,4)	1/3	(24,35)	14/3
(10,1)	1/3	(7,40)	5/3	(4,3)	1	(33,40)	6

These dimensions were published by Duemmel [6] who found them by computer search. Graham and Roberts [10] have a list of rectangle dimensions that overlaps with these. They cite calculus textbooks that have presented the box problem and observe that only the first three rectangles in the second column have actually been used.

Dundas [8] discusses a more practical way of building a box from a rectangular piece of cardboard, with a top and reinforced sides. He shows that this problem involves maximizing the same nice cubic polynomials as the standard box problem discussed here. He also has some interesting discussion about other kinds of boxes.

For triangles here are some triples for $\gamma = 60°$, and $\gamma = 120°$ respectively:

(i,j)	(2,1)	(3,1)	(4,1)	(5,1)	(3,2)	(5,2)
(a,b,c)	(8,5,7)	(15,7,13)	(8,3,7)	(35,11,31)	(21,16,19)	(15,8,13)

(i,j)	(1,1)	(1,2)	(1,3)	(2,1)	(2,3)	(4,1)
(a,b,c)	(3,5,7)	(5,16,19)	(7,33,37)	(8,7,13)	(16,39,49)	(24,11,31).

Finally we look at angles that have a rational cosine. In equation (6) let m/n be d in lowest terms, let $(i,j) = (1,1)$ and multiply by n. This produces the triangle $(a,b,c) = (3n, m+2n, 2m+n)$. Here are some of these values:

$\cos\gamma$	3/4	1/4	5/6	2/3	−1/4	−3/4
d	1/2	3/2	1/3	2/3	5/2	7/2
(a,b,c)	(6,5,4)	(6,7,8)	(9,7,5)	(9,7,8)	3(2,3,4)	(6,11,16).

Other possibilities Obtain the triangle $(a,b,c) = 3(n, n+1, n+2)$ by setting $d = (n+3)/n$ and $\cos\gamma = (n-3)/2n$. The triangle $(a,b,c) = (3n, 3n+1, 3n+2)$ is obtained by setting $d = (n+1)/n$ and $\cos\gamma = (n-1)/2n$. Letting $(i,j) = (1,n)$, yields $(a,b,c) = (1+2n, 2n+mn, 1+m+mn)$. Dundas [7] has some interesting observations about triangles with a given rational value for $\cos\gamma$.

Comments The result from Dickson given in section 2 has a long history. The case $g=0$ is the classical problem of finding all Pythagorean triples. This solution appears in Diophantus, Book II, Theorem 8. According to Heath however, it had essentially already appeared in Euclid Book X, Theorem 28, Lemma 1 [**11**, p. 116–7], [**12**, Vol. 3, p. 63–4]. For $g=-1$ this solution was first given by J. Neuberg in 1874–5 [**4**, p. 405, ref. 36], and for $g=1$ by J. Neuberg and G. B. Mathews in 1887 [**4**, p. 406, ref. 40]. The general theorem was first proved by A. Gérardin in 1911 [**4**, p. 406, ref. 45].

All the sources we have consulted use this form [**5**, p. 42, exercise 1], [**9**, p. 16], [**14**, p. 146]. We prefer our second form. It has the nice property that positive i and j give positive a, b, and c. When i is much bigger, about the same size, or much smaller than j, it gives an a that is respectively much bigger, about the same size, or much smaller than b. At the start of our investigation we discovered this parameterization even before we had begun to generate triples on the computer.

In our investigation of nice cubics the initial assumption that the middle root be zero is arbitrary. Another approach is to assume a pair of roots of equal magnitude and opposite sign. This is achieved through translation by $(a-b)/2$. The resulting polynomial has the form $(x^2-r^2)(x-s)$, where

$$r=(a+b)/2 \qquad a=r+s$$
$$s=(a-b)/2 \qquad b=r-s$$

and the condition that it be nice reduces to

$$3r^2+s^2=c^2.$$

This transformation eliminates the cross term from the left side of equation (2). This could be achieved by a transformation corresponding to a 45° rotation, which has determinant 1 but irrational entries. The transformation above has rational entries, therefore preserving rational points, but has the disadvantage of a determinant $\neq 1$. Graham and Roberts [**10**] use (essentially) this transformation to solve the box problem. They obtain a substantially different parameterization than the two discussed here.

A linear transformation with determinant 1 that fixes the set of integer solutions to a Diophantine equation is known as an *automorph* [**5**, p. 72–3], [**3**, p. 146]. The proof of the formula for Pythagorean triples given in [**14**, p. 146–7] makes use of the factorization of the left-hand side of equation (4) over the quadratic extension of the rational field obtained by adjoining the square root of the discriminant g^2-4. When the discriminant is negative, it is known that multiplication by the roots of unity in this extension gives all automorphs.

With just two exceptions, a negative discriminant yields exactly two roots of unity, namely ± 1. The exceptions correspond precisely to the three equations studied here. With $g=0$ the extension contains the fourth roots of unity. With $g=\pm 1$ the extension contains the sixth roots of unity. This is what lies behind the procedure at the beginning of section 3 in which six nice cubics are generated from each canonical cubic.

A symmetry like interchanging a and b can be very useful even though it has determinant -1 and does not qualify as an automorph. We observed that if a and b yield a nice box, so do $b-a$ and b. This transformation also has determinant -1. It is used in both [**6**] and [**10**].

The relationship of automorphs to the units in a quadratic extension of the rational field is explained in Borevich and Shafarevich [**1**, p. 75–6]. The fact that there are only finitely many units (and therefore automorphs) when the discriminant is negative can be regarded as a very special case of the Dirichlet Unit Theorem [**1**, p. 114].

REFERENCES

1. Z. I. Borevich and I. R. Shafarevich, *Number Theory*, Academic Press, New York, 1966.
2. T. Bruggeman and T. Gush, Nice cubic polynomials for curve sketching, this MAGAZINE 53 (1980), 233–234.
3. H. Davenport, *The Higher Arithmetic*, 5th edition, Cambridge University Press, London, 1982.
4. Leonard E. Dickson, *History of the Theory of Numbers*, Vol. 2, Chelsea Publishing Co., 1952.
5. Leonard E. Dickson, *Introduction to the Theory of Numbers*, Dover Publications, Inc., New York, 1957.
6. James Duemmel, From calculus to number theory, *Amer. Math. Monthly* 96 (1989), 140–143.
7. Kay Dundas, Quasi-Pythagorean triples for an oblique triangle, *The Two-year College Mathematics Journal* 8 (1977), 152–155.
8. Day Dundas, To build a better box, *The College Mathematics Journal* 15 (1984), 30–36.
9. A. O. Gelfond, *The Solution of Equations in Integers*, W. H. Freeman, San Francisco, 1961.
10. George P. Graham and Charles E. Roberts, A Diophantine equation from calculus, this MAGAZINE 62 (1989), 97–101.
11. Thomas L. Heath, *Diophantus of Alexandria*, Dover Publications, Inc., New York, 1964.
12. Thomas L. Heath, *The Thirteen Books of Euclid's Elements*, Cambridge Univ. Press, London, 1926.
13. O. P. Lossers, Integer roots of cubics, *Amer. Math. Monthly* 96 (1989), 841–842.
14. Harry Pollard and Harold G. Diamond, *The Theory of Algebraic Numbers*, 2nd edition, MAA, 1975.

A Calculus Exercise For the Sums of Integer Powers

JOSEPH WIENER
University of Texas-Pan American
Edinburg, TX 78539

Let $S_{k,n} = 1^k + 2^k + \cdots + n^k$, where k, n are positive integers. The usual method of finding $S_{k,n}$ for $k = 1, 2, \ldots$ is by means of the identity

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

where $\binom{k}{i}$ denotes the binomial coefficient $k!/i!(k-i)!$. Recently, D. Acu has obtained (1) and similar formulas involving only even or only odd values of i from certain binomial identities [1].

We shall derive a generalization of (1) by differentiating the function

$$F(x) = e^{ax} + e^{(a+d)x} + e^{(a+2d)x} + \cdots + e^{(a+(n-1)d)x}.$$

Indeed,

$$F^{(k)}(0) = \sum_{m=1}^{n} (a + (m-1)d)^k, \tag{2}$$

REFERENCES

1. Z. I. Borevich and I. R. Shafarevich, *Number Theory*, Academic Press, New York, 1966.
2. T. Bruggeman and T. Gush, Nice cubic polynomials for curve sketching, this MAGAZINE 53 (1980), 233–234.
3. H. Davenport, *The Higher Arithmetic*, 5th edition, Cambridge University Press, London, 1982.
4. Leonard E. Dickson, *History of the Theory of Numbers*, Vol. 2, Chelsea Publishing Co., 1952.
5. Leonard E. Dickson, *Introduction to the Theory of Numbers*, Dover Publications, Inc., New York, 1957.
6. James Duemmel, From calculus to number theory, *Amer. Math. Monthly* 96 (1989), 140–143.
7. Kay Dundas, Quasi-Pythagorean triples for an oblique triangle, *The Two-year College Mathematics Journal* 8 (1977), 152–155.
8. Day Dundas, To build a better box, *The College Mathematics Journal* 15 (1984), 30–36.
9. A. O. Gelfond, *The Solution of Equations in Integers*, W. H. Freeman, San Francisco, 1961.
10. *George P. Graham and Charles E. Roberts, A Diophantine equation from calculus, this MAGAZINE 62* (1989), 97–101.
11. Thomas L. Heath, *Diophantus of Alexandria*, Dover Publications, Inc., New York, 1964.
12. Thomas L. Heath, *The Thirteen Books of Euclid's Elements*, Cambridge Univ. Press, London, 1926.
13. O. P. Lossers, Integer roots of cubics, *Amer. Math. Monthly* 96 (1989), 841–842.
14. Harry Pollard and Harold G. Diamond, *The Theory of Algebraic Numbers*, 2nd edition, MAA, 1975.

A Calculus Exercise For the Sums of Integer Powers

JOSEPH WIENER
University of Texas-Pan American
Edinburg, TX 78539

Let $S_{k,n} = 1^k + 2^k + \cdots + n^k$, where k, n are positive integers. The usual method of finding $S_{k,n}$ for $k = 1, 2, \ldots$ is by means of the identity

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

where $\binom{k}{i}$ denotes the binomial coefficient $k!/i!(k-i)!$. Recently, D. Acu has obtained (1) and similar formulas involving only even or only odd values of i from certain binomial identities [1].

We shall derive a generalization of (1) by differentiating the function

$$F(x) = e^{ax} + e^{(a+d)x} + e^{(a+2d)x} + \cdots + e^{(a+(n-1)d)x}.$$

Indeed,

$$F^{(k)}(0) = \sum_{m=1}^{n} (a + (m-1)d)^k, \tag{2}$$

where $F^{(k)}(0)$ is the kth derivative of $F(x)$ at $x=0$. Furthermore, the formula for the sum of terms of this finite geometric sequence yields

$$F(x)=\frac{\left(e^{(a+nd)x}-e^{ax}\right)}{\left(e^{dx}-1\right)},$$

whence

$$\left(e^{dx}-1\right)F(x)=e^{(a+nd)x}-e^{ax}.$$

Now, differentiate this identity k times and set $x=0$ to obtain

$$\sum_{i=0}^{k-1}\binom{k}{i}d^{k-i}F^{(i)}(0)=(a+nd)^k-a^k. \tag{3}$$

Denote

$$S_{k,n}(a,d)=\sum_{m=1}^{n}(a+(m-1)d)^k; \tag{4}$$

then, by virtue of (2), (3), (4), we get the formula

$$\sum_{i=0}^{k-1}\binom{k}{i}d^{k-i}S_{i,n}(a,d)=(a+nd)^k-a^k, \tag{5}$$

which generalizes (1) since $S_{k,n}=S_{k,n}(1,1)$. This method of deriving identity (5) has been first discovered in [2] and used later [3] to evaluate some other sums involving integer powers.

Furthermore, to obtain formulas of type (5), where the index i varies only over even integers, we introduce the even function $G(x)=F(x)+F(-x)$ and differentiate the identity

$$\left(e^{dx}-1\right)G(x)=e^{(a+nd)x}-e^{-(a+(n-1)d)x}$$
$$-e^{ax}+e^{(d-a)x},$$

k times at $x=0$ to obtain

$$\sum_{i=0}^{k-1}\binom{k}{i}d^{k-i}G^{(i)}(0)=(a+nd)^k$$
$$+(-1)^{k-1}(a+(n-1)d)^k-a^k+(d-a)^k.$$

Since $G^{(i)}(0)=F^{(i)}(0)+(-1)^iF^{(i)}(0)$ and $F^{(i)}(0)=S_{i,n}(a,d)$, then

$$\sum_{i=0}^{k-1}\binom{k}{i}d^{k-i}\left(1+(-1)^i\right)S_{i,n}(a,d)$$
$$=(a+nd)^k+(-1)^{k-1}(a+(n-1)d)^k-a^k+(d-a)^k.$$

The substitution $i=2p$ changes this formula to

$$2\sum_{p=0}^{[(k-1)/2]}\binom{k}{2p}d^{k-2p}S_{2p,n}(a,d)$$
$$=(a+nd)^k+(-1)^{k-1}(a+(n-1)d)^k-a^k+(d-a)^k, \tag{6}$$

where $[\cdot]$ denotes the greatest integer function. For $a = d = 1$, formula (6) yields

$$2\sum_{p=0}^{[(k-1)/2]}\binom{k}{2p}S_{2p,n} = (n+1)^k + (-1)^{k-1}n^k - 1.$$

To obtain formulas of type (5), where the index i varies only over odd integers, we introduce the odd function $H(x) = F(x) - F(-x)$ and differentiate the identity

$$(e^{dx}-1)H(x) = e^{(a+nd)x} + e^{-(a+(n-1)d)x} - e^{ax} - e^{(d-a)x}$$

k times at $x = 0$. Since $H^{(i)}(0) = F^{(i)}(0) - (-1)^i F^{(i)}(0)$, then

$$\sum_{i=0}^{k-1}\binom{k}{i}d^{k-i}(1-(-1)^i)S_{i,n}(a,d) = (a+nd)^k + (-1)^k(a+(n-1)d)^k - a^k - (d-a)^k.$$

The substitution $i = 2p - 1$ changes this formula to

$$2\sum_{p=1}^{[k/2]}\binom{k}{2p-1}d^{k-2p+1}S_{2p-1,n}(a,d) = (a+nd)^k + (-1)^k(a+(n-1)d)^k - a^k - (d-a)^k. \tag{7}$$

For $a = d = 1$, formula (7) yields

$$2\sum_{p=1}^{[k/2]}\binom{k}{2p-1}S_{2p-1,n} = (n+1)^k + (-1)^k n^k - 1.$$

Acknowledgements. Research partially supported by U.S. Army Grant DAAL03-89-G-0107. I am thankful to the referees for a number of useful suggestions.

REFERENCES

1. D. Acu, Some algorithms for sums of integer powers, this MAGAZINE 61 (1988), 189–191.
2. J. Wiener and A. Wojczechowsky, Evaluating certain finite sums by differentiation, *U.S.S.R. Journal Mathematics in School* 2 (1968), 84–87.
3. J. Wiener, Solution of Problem E. 2770, *Amer. Math. Monthly* 87 (1980), 491.

An Inductive Proof for Extremal Simplexes

M. SAYRAFIEZADEH
Medgar Evers College
of the City University of New York
Brooklyn, NY 11225

Among the triangles that circumscribe a given circle, the one with minimum area adopts a particularly balanced position: It touches the circle at the midpoints of its sides. This is not surprising, and one may think that it is true due to the symmetry in the circle. It is not! For example, take the circle with its minimal circumscribing triangle and look at their shadow on the wall. Circle turns into an ellipse, the sides of the triangle change size, but they still touch the ellipse at their midpoints. In fact the above proposition remains valid if the circle is replaced by any convex figure.

The key element in the proposition is the minimization of the area function, and as such it is reminiscent of extrema problems of calculus. However, if the student routinely rushes to do the problem using the techniques of calculus he or she will soon recognize that the problem is far from being routine. If the convex figure is not smooth, then the usual techniques of calculus that require differentiability of functions will not apply.

Using continuity arguments and affine transformations Chakerian and Lange gave an elementary proof of the proposition in this MAGAZINE [2]. Day has proved a generalization of the problem to higher dimensions. He shows that if a polyhedron circumscribes a convex body but does not touch it at the centroid of one of its faces, then there is a polyhedron of smaller volume that also circumscribes the body [3]. The proof involves analysis in space and is not elementary. The reader will find some other, closely related results on this topic in the articles by Bailey [1] and Lange [7, 8].

In this note we will use elementary geometric transformations and "cutting and pasting" to prove the proposition, and then we will generalize it to n-dimensions by induction. Propositions will be formulated in terms of a *convex body*, which is a closed and bounded convex set with an interior point in n-dimensional Euclidean space, $n \geqslant 2$.

PROPOSITION 1. *A convex body K in the plane that is contained in a triangle of minimum area, contains the midpoints of the sides of the triangle.*

Proof. The proposition will be proved by showing that if a triangle PQR contains K, but the midpoint of one of its sides is not contained in K, then it can not be minimal. If the midpoint G of side, say QR, is outside K, there will be a line 1 that separates the convex body from G (FIGURE 1a). Now we proceed to construct a smaller triangle than PQR, which contains K. QR must touch K at some point, otherwise the triangle is not minimal. This implies that 1 intersects QR at a point O, and another side of the triangle, say PQ, at a point N. The reflection through point O transforms ONQ into the congruent triangle $ON'Q'$, which has the point R in the interior of the side OQ'. Extension of PR will meet another side of the triangle $ON'Q'$ at some point M'.

Case (a): M' is on ON', as in FIGURE 1a. In this case: area$(OM'R) <$ area$(ON'Q') =$ area(ONQ). Hence, cutting off the triangle ONQ and adding $OM'R$ yields a new triangle PNM' of smaller area, which also contains K.

Case (b): M' is on $Q'N'$, as in FIGURE 1b. Let M be the reflection of the point M' through O. The following relations: area$(O'MR) <$ area$(OM'Q') =$ area(OMQ), again

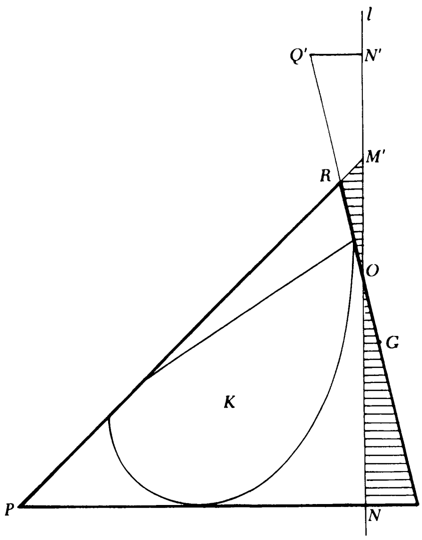


FIGURE 1a

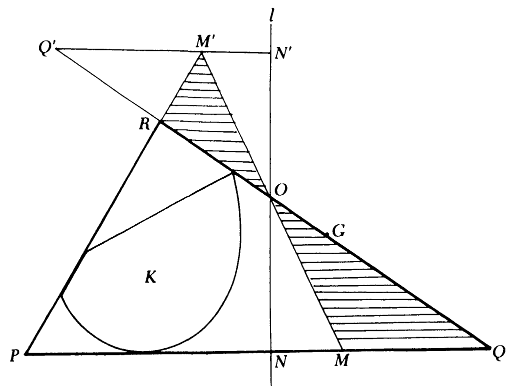


FIGURE 1b

show that cutting off OMQ and adding $OM'R$ yields a new triangle that contains K and has a smaller area than PQR. This exhausts the cases and proves the proposition. It may be noted that the proof remains valid if the statement of the theorem is extended to replace the triangle by a polygon.

Behind the simplicity of the proof lurks an existence question. In the set of triangles that contain a convex body is there always one with minimum area? The answer is affirmative and follows the usual compactness arguments: A certain distance is defined between two triangles that turns the set of triangles into a metric space. In this space the set of triangles that contain K and have an upper bound on their areas is compact, and consequently the area function being continuous on this set assumes its minimum there. These arguments may be generalized to higher dimensions to establish the existence of simplexes of minimum volume that contain a convex body, which will be dealt with in the next theorem. See [5] pp. 40–47.

Before proceeding with a generalization of Proposition 1 we will define the notation: An n-dimensional *simplex* in the Euclidean n-space E_n is the smallest convex body $\triangle$, if it exists, that contains the $n+1$ given points $\{V_i\}$, $i = 0, 1, \ldots, n$. V_i is a *vertex* of $\triangle$, and the $n-1$ dimensional simplex $\triangle_i$ formed by the rest of the vertices is the *face* of $\triangle$ opposite to V_i. The *centroid* G of $\triangle$ is the unique point that is inductively determined by $G = \bigcap_{i=0}^{n} V_i G_i$, $n > 1$, where $V_i G_i$ is the line joining the vertex V_i to the centroid of its opposite face G_i. The *height* h_i is the smallest distance between V_i and its opposite face $\triangle_i$. The *n-dimensional volume* of $\triangle$ in E_n, may be defined inductively as $(1/n)\, h$ multiplied by the $n-1$ dimensional volume of $\triangle_i$. A convex body is *strictly convex* if its boundary contains no line segments. By compactness arguments it may be shown that each convex body in E is contained in a simplex of minimum volume.

PROPOSITION 2. *A strictly convex body in E_n and a simplex of minimum volume containing it meet at the centroids of the faces of the simplex.*

Proof. We use induction on n, the dimension of the space E_n. Proposition 1 established the validity of this proposition for $n = 2$. Assume that it is valid in E_{n-1}, $n > 2$. Let K be a strictly convex body in E_n, and let $\triangle$ be a simplex of minimum volume that contains it. Consider any vertex V_k and its opposite face $\triangle_k$. We will show that the centroid G_k of Δ_k belongs to K. Minimality of Δ and strict convexity of K imply that there is a unique point M_k on the face Δ_k that belongs to K. To complete the proof it is sufficient to show that $M_k = G_k$.

Through V_i, $i \neq k$, make a projection of K into Δ_i to obtain K_i, in other words $K_i = \{P | P = \overline{V_i Q} \cap \Delta_i, Q \in K, P \in E_n\}$, (FIGURE 2). K_i is a strictly convex body in E_{n-1} [4]; and Δ_i is an $n-1$ dimensional simplex of minimum volume that contains it. It is minimum because any other $n-1$ dimensional simplex Δ'_i that contains K_i gives rise to an n-dimensional simplex Δ', with a face Δ'_i and opposite vertex V_i that also contains K_i. Then Δ' and Δ have the same height, and minimality of Δ implies minimality of Δ_i. By the induction hypothesis, the centroids of the faces of Δ_i meet K_i. In particular, G_{ik}, the centroid of the face $\Delta_i \cap \Delta_k$ of Δ_i, $i \neq k$, belongs to K_i. This implies that M_k is on the line $V_i G_{ik}$. Now repeat the same process with V_i, $i \neq j \neq k$, and project K into Δ to find that the line $V_i G_{jk}$, where G_{jk} is the centroid of $\Delta_j \cap \Delta_k$, contains M_k. The point M_k thus belongs to two lines which intersect at G_k. This yields $M_k = G_k$, and proves the theorem. Strict convexity of K, while indispensable in this proof, is not necessary for the validity of the proposition [3].

Not every simplex that contains K and meets it at the centroids of its faces is necessarily minimal. As a strictly convex example in the plane consider the Reuleaux triangle K of width h. (K is the intersection of three disks of radius h, each having its center on a vertex of an equilateral triangle PQR of side length h). A homothecy through the center of PQR and ratio -2 transforms PQR into another triangle Δ. The new triangle contains K and meets it at the midpoints of its sides, namely P, Q, and R, but it is not minimal. To see this we use a result by Gross [2, p. 67], which states that for Δ to be minimal the inequality area $(\Delta) \leqslant 2$ area(K) must hold. In our case area $(\Delta) = \sqrt{3}h^2$; area $(K) = (1/2)(\pi - 3)h^2$; which yields the wrong inequality, area $(\Delta) > 2$ area(K).

However, if K is an ellipse meeting a circumscribed triangle Δ at the midpoints of its sides, then Δ is minimal. This is due to the affine invariance of the proposition that allows K to be taken as a circle and Δ to be an equilateral triangle and thus minimal. This argument may be extended to higher dimensions. For example, let $\Delta = V_0V_1V_2V_3$ be a tetrahedron (simplex in E_3) that is circumscribed to a sphere S and meets it at the centroids G_0, G_1, G_2, G_3. The three centroids G_1, G_2, G_3 determine a plane π that is parallel to Δ_0, the face of Δ opposite the vertex V_0. The plane

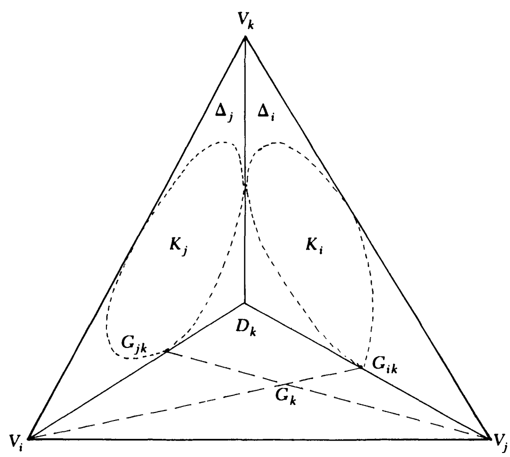


FIGURE 2

π intersects $\triangle$ and S in a triangle T, and a circle C, respectively. Triangle T contains C and meets it at the midpoints of its sides. This is the plane case of the problem implying that T is equilateral. $\triangle_0$ is obtained from T by a homothecy with center V_0; hence $\triangle_0$ is equilateral. Similarly, other faces of $\triangle$ are equilateral, and $\triangle$ is regular (all sides are congruent). Projections of the type used in the proof of Proposition 2 show that simplexes of minimum volume circumscribed to a sphere must be regular; and conversely, any regular simplex that circumscribes a sphere must be minimal. Then induction yields the following proposition:

PROPOSITION 3. *A simplex that circumscribes an ellipsoid K is minimal in volume if and only if it meets K at the centroids of its faces.*

Acknowledgements. I am grateful to Professor Heinrich Guggenheimer whose final examination in an elementary course on convexity motivated me to work on this problem [6]. This was many years ago, but his creative approach to teaching has remained to be a source of inspiration for me. The author would like to thank the referees for their helpful suggestions and their supplying several references.

REFERENCES

1. H. Bailey, A surprising max-min result, *College Math. J.* 18 (1987), 225–229.
2. G. D. Chakerian and L. H. Lange, Geometric extremum problems, this MAGAZINE 44 (1971), 57–69.
3. M. M. Day, Polygons circumscribed about closed convex curves, *Trans. Amer. Math. Soc.* 62 (1957), 315–319.
4. H. G. Eggleston, *Convexity*, Cambridge Univ. Press, Cambridge, 1958.
5. H. Guggenheimer, *Applicable Geometry*, Krieger, Malabar, FL, 1977.
6. H. Guggenheimer, Some problems on convex figures, *Amer. Math. Monthly* 70 (1963), 656–659.
7. L. H. Lange, Several hyperbolic encounters, *Two-Year College Math. J.* 7 (1976), 2–6.
8. L. H. Lange, Cutting certain minimum corners, *Amer. Math. Monthly* 83 (1976), 361–365.

A Characterization of Continuity

JINGCHENG TONG
University of North Florida
Jacksonville, FL 32216

The following two examples of discontinuous functions from $(-\infty, +\infty)$ into $(-\infty, +\infty)$ are often quoted in books on calculus.

$$f(x)=\begin{cases}0 & x \text{ is rational,}\\ 1 & x \text{ is irrational;}\end{cases}$$

$$g(x)=\begin{cases}x & x\leqslant 0,\\ 1+x & x>0.\end{cases}$$

In this note we show that f and g each enjoy one of two weaker properties, which we call "almost continuity" and "near continuity." These two properties between them imply continuity.

Let us make an analysis.

π intersects $\triangle$ and S in a triangle T, and a circle C, respectively. Triangle T contains C and meets it at the midpoints of its sides. This is the plane case of the problem implying that T is equilateral. $\triangle_0$ is obtained from T by a homothecy with center V_0; hence $\triangle_0$ is equilateral. Similarly, other faces of $\triangle$ are equilateral, and $\triangle$ is regular (all sides are congruent). Projections of the type used in the proof of Proposition 2 show that simplexes of minimum volume circumscribed to a sphere must be regular; and conversely, any regular simplex that circumscribes a sphere must be minimal. Then induction yields the following proposition:

PROPOSITION 3. *A simplex that circumscribes an ellipsoid K is minimal in volume if and only if it meets K at the centroids of its faces.*

Acknowledgements. I am grateful to Professor Heinrich Guggenheimer whose final examination in an elementary course on convexity motivated me to work on this problem [6]. This was many years ago, but his creative approach to teaching has remained to be a source of inspiration for me. The author would like to thank the referees for their helpful suggestions and their supplying several references.

REFERENCES

1. H. Bailey, A surprising max-min result, *College Math. J.* 18 (1987), 225–229.
2. G. D. Chakerian and L. H. Lange, Geometric extremum problems, this MAGAZINE 44 (1971), 57–69.
3. M. M. Day, Polygons circumscribed about closed convex curves, *Trans. Amer. Math. Soc.* 62 (1957), 315–319.
4. H. G. Eggleston, *Convexity*, Cambridge Univ. Press, Cambridge, 1958.
5. H. Guggenheimer, *Applicable Geometry*, Krieger, Malabar, FL, 1977.
6. H. Guggenheimer, Some problems on convex figures, *Amer. Math. Monthly* 70 (1963), 656–659.
7. L. H. Lange, Several hyperbolic encounters, *Two-Year College Math. J.* 7 (1976), 2–6.
8. L. H. Lange, Cutting certain minimum corners, *Amer. Math. Monthly* 83 (1976), 361–365.

A Characterization of Continuity

JINGCHENG TONG
University of North Florida
Jacksonville, FL 32216

The following two examples of discontinuous functions from $(-\infty, +\infty)$ into $(-\infty, +\infty)$ are often quoted in books on calculus.

$$f(x) = \begin{cases} 0 & x \text{ is rational,} \\ 1 & x \text{ is irrational;} \end{cases}$$

$$g(x) = \begin{cases} x & x \leqslant 0, \\ 1+x & x > 0. \end{cases}$$

In this note we show that f and g each enjoy one of two weaker properties, which we call "almost continuity" and "near continuity." These two properties between them imply continuity.

Let us make an analysis.

The function f is nowhere continuous since for each open set V containing 0 or 1 but not both, $f^{-1}(V)$ is not open. However, for any such open set V, $f^{-1}(V) \subset \operatorname{Int}\operatorname{Cl} f^{-1}(V) = (-\infty, +\infty)$. Thus f satisfies the following definition due to T. Husain ([1] or [2]).

Definition 1. Let X and Y be topological spaces. A function $f\colon X \to Y$ is said to be *almost continuous* if for each open set V in $Y, f^{-1}(V) \subset \operatorname{Int}\operatorname{Cl} f^{-1}(V)$.

The function g is not continuous at $x = 0$ since for each open set $V = (a, b)$ containing 0 with $b < 1$, $g^{-1}(V) = (a, 0]$, which is not open in $(-\infty, +\infty)$. However, $g^{-1}(V) = (a, 0]$ is open in the subspace $\operatorname{Cl} g^{-1}(V) = [a, 0]$, i.e., there is an open set U such that $g^{-1}(V) = (\operatorname{Cl} g^{-1}(V)) \cap U$. For instance, we can let $U = (a, 1)$. Thus we have the following definition.

Definition 2. Let X and Y be topological spaces. A function $f\colon X \to Y$ is said to be *nearly continuous* if for each open set V in $Y, f^{-1}(V)$ is an open set in the subspace $\operatorname{Cl} f^{-1}(V)$.

Obviously a continuous function is both almost continuous and nearly continuous. The converse is also true.

THEOREM. *Let X and Y be topological spaces. A function $f\colon X \to Y$ is continuous if and only if it is almost continuous and nearly continuous.*

Proof. We need only prove the sufficiency.

Let V be an arbitrary open set in Y. Since f is nearly continuous, there is an open set $U \subset X$ such that $f^{-1}(V) = (\operatorname{Cl} f^{-1}(V)) \cap U$. Hence $f^{-1}(V) \subset U$. The function f is almost continuous, therefore $f^{-1}(V) \subset \operatorname{Int}\operatorname{Cl} f^{-1}(V)$. We have

$$\begin{aligned}
\operatorname{Int} f^{-1}(V) &= \operatorname{Int}\big((\operatorname{Cl} f^{-1}(V)) \cap U\big) \\
&= \operatorname{Int}(\operatorname{Cl} f^{-1}(V)) \cap \operatorname{Int}(U) \\
&= \operatorname{Int}\operatorname{Cl} f^{-1}(V) \cap U \\
&\supseteq f^{-1}(V) \cap f^{-1}(V) \\
&= f^{-1}(V).
\end{aligned}$$

Thus $f^{-1}(V)$ is open in X, and hence f is continuous.

Acknowledgement. The author thanks the referees for their valuable suggestions to improve this note.

REFERENCES

1. T. Husain, Almost continuous mappings, *Prace Math.*, 10 (1966), 1–7.
2. T. Husain, *Topology and Maps*, Plenum Press, New York, 1977.

A Note on Topological Continuity

SCOTT J. BESLIN
Nicholls State University
Thibodaux, LA 70310

Let X and Y be topological spaces, and f a function from X into Y. In an elementary topology course, it is usually shown that the continuity of f is equivalent to the condition

(I) For every $A \subset X$, $f(\bar{A}) \subset \overline{f(A)}$.

Here $\bar{A}$ denotes the closure of A. In fact, in the above statement, one may replace $\bar{A}$ by A', the set of limit points, or derived set, of A. Some texts (Dugundji [3], for instance) show that continuity is equivalent to

(II) For every $B \subset Y$, $\overline{f^{-1}(B)} \subset f^{-1}(\bar{B})$.

Both of these conditions involve the closure operator. Recently, a student in an undergraduate topology class asked whether or not there exists a set condition equivalent to continuity which involves the interior operator, denoted here by "int". In Bourbaki [1], one may find as an exercise that the function f from X into Y is continuous if and only if the following is true:

(III) For every $B \subset Y$, $f^{-1}(\text{int}(B)) \subset \text{int}(f^{-1}(B))$.

This criterion is also stated in Brown [2] as a proposition. Most of the other commonly used texts do not consider condition (III) at all.

The pair of conditions (I) and (II) is analogous to the pair (III) and

(IV) For every $A \subset X$, $\text{int}(f(A)) \subset f(\text{int}(A))$.

It is thus somewhat surprising that, as we shall see, (IV) neither implies nor is implied by continuity. However, (IV) is nevertheless closely related to continuity.

In what follows, the symbol R stands for the set of real numbers, while (R, usual) denotes the usual topology on R. The symbol R^2 represents the product space (R, usual) $\times$ (R, usual). The discrete topology on a set X consists of all the subsets of X, while the indiscrete topology consists only of the empty set $\varnothing$ and X itself.

For an example of a continuous function that does not satisfy (IV), let f: (R, usual) $\to$ (R, discrete) be defined by $f(x) = 0$ for all x. Then f is continuous, but if A is the set of rationals, $\text{int}(f(A)) = \{0\}$ while $f(\text{int}(A)) = \varnothing$.

A constant function is the "extreme" example of a function that is not injective. If f is an injective function, then for each $A \subset X$, $f^{-1}(f(A)) = A$. This fact allows us to prove the following theorem.

THEOREM 1. *Let f be an injective continuous function from X into Y. Then f satisfies (IV).*

Proof. Let x be in A, a subset of X, so that $f(x)$ is in $\text{int}(f(A))$. Then there exists an open set U of Y such that $f(x)$ is in U, and $U \subset f(A)$. Since f is injective, $f^{-1}(f(A)) = A$, and hence $x \in f^{-1}(U) \subset A$. The set $f^{-1}(U)$ is open in X by the continuity of f. So $x \in \text{int}(A)$ and $f(x) \in f(\text{int}(A))$.

On the other hand, condition (IV) does not imply the continuity of f, even if f is injective. Indeed if f: $X \to Y$ is such that $\text{int}(f(X)) = \varnothing$, then (IV) is satisfied. Thus if

f: (R, usual) $\to R^2$ is given by $f(x)=(x,0)$ if $x \leqslant 0$ and $f(x)=(x,1)$ if $x>0$, then (IV) is satisfied, but f is not continuous.

Moreover, a constant function from R with its discrete topology into R with its indiscrete topology shows that a continuous function satisfying (IV) need not be injective.

With a proof as straightforward as that of Theorem 1, we may prove

THEOREM 2. *Suppose f: $X \to Y$ is surjective and satisfies (IV). Then f is continuous.*

The student may find it helpful and instructive to prove and/or disprove other conditions for continuity formulated by taking combinations of the closure and interior operators.

The author is grateful to the referees and the editor for their thorough and valuable comments, especially those leading to Theorem 2.

REFERENCES

1. N. Bourbaki, *Elements of Mathematics: General Topology*, part I, Addison-Wesley Publishing Co., Inc. Reading, MA, 1966.
2. R. Brown, *Elements of Modern Topology*, McGraw-Hill, London, 1968.
3. J. Dugundji, *Topology*, Allyn and Bacon, Inc., Boston, MA, 1966.

An Inequality of Orthogonal Complements

IRVING J. KATZ
George Washington University
Washington, DC 20052

Let V be an infinite-dimensional inner product space with (x,y) denoting the inner product and $S^{\perp}=\{v \in V|(v,s)=0$ for all $s \in S\}$, where S is a subspace of V. The standard example illustrating that $(S^{\perp})^{\perp}=S$ is not necessarily true makes use of the Weierstrass approximation theorem [**3**, pp. 85, 124]. We present an example of the inequality that avoids the use of the Weierstrass theorem by restricting attention to spaces of polynomials. In particular, let V be the infinite dimensional vector space of polynomials with real coefficients. For $f(x)$, $g(x) \in V$ define $(f,g)=\int_{-1}^{1} f(x)g(x)\,dx$. Let U be the subspace $\{\Sigma a_i x^{4i}\}$ (that is, polynomials in which all exponents are multiples of 4). We shall prove that $(U^{\perp})^{\perp} \neq U$.

This note was motivated by an earlier note of Fowler [**1**] that provided a different example of an infinite dimensional vector space V with a subspace S satisfying $(S^{\perp})^{\perp} \neq S$. The present example seems to be more accessible to students in an introductory linear algebra course, so it may make possible consideration of a point generally not discussed at that stage.

Denote by W_0 the subspace of V consisting of the even polynomials, by W_1, the subspace of V consisting of the odd polynomials. We assert that (1) $V=W_0+W_1$,

f: (R, usual) $\to R^2$ is given by $f(x) = (x, 0)$ if $x \leqslant 0$ and $f(x) = (x, 1)$ if $x > 0$, then (IV) is satisfied, but f is not continuous.

Moreover, a constant function from R with its discrete topology into R with its indiscrete topology shows that a continuous function satisfying (IV) need not be injective.

With a proof as straightforward as that of Theorem 1, we may prove

THEOREM 2. *Suppose* $f: X \to Y$ *is surjective and satisfies (IV). Then* f *is continuous.*

The student may find it helpful and instructive to prove and/or disprove other conditions for continuity formulated by taking combinations of the closure and interior operators.

The author is grateful to the referees and the editor for their thorough and valuable comments, especially those leading to Theorem 2.

REFERENCES

1. N. Bourbaki, *Elements of Mathematics: General Topology*, part I, Addison-Wesley Publishing Co., Inc. Reading, MA, 1966.
2. R. Brown, *Elements of Modern Topology*, McGraw-Hill, London, 1968.
3. J. Dugundji, *Topology*, Allyn and Bacon, Inc., Boston, MA, 1966.

An Inequality of Orthogonal Complements

IRVING J. KATZ
George Washington University
Washington, DC 20052

Let V be an infinite-dimensional inner product space with (x, y) denoting the inner product and $S^{\perp} = \{v \in V | (v, s) = 0 \text{ for all } s \in S\}$, where S is a subspace of V. The standard example illustrating that $(S^{\perp})^{\perp} = S$ is not necessarily true makes use of the Weierstrass approximation theorem [3, pp. 85, 124]. We present an example of the inequality that avoids the use of the Weierstrass theorem by restricting attention to spaces of polynomials. In particular, let V be the infinite dimensional vector space of polynomials with real coefficients. For $f(x), g(x) \in V$ define $(f, g) = \int_{-1}^{1} f(x)g(x)\,dx$. Let U be the subspace $\{\Sigma a_i x^{4i}\}$ (that is, polynomials in which all exponents are multiples of 4). We shall prove that $(U^{\perp})^{\perp} \neq U$.

This note was motivated by an earlier note of Fowler [1] that provided a different example of an infinite dimensional vector space V with a subspace S satisfying $(S^{\perp})^{\perp} \neq S$. The present example seems to be more accessible to students in an introductory linear algebra course, so it may make possible consideration of a point generally not discussed at that stage.

Denote by W_0 the subspace of V consisting of the even polynomials, by W_1, the subspace of V consisting of the odd polynomials. We assert that (1) $V = W_0 + W_1$,

(2) $W_0^\perp = W_1$, (3) $W_1^\perp = W_0$, and (4) $U^\perp = W_1$. The validity of (1) is evident. To prove (2) choose $f \in W_0$, $g \in W_1$. Then $fg \in W_1$ so that an antiderivative of fg is an even polynomial, and hence it follows that $\int_{-1}^{1} f(x)g(x)\,dx = 0$. This shows that $W_1 \subset W_0^\perp$. Next choose $p \in W_0^\perp$ and write $p = p_0 + p_1$, $p_0 \in W_0$, $p_1 \in W_1$. Then $0 = (p_0, p) = (p_0, p_0) + (p_0, p_1) = (p_0, p_0) = \int_{-1}^{1} p_0^2(x)\,dx$. But $p_0^2(x) \geqslant 0$, so the integral can be zero only if $p_0 = 0$. Thus $W_0^\perp \subset W_1$ and (2) is established. The proof of (3) is similar.

Now we prove (4). Since $U \subset W_0$, it follows that $W_1 = W_0^\perp \subset U^\perp$. The proof of the reverse inclusion requires only a bit more effort. Choose $p \in U^\perp$ and write $p = p_0 + p_1$, where $p_0 \in W_0$, $p_1 \in W_1$. Then $p_1 \in W_1 = W_0^\perp \subset U^\perp$, so $p_0 \in U^\perp$. If $p_0(x) = c_0 + c_1x^2 + \cdots + c_nx^{2n}$, $c_0, \dots, c_n$ real numbers then $(p_0, x^{4t}) = 0$ for $t = 0, 1, 2, \dots$. That is $\int_{-1}^{1} x^{4t} p_0(x)\,dx = 0$ for $t = 0, 1, 2, \dots$. Evaluating the integrals for $t = 0, 1, \dots,$ we find

$$\begin{cases} c_0 + c_1/3 + c_2/5 + \cdots & +c_n/(2n+1) = 0 \\ c_0/5 + c_1/7 + \cdots & +c_n/(2n+5) = 0 \\ \vdots & \\ c_0/(4n+1) + c_1/(4n+3) + \cdots & +c_n/(6n+1) = 0 \end{cases} \qquad (*)$$

Now define the numbers

$$\begin{aligned} a_i &= 2i - 1 & i &= 1, \dots, n+1, \\ b_i &= 4(i-1) & i &= 1, \dots, n+1, \\ m_{ij} &= 1/(b_i + a_j) & i, j &= 1, \dots, n+1. \end{aligned}$$

If we consider $c_0, c_1, \dots, c_n$ as variables, then $(*)$ is a system of linear homogeneous equations with matrix (m_{ij}). Using the result in [**2**, p. 92, number 3][†], it follows that $\det(m_{ij}) \neq 0$ and so $(*)$ has no nontrivial solutions; that is $c_0 = c_1 = \cdots = c_n = 0$. Thus $p_0 = 0$, and $p = p_1 \in W_1$.

This shows that $U^\perp \subset W_1$ and completes the proof of (4).

We now have the desired counterexamples.

$$(U^\perp)^\perp = W_1^\perp = W_0 \neq U \qquad \text{and} \qquad U \oplus U^\perp \neq V.$$

REFERENCES

1. Northrup Fowler III. Elementary Counterexamples in Infinite Dimensional Inner Product Spaces, this MAGAZINE, 52, (1979), 96–97.
2. G. Pólya and G. Szegő, *Problems and Theorems in Analysis*, Vol. 2, Springer-Verlag, New York, 1976.
3. Albert Wilansky, *Functional Analysis*, Blaisdell Publishing Co., New York, 1964.

[†]I am indebted to Robert Hartwig for drawing my attention to [**2**].

The pth Root of an Analytic Function

LAWRENCE J. WALLEN
University of Hawaii at Manoa
Honolulu, HI 96822

This note is addressed to those who, like me, begin to feel uneasy when confronted with the necessity of extracting analytic pth roots of analytic functions. Recently, while teaching an introductory function theory course, I decided to confront the issue squarely by asking the following painfully unsubtle question: Precisely when does an analytic function have an analytic pth root?

A perusal of many texts, popular and unpopular, extant and extinct, failed to produce any that addressed the question. A (very limited) poll of function theorists was similarly unproductive though I finally found a Riemann surface expert who knew the answer. A problem in the popular text of Ahlfors (see [**1**, p. 147, problem 4]) reads in part: Show that a single-valued analytic branch of $\sqrt{1-z^2}$ can be defined in any region such that the points ± 1 are in the same component of the complement. This is clearly in the right direction but is never pursued further. At any rate, my conclusion was that although neither the question nor the answer is novel, they deserve to be more widely known by both professionals and students. The theorem we prove below permits, in a large variety of cases, a quick and painless determination of just which pth roots exist. First we need a little notation and one elementary fact.

D will denote a connected open set in the finite complex plane $\mathbb{C}$, and $A(D)$ will denote the class of functions analytic on D. To save space, we agree that "curve" means "piecewise smooth curve." Also, we write dz' for $dz/2\pi i$. We will use the fact (see [**2**, p. 219]), that for each closed curve C lying in D, $\int_C(f'(z)/f(z))\,dz'$ is an integer if $f \in A(D)$.

THEOREM. *If $f \in A(D)$, then f has a pth root in $A(D)$ if and only if, for each closed curve C lying in D and not passing through any zeros of f, p divides $\int_C(f'(z)/f(z))\,dz'$.*

Proof. The necessity of the condition is immediate, for if $g \in A(D)$ and $g^p = f$, then

$$\int_C \frac{f'(z)}{f(z)}\,dz' = p\int_C \frac{g'(z)}{g(z)}\,dz'$$

and since the last integral is an integer, necessity is proved.

For the sufficiency, first assume f has no zeros in D. Pick $z_0 \in D$ and let $\alpha^p = f(z_0)$. If $z \in D$ and γ is a curve going from z_0 to z, define

$$g(z,\gamma) = \alpha \exp\left(\frac{1}{p}\int_\gamma \frac{f'(\zeta)}{f(\zeta)}\,d\zeta\right).$$

If γ' is another such curve,

$$g(z,\gamma)/g(z,\gamma') = \exp\left(\frac{2\pi i}{p}\int_{\gamma-\gamma'} \frac{f'(\zeta)}{f(\zeta)}\,d\zeta'\right) = 1$$

by hypothesis. Consequently, we may omit the γ, and write

$$g(z)=\alpha\exp\left(\frac{1}{p}\int_{z_0}^{z}\frac{f'(\zeta)}{f(\zeta)}\,d\zeta\right).$$

The function g belongs to $A(D)$ since if $|w-z|<\varepsilon$,

$$g(w)=g(z)\exp\left(\frac{1}{p}\int_{z}^{w}\frac{f'(\zeta)}{f(\zeta)}\,d\zeta\right)$$

and the integral is an analytic function of w in this disc.

To show $g^p=f$, it suffices to show this in a small disc about z_0. Let

$$F(z)=g^p(z)=f(z_0)\exp\left[\int_{z_0}^{z}\frac{f'(\zeta)}{f(\zeta)}\,d\zeta\right].$$

Then,

$$F'=Ff'/f \quad\text{or}\quad F^2\frac{d}{dz}\left(\frac{f}{F}\right)=0$$

so $F=f$ (we are being a little fussy here to avoid mentioning "log"). Moreover, if $h^p=f$ with $h(z_0)=g(z_0)$, then $|g/h|\equiv 1$ so $g=h$.

Now we permit f to have finitely many zeros located at $z_1,\dots,z_n$. If C is a small circle about z_j, then our hypothesis shows z_j has multiplicity $m_j=pl_j$. If $P(z)$ is the polynomial $\prod_{j=1}^{n}(z-z_j)^{m_j}$, the function $h=f/P\in A(D)$ and has no zeros. Since $h'/h=(f'/f)-(P'/P)$ and $P=Q^p$ where $Q(z)=\prod_{j=1}^{n}(z-z_j)^{l_j}$, it follows that for all closed C, $\int_C(h'(z)/h(z))\,dz'$ is divisible by p. Hence, there is a $g\in A(D)$ with $g^p=h$ and $(Qg)^p=Pg^p=f$. Moreover, there is only one pth root of f whose value at z_0 is α.

In the general case where we allow infinitely many zeros but $f\not\equiv 0$, write $D=\bigcup_{k=1}^{\infty}D_k$ where each D_k is a connected open set, $\overline{D}_k\subseteq D$, $\overline{D}_k$ is compact, and $D_k\subseteq D_{k+1}$. Note that f has at most finitely many zeros on D_k. Let $z_0\in D_1$ with, again $\alpha^p=f(z_0)\neq 0$, and let g_k be the unique element in $A(D_k)$ such that $g_k^p=f$ on D_k and $g_k(z_0)=\alpha$. It is now immediate that the function g defined to be g_k on D_k is well defined, belongs to $A(D)$, and $g^p=f$. This finishes the proof.

As an example in the vein of the one mentioned in the second paragraph, let $f(z)=\prod_{j=1}^{6}(z-j)$. If the domain of f is $\mathbb{C}\setminus[1,2]\cup[3,4]\cup[5,6]$, then f has a square root but no other pth root. If the domain is $\mathbb{C}\setminus[1,3]\cup[4,6]$, f has only a cube root. Finally, for the domain $\mathbb{C}\setminus[1,6]$, f has a sixth root, and thus obviously a square and cube root also.

As regards the less thorny logarithm, we leave it to the reader to prove that f has an analytic logarithm if and only if f never vanishes and $\int_C(f'(z)/f(z))\,dz'=0$ for all closed C. Hence, we conclude that f has a logarithm precisely when it has a pth root for all p, provided $f\not\equiv 0$.

One final remark is in order. We have avoided mentioning geometric notions such as that of winding number. Of course, the theorem above can be stated in terms of this notion. A short proof of the theorem can be given using covering space arguments but this seems less elementary than the analytic argument we chose to give.

REFERENCES

1. L. Ahlfors, *Complex Analysis*, 2nd edition, McGraw Hill, New York, 1966.
2. W. Rudin, *Real and Complex Analysis*, 2nd edition, McGraw Hill, New York 1974.

Product Graphs are Sum Graphs

DEBORAH BERGSTRAND, KEN HODGES, GEORGE JENNINGS,
LISA KUKLINSKI, JANET WIENER
Williams College
Williamstown, MA 01267

FRANK HARARY
New Mexico State University
Las Cruces, NM 88003

Abstract. Given a set S of positive integers, the sum graph $G^+(S)$ has S as its node set with two nodes x and y adjacent if and only if $x+y \in S$. Then a sum graph is isomorphic to the sum graph of some S. Provided $1 \notin S$, the product graph of S and a product graph are defined similarly. We show that every sum graph is a product graph, and conversely.

1. Introduction Given a set S of positive integers, *the sum graph* $G^+(S)$ has S as its node set, with two nodes x and y adjacent whenever $x+y \in S$. Then a sum graph is isomorphic to the sum graph of some S. This concept was introduced in [2]. The notation and terminology of [1] is used. With the proviso that $1 \notin S$, the product graph $G^\times(S)$ and a product graph are now defined as expected.

The sum and product graphs for $S = \{2,3,4,6,8,12,24\}$ are illustrated in FIGURE 1.

THEOREM 1. *A graph G is a product graph if and only if G is a sum graph.*

Proof. The "easy half" of the proof is to show that every sum graph G is a product graph. Let $G = G^+(S)$ where $S = \{u, v, w, \dots\}$. Define the set $M = \{2^u, 2^v, 2^w, \dots\}$.

Since $0 \notin S$ we have $1 \notin M$. If $u+v=w$ then u, v are adjacent in G. As $2^u \times 2^v = 2^{u+v} = 2^w$, the nodes 2^u and 2^v of the product graph $G^\times(M)$ are adjacent. Hence in general two nodes are adjacent in $G^+(S)$ if and only if their corresponding nodes are adjacent in $G^\times(M)$. Thus we know that $G^\times(M) = G$, so G is also a product graph.

Conversely, we show that all product graphs are sum graphs. Let $G = (V, E)$ be a product graph with $V = \{x, y, z, \dots\}$ and $G = G^\times(M)$ where

$$M = \{m(x), m(y), m(z), \dots\},$$

and m is the multiplicative labeling function. Each label $m(x)$ has a unique prime factorization of the form,

$$m(x) = p_1^{r_1} p_2^{r_2} \cdots p_k^{r_k} \tag{1}$$

using the following notation:

(i) the numbers p_i are the first k distinct primes,

(ii) p_k is the largest prime factor of the product,

$$m(x) \cdot m(y) \cdot m(z) \cdot \cdots$$

The research of the authors was supported in part by: NSF grant DMS-8808695, NSF grant DMS-8802266, and the Computing Research Laboratory of New Mexico State University.

FIGURE 1
A sum graph and a product graph.

(iii) the r_i are nonnegative integers, $r_i = r_i(x)$. For each x, not all $r_i = 0$ since $m(x) \neq 1$. Let $b \in N$ be the minimum number such that

$$10^b > 2\max\{r_i(x): x \in V, i \in 1, \dots, k\}. \tag{2}$$

Now for each $i = 1, \dots, k$ define

$$a_i = 10^{bi} \tag{3}$$

and define the new node label $a(x)$, with the letter, a, standing for additive,

$$a(x) = r_1 a_1 + r_2 a_2 + \cdots + r_k a_k. \tag{4}$$

Note that $a(x)$ is positive since not all $r_i = 0$.

We illustrate this process with the product graph $G^\times(S)$ of FIGURE 1. Notice that each label of this graph is a product of powers of 2 and 3 and that the largest exponent required is 3. Thus $b = 1$, $a_1 = 10$ and $a_2 = 100$. The new additive labels are, respectively,

$$\{10, 100, 20, 110, 30, 120, 130\}.$$

To complete the proof we will show that for all nodes $x, y, z \in V$,

$$m(x)m(y) = m(z) \quad \text{if and only if } a(x) + a(y) = a(z).$$

Obviously if $m(u)m(v) = m(w)$, then $a(u) + a(v) = a(w)$ so all edges of the product graph are also edges of the sum graph in which the node values are given by (4).

We still need to prove the converse, i.e., that for all $x, y, z \in V$, $a(x) + a(y) = a(z)$ implies $m(x)m(y) = m(z)$.

The following observation is critical. Since (1) gives any product label, the values of a_i from (3) guarantee that $a(u)$ in (4) is the unique linear combination of $a_1, \dots, a_k$. This follows because each exponent $r_i(x)$ in (1) will appear literally as a sequence of digits in the decimal expansion of $a(x)$. This sequence will end at the (10^{bi})'th place of $a(x)$ and will not overlap with the digit sequence contributed by any other exponent due to the choice of b in (2).

Thus if $a(x) + a(y) = a(z)$ with

$$a(x) = \sum r_i a_i, a(y) = \sum s_i a_i, a(z) = \sum t_i a_i \tag{5}$$

then we must have for all $i = 1, \dots, k$

$$r_i + s_i = t_i \tag{6}$$

so $m(x)m(y) = m(z)$.

Remark. Sum graphs can also be defined over all the nonnegative integers including 0; likewise for product graphs over all the positive integers including 1. In either case, the additional sum graphs and product graphs that result are joins [1, p. 21] of the form $K_1 + G$ where G is a sum graph, i.e., G plus one new node adjacent to every node of G.

Acknowledgements. Ken Hodges, George Jennings, Lisa Kuklinski and Janet Wiener were undergraduates whose research under Dr. Bergstrand was part of the 1988 SMALL Geometry Project at Williams College.

REFERENCES

1. F. Harary, *Graph Theory*, Addison-Wesley Publishing Co, Reading, MA, 1969.
2. F. Harary, Sum graphs and difference graphs, *Congressus Numerantium* 72 (1990), 101–108.

Life on the Number Line

LEONARD GAMMA
(as told to Hans Sagan)

I was merrily skipping along the sequence $\{1 + 1/2 + 1/3 + \cdots + 1/n - \ln(n)\}$ when word reached me of a surprise Dedekind cut in my neighborhood. Fortunately, 0.57 covered for me 'til I could get into position. The rationals got so upset about it that they let some common factors creep into their numerators and denominators. It is the purpose of these unexpected drills to discourage us from moving around too much. Once, π got caught out of place in a nested interval exercise and, as punishment, was condemned to satisfy an algebraic equation of 17th degree. He is, of course, still at it and Algie 17 is not anywhere near to being satisfied. The Lindies, a cult devoted to transcendental meditation, have it that she never will be! The rationals get everywhere dense about these drills but they have a field day when we play open sets. It makes them feel wanted.

Our positive class consciousness prevents us from visiting on the other side of zero. Besides, who wants to associate with elements that have their roots in the complex plane. Rumor has it that the complex numbers are not even ordered! A bunch of primes keeps loitering in the interval $[0, 1)$, gazing in the imaginary direction and waiting for the zeta function to vanish. Maybe it will. Who is to say?

Binary drills are fun. Once, by mistake, $\sqrt{7}/5$ got into her ternary form and, with her twos sticking out, almost got pushed off the number line. Certain rationals consider it fashionable to appear as finite ternaries. Some are so scandalously short that one can almost see their ternary point. I just love to watch $2/3$ bounce around the Cantor set and brag about its cardinality!

Caught in a set of measure zero and trying to get out from under before a looming integral, I was whipped back into place with a Lebesgue measuring tape. Exhausted from the experience, I cower in my deleted neighborhood, wrapped into my integral representation and wonder about my rationality!*

---

*It is not known whether Euler's constant $\gamma = \lim\{1 + 1/2 + 1/3 + \cdots + 1/n - \ln n\} = -\int_0^\infty \ln t e^{-t}\, dt$ is rational or irrational.

Remark. Sum graphs can also be defined over all the nonnegative integers including 0; likewise for product graphs over all the positive integers including 1. In either case, the additional sum graphs and product graphs that result are joins [**1**, p. 21] of the form $K_1 + G$ where G is a sum graph, i.e., G plus one new node adjacent to every node of G.

Acknowledgements. Ken Hodges, George Jennings, Lisa Kuklinski and Janet Wiener were undergraduates whose research under Dr. Bergstrand was part of the 1988 SMALL Geometry Project at Williams College.

REFERENCES

1. F. Harary, *Graph Theory*, Addison-Wesley Publishing Co, Reading, MA, 1969.
2. F. Harary, Sum graphs and difference graphs, *Congressus Numerantium* 72 (1990), 101–108.

Life on the Number Line

LEONARD GAMMA
(as told to Hans Sagan)

I was merrily skipping along the sequence $\{1 + 1/2 + 1/3 + \cdots + 1/n - \ln(n)\}$ when word reached me of a surprise Dedekind cut in my neighborhood. Fortunately, 0.57 covered for me 'til I could get into position. The rationals got so upset about it that they let some common factors creep into their numerators and denominators. It is the purpose of these unexpected drills to discourage us from moving around too much. Once, π got caught out of place in a nested interval exercise and, as punishment, was condemned to satisfy an algebraic equation of 17th degree. He is, of course, still at it and Algie 17 is not anywhere near to being satisfied. The Lindies, a cult devoted to transcendental meditation, have it that she never will be! The rationals get everywhere dense about these drills but they have a field day when we play open sets. It makes them feel wanted.

Our positive class consciousness prevents us from visiting on the other side of zero. Besides, who wants to associate with elements that have their roots in the complex plane. Rumor has it that the complex numbers are not even ordered! A bunch of primes keeps loitering in the interval $[0, 1)$, gazing in the imaginary direction and waiting for the zeta function to vanish. Maybe it will. Who is to say?

Binary drills are fun. Once, by mistake, $\sqrt{7}/5$ got into her ternary form and, with her twos sticking out, almost got pushed off the number line. Certain rationals consider it fashionable to appear as finite ternaries. Some are so scandalously short that one can almost see their ternary point. I just love to watch $2/3$ bounce around the Cantor set and brag about its cardinality!

Caught in a set of measure zero and trying to get out from under before a looming integral, I was whipped back into place with a Lebesgue measuring tape. Exhausted from the experience, I cower in my deleted neighborhood, wrapped into my integral representation and wonder about my rationality!*

*It is not known whether Euler's constant $\gamma = \lim\{1 + 1/2 + 1/3 + \cdots + 1/n - \ln n\} = -\int_0^\infty \ln t e^{-t}\,dt$ is rational or irrational.

PROBLEMS

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

GEORGE GILBERT, *associate editor*
Texas Christian University

Proposals

To be considered for publication, solutions should be received by March 1, 1993.

1403. *Proposed by Richard Friedlander, University of Missouri–St. Louis, and Stan Wagon, Macalester College, Saint Paul, Minnesota.*

Simpson's aggregation paradox admits a simple baseball interpretation. It is possible for there to be two batters, Veteran and Youngster, and two pitchers, Righty and Lefty, such that Veteran's batting average against Righty is better than Youngster's average against Righty, and Veteran's batting average against Lefty is better than Youngster's average against Lefty, but yet Youngster's combined batting average against the two pitchers is better than Veteran's. Question: Can there be a double Simpson's paradox? That is, is it possible to have the situation just described and, *at the same time*, have it be the case that Righty is a better pitcher than Lefty against either batter, but Lefty is a better pitcher than Righty against both batters combined?

1404. *Proposed by Hillel Gauchman and Ira Rosenholtz, East Illinois University, Charleston, Illinois.*

Find the smallest prime which is not the difference (in some order) of a power of 2 and a power of 3.

1405. *Proposed by Hüseyin Demin, Middle East Technical University, Ankara, Turkey.*

Two circles inscribed in distinct angles of a triangle are *isogonally related* if the tangents from the third vertex not coinciding with the sides are symmetric with respect to the bisector of the third angle. Given three circles inscribed in distinct angles of a triangle, prove that if any two of the three pairs of circles are isogonally related then so is the third pair.

 We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals should be accompanied by solutions, if at all possible, and by any other information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution. An asterisk (*) next to a problem number indicates that neither the proposer nor the editors supplied a solution.

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

Solutions and new proposals should be mailed in duplicate to Loren Larson, Department of Mathematics, St. Olaf College, 1520 St. Olaf Ave., Northfield, MN 55057-1098 or mailed electronically via fax: (507) 663-3549 or e-mail: larson@stolaf.edu.

1406. *Proposed by Detlef Laugwitz, Fachbereich Mathematik Technische Hochschule, Darmstadt, Germany.*

Define a sequence $(a_n)_{n\geq 1}$ by

$$a_1=\sqrt{3}, \qquad a_n=a_{n+1}(3-a_{n+1}^2), \qquad 0<a_n\leq a_1 \qquad \text{for } n=1,2,3,\dots .$$

Show that $\lim_{n\to\infty} 3^n a_n$ exists and find its value.

1407. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

Determine the maximum value of the sum

$$x_1^p+x_2^p+\cdots+x_n^p-x_1^q x_2^r-x_2^q x_3^r-\cdots-x_n^q x_1^r,$$

where p,q,r are given numbers with $p\geq q\geq r>0$ and $0\leq x_i\leq 1$ for all i.

Quickies

Answers to the Quickies are on page 272.

Q794. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

The general problem of Appolonius is to draw a circle tangent to three given circles. Special cases ensue when all or some of the circles are replaced by points or lines. Solve the problem in the case of two points O, Q, and a circle C, where O is the center of C and Q is an interior point of C.

Q795. *Proposed by Norman Schaumberger, Douglaston, New York.*

Find a solution in positive integers of

$$a^4+b^7+c^9=d^{11}.$$

Q796. *Proposed by Tadhg Creedon (student) and Desmond MacHale, University College, Cork, Ireland.*

Let $G=\{a_1,a_2,\dots,a_n\}$ be a finite group with binary operation $*$. Define an $n\times n$ matrix $M=(m_{ij})$ by $m_{ij}=1$ if $a_i * a_j = a_j * a_i$ and $m_{ij}=0$ otherwise. Show that M is invertible if and only if $n=1$.

Solutions

Triangular numbers **October 1991**

1378. *Proposed by Stephen G. Penrice, Arizona State University, Tempe, Arizona.*

Let $(i)_j$ denote the falling product $i(i-1)\cdots(i-j+1)$ and let $(i)_0=1$. Show that for all positive integers n and k

$$\frac{(n+k)_{k+1}}{2(k!)^2}\sum_{i=1}^{k}(k)_{i-1}(n+k-i)_{k-i}$$

is a triangular number.

I. *Solution by R. S. Tiberio, Natick, Maine.*

Writing $(i)_j$ as $i!/(i-j)!$, we may write the given expression as

$$\frac{1}{2}\binom{n+k}{n}\sum_{i=1}^{k}\binom{n+k-i}{n-1}.$$

Now using the standard additive property of binomial coefficients, we can write this as

$$\frac{1}{2}\binom{n+k}{n}\sum_{i=1}^{k}\left(\binom{n+k-i+1}{n}-\binom{n+k-i}{n}\right),$$

which telescopes to

$$\frac{1}{2}\binom{n+k}{n}\left(\binom{n+k}{n}-1\right),$$

a triangular number.

II. *Solution by Reiner Martin (student), University of California at Los Angeles, Los Angeles, California.*

We claim that

$$n\sum_{i=1}^{k}(k)_{i-1}(n+k-i)_{k-i}=(n+k)_k-k!.$$

The right side gives the number of ways of choosing k elements ordered and without repetition from $\{1,2,\dots,n+k\}$, choosing at least one element from $\{1,2,\dots,n\}$. For such a choice, let the i-th element be the first one chosen from $\{1,2,\dots,n\}$. There are $(k)_{i-1}$ possibilities choosing the first $i-1$ elements from $\{n+1,n+2,\dots,n+k\}$, and n possibilities to choose the ith element from $\{1,2,\dots,n\}$, and $(n+k-i)_{k-i}$ possibilities to choose the last $k-i$ elements from the remaining ones. This proves our identity.

Using $(n+k)_{k+1}=n(n+k)_k$, we now get

$$\frac{(n+k)_{k+1}}{2(k!)^2}\sum_{i=1}^{k}(k)_{i-1}(n+k-i)_{k-i}=\frac{(n+k)_k}{2k!}\left(\frac{(n+k)_k}{k!}-1\right)$$
$$=\frac{1}{2}\binom{n+k}{k}\left(\binom{n+k}{k}-1\right),$$

which is a triangular number.

Also solved by Michael H. Andreoli, Seung-Jin Bang (Korea), Harvey L. Berger, J. C. Binz (Switzerland), William Chen and Edward T. H. Wang, Con Amore Problem Group (Denmark), Robert L. Doucette, Arne Fransén (Sweden), Russell Jay Hendel, Albert Kurz (student), Carl Libis, Norman Lindquist, Jack McCown, James L. Parish, Volkhard Schindler (Germany), Heinz-Jürgen Seiffert (Germany), John S. Sumner, Michael Vowe (Switzerland), Robert J. Wagner, Chris Wildhagen (The Netherlands), David Zhu (student), and the proposer.

Self-seeking operations **October 1991**

1379. *Proposed by John O. Kiltinen, Northern Michigan University, Marquette, Michigan.*

Call an operation $*$ on a nonempty set A *self-seeking* if every permutation of A is an automorphism from $(A, *)$ to $(A, *)$. Such operations have no isomorphic copies on the set other than themselves. Describe all the self-seeking operations, if any, on A.

Solution by Fred Dodd, University of South Alabama, Mobile, Alabama.

The operations $*_1$ and $*_2$ given by $x *_1 y = x$ and $x *_2 y = y$ for each $x, y \in A$ are self-seeking since $\sigma(x *_1 y) = \sigma(x) = \sigma(x) *_1 \sigma(y)$ and $\sigma(x *_2 y) = \sigma(y) = \sigma(x) *_2 \sigma(y)$ for each permutation σ. We show that if $|A| \neq 2, 3$ these exhaust the self-seeking operations; and for $|A| = 3$, $|A| = 2$, there are three and four self-seeking operations respectively. The case $|A| = 1$ is trivial and so we assume that $|A| \geq 2$.

Lemma. If a, b are two distinct elements of A and $*$ is self-seeking then $*$ is completely determined by the values $a * a$ and $a * b$. Moreover, if $|A| > 2$ then $a * a = a$; and if $|A| > 3$ then $a * b = a$ or $a * b = b$.

Proof of Lemma. If x and y are any two distinct elements of A there is a permutation τ with $\tau(a) = x$ and $\tau(b) = y$. Thus, $x * x = \tau(a) * \tau(a) = \tau(a * a)$ and $x * y = \tau(a) * \tau(b) = \tau(a * b)$. If $|A| > 2$ there is a permutation ϕ such that $\phi(a) = a$ and $\phi(a * a) \neq a * a$ when $a * a \neq a$. Thus, since $a * a = \phi(a) * \phi(a) = \phi(a * a)$ it follows that $a * a = a$ when $|A| > 2$. If $|A| > 3$ there is a permutation θ such that $\theta(a) = a$, $\theta(b) = b$, and $\theta(a * b) \neq a * b$ when $a * b \notin \{a, b\}$. Thus, since $a * b = \theta(a) * \theta(b) = \theta(a * b)$ we must have $a * b = a$ or $a * b = b$ when $|A| > 3$.

It is immediate from the lemma that $*_1$ and $*_2$ are the only self-seeking operations when $|A| \neq 2, 3$. If $A = \{a, b, c\}$, the lemma says there are at most three self-seeking operations. It is easy to verify that the operation $*$ given by $x * x = x$, $x * y = z$ where $\{x, y, z\} = \{a, b, c\}$ is a third self-seeking operation. Finally, if $A = \{a, b\}$, the lemma says there are at most four self-seeking operations. It is clear that the two operations given by

$$x *_3 x = y, \qquad x *_3 y = x$$

and

$$x *_4 x = y, \qquad x *_4 y = y$$

where $\{x, y\} = \{a, b\}$ give third and fourth self-seeking operations.

Also solved by S. F. Barger, David Callan, Con Amore Problem Group (Denmark), Robert L. Doucette, Furman University Problem Solving Group, William E. Gould, Colonel Johnson Jr., Emil F. Knapp, Dean Larson, Reiner Martin (student), Jean-Marie Monier (France), Raphael S. Ryger, John S. Sumner, Trinity University Problem Group, C. Wildhagen, and the proposer.

Lower bound for positive root **October 1991**

1380. *Proposed by Thoddi C. T. Kotiah, Utica College of Syracuse University, Utica, New York.*

Let $P(x) = x^n - a_1 x^{n-1} - a_2 x^{n-2} - \cdots - a_n$, where the a_i are nonnegative real numbers, not all zero. Let $s = \sum_{i=1}^n a_i$ and $v = \sum_{i=1}^n i a_i$. Prove that a lower bound for the positive real zero of $P(x)$ is $s^{s/v}$.

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

We have to prove that if $t > 0$, and

$$(1/t)^n - a_1(1/t)^{n-1} - a_2(1/t)^{n-2} - \cdots - a_n = 0,$$

or

$$a_1 t + a_2 t^2 + \cdots + a_n t^n = 1, \tag{1}$$

then $s^{s/v} \le 1/t$, or,

$$s^s t^v \le 1. \tag{2}$$

So assume (1), and for $i = 1, 2, \ldots, n$ define

$$\alpha_i = a_i/s. \tag{3}$$

Then

$$\alpha_i \ge 0 \ (i = 1, 2, \ldots, n), \quad \text{and} \quad \sum_{i=1}^{n} \alpha_i = 1. \tag{4}$$

Now (2) may be written as

$$(st)^{a_1}(st^2)^{a_2} \cdots (st^n)^{a_n} \le 1,$$

or (taking the sth root on both sides, and using (3)),

$$(st)^{\alpha_1}(st^2)^{\alpha_2} \cdots (st^n)^{\alpha_n} \le 1,$$

or again, as

$$\alpha_1(-\log(st)) + \alpha_2(-\log(st^2)) + \cdots + \alpha_n(-\log(st^n)) \ge 0. \tag{5}$$

Now $-\log x$ is a convex function so by Jensen's inequality and (4), the left side of (5) is at least

$$-\log\left(\alpha_1 st + \alpha_2 st^2 + \cdots + \alpha_n st^n\right),$$

or (see (3)),

$$-\log\left(a_1 t + a_2 t^2 + \cdots + a_n t^n\right),$$

which is 0 (see (1)), so (5) is true and then so is (2).

II. *Solution by Reiner Martin (student), University of California at Los Angeles, Los Angeles, California.*

By an obvious induction on n, using the derivative, we see that P has exactly one positive real zero, and $P' > 0$ at this point. So it is enough to show that $P(s^{s/v}) \le 0$.

By the weighted arithmetic mean–geometric mean inequality, we have

$$\frac{\sum_{i=1}^{n} a_i s^{(n-i)s/v}}{s} \le \left(\prod_{i=1}^{n} s^{a_i(n-i)s/v}\right)^{1/s} = s^{\sum_{i=1}^{n} a_i(n-i)/v} = s^{ns/v-1}.$$

Thus

$$P(s^{s/v}) = s^{ns/v} - \sum_{i=1}^{n} a_i s^{(n-i)s/v} \ge 0.$$

Also solved by Arnold Adelberg and Eugene A. Herman, Robert A. Agnew, Paul Bracken (Canada), Robert L. Doucette, F. J. Flanigan, G. Ladas, F. C. Rembis, Heinz-Jürgen Seiffert (Germany, two solutions), John S. Sumner, and the proposer.

Ladas cites a paper by David H. Anderson, "Estimation and Computation of the Growth Rate in Leslie's and Lotka's Population Models," *Biometrics* 31, September 1975, pp. 701–718, which contains this, and several other bounds, for the unique positive root of this polynomial.

Diophantine equation **October 1991**

1381. *Proposed by Mihály Bencze, Braşov, Romania.*

Find all integer solutions of the following (n and k are positive integers).

a. $(x+y)^{2n+1}=x^{2n}+y^{2n}$

b*. $(x_1+x_2+\cdots+x_k)^{kn+k-1}=x_1^{kn}+x_2^{kn}+\cdots+x_k^{kn}$

I. *Solution to (a) by Keith Neu (student), Louisiana State University in Shreveport, Shreveport, Louisiana.*

We will show that the solutions to (a) are $x=y=0$, or,

$$x=(b+1)\left((b+1)^{2n}+b^{2n}\right),$$

$$y=-b\left((b+1)^{2n}+b^{2n}\right),$$

where b is any integer.

The binomial theorem implies that x and y must be of opposite signs, unless both are zero. Since the equation is symmetric in x and y, without loss of generality, we may assume $x\geq 0$ and $y\leq 0$. Let $z=-y$, so that the equation becomes $(x-z)^{2n+1}=x^{2n}+z^{2n}$, with $x\geq 0$ and $z\geq 0$. The right side of this equation being nonnegative implies $x\geq z\geq 0$. Clearly $x=z=0$ is a solution, and if $x=z$, then both must equal 0. So assume that $x>z>0$.

Let $x=ag$, $z=bg$, where $\gcd(a,b)=1$. We may assume that $g>0$, and the above assumptions on x and z imply that $a>b>0$. Inserting $x=(a/b)z$ in our equation and cancelling a factor of z^{2n} reveals

$$z=\frac{b(a^{2n}+b^{2n})}{(a-b)^{2n+1}}.$$

Since a and b are relatively prime, so are $a-b$ and b, as well as $(a-b)^{2n+1}$ and b. Thus, for z to be an integer, $(a-b)^{2n+1}$ divides $a^{2n}+b^{2n}$. We claim this implies $a-b=1$. Assume $a-b=1$. Form

$$a-b=\prod_{m=1}^{q}p_m^{i_m}\quad\text{and}\quad a^{2n}+b^{2n}=\prod_{m=1}^{q}p_m^{r_m}\prod_{m=q+1}^{K}p_m^{r_m},$$

with $r_m\geq i_m(2n+1)\geq 3$ for $m=1,2,\dots q$.

We claim that if p_l is one of the primes in the factorization of $a-b$, then $p_l=2$. Since $a-b\equiv 0\pmod{p_l}$ and since $\gcd(a,b)=1$, we may assume $a\equiv j\pmod{p_l}$, and $b\equiv j\pmod{p_l}$, where $0<j<p_l$. This implies that $a^m+b^m\equiv j(a^{m-1}+b^{m-1})\pmod{p_l}$, for $m=1,2,\dots,2n$. By assumption $a^{2n}+b^{2n}\equiv 0\pmod{p_l}$. Since $\gcd(j,p_l)=1$, this implies $a^{2n-1}+b^{2n-1}\equiv 0\pmod{p_l}$. Continuing in this fashion, it is clear that $2\equiv 0\pmod{p_l}$, which implies that p_l divides 2, as claimed.

We have shown that

$$a^{2n}+b^{2n}=2^R\prod_{m=q+1}^{K}p_m^{r_m},$$

for some $R \geq 3$. In particular, 8 divides $a^{2n} + b^{2n}$. On the other hand, since both a and b must be odd, 2 divides $a^{2n} + b^{2n}$, but 4 does not. This contradiction implies that $a - b = 1$.

Inserting $a = b + 1$ into our expression for z and x and recalling that $y = -z$ yields the solution as claimed.

II. *Solution to (a) by the Con Amore Problem Group, Royal Danish School of Educational Studies, Copenhagen, Denmark.*

Let $u = x + y$. For $u = 0$, the equation has one solution, namely $x = 0$, $y = 0$. For $u \neq 0$, let $s = x/u$ and $t = y/u$ and consequently $s + t = 1$. With $x = su$ and $y = tu = (1 - s)u$ the equation becomes

$$u^{2n+1} = s^{2n}u^{2n} + (1 - s)^{2n}u^{2n},$$

which reduces to

$$u = s^{2n} + (1 - s)^{2n} = s^{2n} + (1 - s)^{2n},$$

and we find

$$x = s\left(s^{2n} + (s - 1)^{2n}\right),$$

$$y = -(s - 1)\left(s^{2n} + (s - 1)^{2n}\right). \qquad (*)$$

If s is an integer, x and y are integers. Conversely, if x and y are integers, $u = x + y$ is also an integer, and $s = x/u$ is a rational number. Let $s = p/q$, where p and q are relatively prime integers, and q is positive. We will show that $q = 1$, so that s is an integer.

We have

$$u = \left(\frac{p}{q}\right)^{2n} + \left(\frac{p}{q} - 1\right)^{2n} = \frac{p^{2n} + (p - q)^{2n}}{q^{2n}},$$

and therefore,

$$uq^{2n} = p^{2n} + \left(p^{2n} - \binom{2n}{1}p^{2n-1}q + \cdots - \binom{2n}{2n-1}pq^{2n-1} + q^{2n}\right)$$

$$= 2p^{2n} + qP(p, q),$$

where $P(p, q)$ is a polynomial in p and q with integer coefficients. Therefore q divides $2p^{2n}$, and since p and q are relatively prime, it follows that q divides 2; and because of $q > 0$ we then have $q = 2$ or $q = 1$.

If $q = 2$ then p is odd, and

$$u = \frac{p^{2n} + (p - 2)^{2n}}{4^n}.$$

One of the two consecutive odd numbers p and $p - 2$ must be congruent to 1 and the other to -1 modulo 4. Both p^{2n} and $(p - 2)^{2n}$ are then congruent to 1, and the numerator $p^{2n} + (p - 2)^{2n}$ is congruent to 2 modulo 4. Therefore the denominator 4^n cannot divide $p^{2n} + (p - 2)^{2n}$, and this is a contradiction, since u is an integer. This shows that $q = 2$ is impossible, and we are left with $q = 1$, so that s is an integer.

Besides $x=0$, $y=0$, the only possible solutions to the original equation are therefore those given by $(*)$, where s is an arbitrary integer, and it is straightforward to check that they are indeed solutions.

Also solved by Seung-Jin Bang (Korea), Pierre Barnouin (France), Ty Le (two solutions), Jeff Nelson (student), David Stone, John S. Sumner and Kevin L. Dove, Michael Vowe, and the proposer. No solutions were received to part (b).

Zeros of order n **October 1991**

1382. *Proposed by Michael Golumb, Purdue University, West Lafayette, Indiana.*

Suppose f is a real-valued function of class C^∞ near $x_0 \in \mathbf{R}$, and g is a real-valued function of class C^∞ near $f(x_0)$. Prove that if $g \circ f - e$ (e the identity function) has a zero of order n ($1 \le n \le \infty$) at x_0, then $f \circ g - e$ has a zero of the same order at $f(x_0)$.

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

Let $\alpha = g \circ f - e$ and $\beta = f \circ g - e$. For x near x_0, $\beta(f(x)) = f(x+\alpha(x)) - f(x)$.

First assume that $n=1$. We have $\alpha(x_0)=0$, which implies $g(f(x_0)) = x_0$. With these we may show that $\beta(f(x_0)) = 0$ and $\beta'(f(x_0)) = \alpha'(x_0) \neq 0$.

Next assume that $1 < n < \infty$. By induction we can show that, for all finite $k \ge 1$, there are functions $\lambda_0, \dots, \lambda_{k-1}$; $\varphi_0, \dots, \varphi_{k-1}$ of class C^∞ near x_0 such that the k-th derivative of $\beta(f(x))$ equals both

$$\alpha^{(k)}(x) f'(x+\alpha(x)) + f^{(k)}(x+\alpha(x)) - f^{(k)}(x) + \sum_{i=0}^{k-1} \alpha^{(i)}(x) \lambda_i(x) \tag{1}$$

and

$$\beta^{(k)}(f(x))(f'(x))^k + \sum_{i=0}^{k-1} \beta^{(i)}(f(x)) \varphi_i(x). \tag{2}$$

Assuming that $\beta^{(0)}(f(x_0)) = \beta^{(1)}(f(x_0)) = \cdots = \beta^{(k-1)}(f(x_0)) = 0$, $1 \le k \le n$, we have, by equating (1) and (2) and evaluating at $x = x_0$,

$$\beta^{(k)}(f(x_0))(f'(x_0))^k = \begin{cases} 0 & 1 \le k < n, \\ \alpha^{(n)}(x_0) f'(x_0) & k = n. \end{cases}$$

Now $\alpha'(x_0) = 0$ implies that $f'(x_0) \neq 0$. Since $\beta^{(0)}(f(x_0)) = 0$, we have, by induction,

$$\beta^{(i)}(f(x_0)) = 0, \qquad i = 0, 1, \dots, n-1, \quad \text{and} \quad \beta^{(n)}(f(x_0)) \neq 0.$$

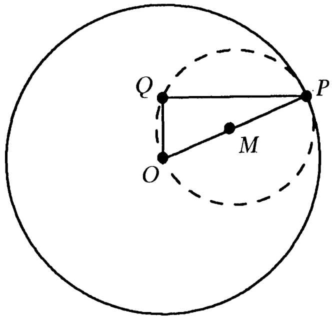
For $n = \infty$ the above argument can be modified to yield $\beta^{(i)}(f(x_0)) = 0$, $i \ge 0$.

Also solved by David Callan, Jean-Marie Monier (France), John S. Sumner, and the proposer.

Answers

Solutions to the Quickies on page 266.

Q794. The center of the desired tangent circle is the midpoint of the hypotenuse of the right triangle drawn as in the figure.



Q795. Since

$$\left(3^{(10!+1)/11}\right)^{11}=3^{10!+1}=3^{10!}+3^{10!}+3^{10!}=\left(3^{10!/4}\right)^{4}+\left(3^{10!/7}\right)^{7}+\left(3^{10!/9}\right)^{9},$$

we can take $a=3^{10!/4}$, $b=3^{10!/7}$, $c=3^{10!/9}$, and $d=3^{(10!+1)/11}$.

Q796. The result is trivially true for $n=1$, so suppose $n>1$. If there exists $x\in G$, $x\neq x^{-1}$, then, since the centralizer of x^{-1} is equal to the centralizer of x, M has two equal rows, so M is not invertible. So suppose that $x=x^{-1}$ for all $x\in G$. Then, for all $x,y\in G$, $x*y=(x*y)^{-1}=y^{-1}*x^{-1}=y*x$, so G is abelian. But this means that all rows of M are equal, so M is not invertible.

REVIEWS

PAUL J. CAMPBELL, *editor*
Beloit College

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

Kolata, Gina, New short cut found for long math proofs, *New York Times* (7 April 1992) (National Edition) B5, B9. Cipra, Barry, "Transparent" proofs help solve opaque problems, *Science* 256 (15 May 1992) 970–971. Peterson, Ivars, Holographic proofs, *Science News* 141 (6 June 1992) 382–383. Babai, Lászl6, Transparent proofs, *Focus* 12(3) (June 1992) 1–2.

These articles describe an exciting new result in computer science that may have important ramifications for pure mathematics. The result shows how to transform a formal proof (of, say, n symbols) into a form (with $n^{1.5}$ symbols) in which any error is amplified to the point of being easily detectable. The error is easily detectable because it appears in the transformed proof almost (though not absolutely) everywhere. For a given level of confidence, a fixed amount of spot checking—regardless of the length of the proof—will suffice, so it is no harder to check a long proof than a short one. (The constant amount of checking is an improvement made after Babai wrote his article.) Because an error will appear almost everywhere, the new proofs have been dubbed *transparent* or *holographic* proofs. The same approach can be used to verify the results of computations by unreliable software on unreliable hardware, using (in Babai's words) "a goldplated Macintosh monitoring the operation of a herd of Crays." Babai does not delve into the equally important implications for computer science, while Cipra explains them at a level beyond the vagueness of Kolata and Peterson. The transparent proofs result can be used to show that it is impossible to compute even approximate solutions to the maximum clique problem (an NP-complete problem to find the largest group of people at a party, each of whom knows each other person in the group). In fact, any algorithm for estimating maximum clique size would lead to solving the original problem. In other words, if there is a polynomial-time algorithm for the approximate maximum clique size problem, then $\mathcal{P} = \mathcal{NP}$ (which equivalence virtually all researchers are convinced is false). The methods leading to the new result are based on the efforts of a number of researchers and on the concepts of zero-knowledge proofs and interactive proofs. Pure mathematicians can take heart in Babai's reassurance that "We are not able to prove theorems. We are only able to check proofs."

Holland, John H., Genetic algorithms, *Scientific American* 267:1 (July 1992) 66–72; Riolo, Rick L., The Amateur Scientist: Survival of the fittest bits, *ibid.*, 114–116.

A genetic algorithm is a kind of iterative process for finding good (and ideally optimal) solutions to an optimization problem. The algorithm mimics evolution, operating on successive sets of data ("generations" of "individuals") by using analogues of gene crossover, natural selection, and mutation to produce each generation from the previous one. Selection is based on a fitness function, with fitness correlated with closeness to optimal solution to the problem. Holland, one of the founders of the field of genetic algorithms, may cause some confusion by referring both to evolution of the data (as a genetic algorithm proceeds) and to the computer program itself evolving (as if it were self-modifying code). Riolo describes a specific example of a genetic algorithm to find the maximum value of a function, including fragments of code in C (the entire program is available on request).

Peterson, Ivars, Striking pay dirt in prime-number terrain, *Science News* 141 (4 April 1992) 213. Merci—a new Mersenne prime, *Science* 256 (10 April 1992) 175. New Mersenne prime discovered, *Focus* 12(3) (June 1992) 3.

The March discovery of the thirty-second known Mersenne prime, 2^{756839}, which is considerably larger than the ones previously discovered, is described in greater detail. Eberhart's conjecture on prevalence of Mersenne primes (see Paulo Ribenboim, *The Book of Prime Number Records*, 2nd ed., Springer-Verlag, 1989, pp. 332–333) suggests that this one in fact may be M_{33} or even M_{34}, leaving one or more yet to be discovered in the gap between M_{31} and the new giant.

Stewart, Ian, Mathematical Recreations: The riddle of the vanishing camel, *Scientific American* 266:6 (June 1992) 122–124.

Ian Stewart has mastered the knack of wrapping a mathematical problem and its solution into a fictional tale and dialogue, usually with a fillip at the end. Here he considers the problem of dividing a legacy of camels among heirs, in proportions that do not work out into integer numbers of camels but also do not add to 1. The trick solution is to add an extra camel, perform the division, and have the extra camel left over. For what proportions, and what numbers of camels, is this possible? The answer involves enumerating solutions to Egyptian-fraction problems.

Stewart, Ian, Mathematical Recreations: The Interplanetary Olympics, *Scientific American* 267:2 (August 1992) 122–124.

Stewart once again spins an engaging tale; this time, it is around the question of how to compare athletic performances of individuals on different celestial bodies. The gravity, diameter, and air resistance vary from one body to another: Which of these influence what sports? How can performances be compared? And despite the attempted equalization, how come the Lunar team won every event? Students may enjoy such motivation for the simple physics discussed in calculus, including the optimal angle to maximize distance of a projectile.

Chaum, David, Achieving electronic privacy, *Scientific American* 267:2 (August 1992) 96–101.

U.S. telephone companies have begun to provide caller ID service, and customers have responded by expressing a desire for "electronic privacy." This article describes a way to realize such privacy in many other realms in which we now do business, by avoiding the linking of databases on a common key such as a Social Security number. A "smart" credit card (one containing a microcomputer) could use contemporary cryptographic methods—in particular, the "blind digital signature"—to ensure electronic privacy of the parties to a transaction, as well as to do a better job of preventing fraud.

Ng, K.C., and Keith H. Bierman, Getting the right answer for trigonometric functions, *SunProgrammer* 1(2) (Spring 1992) 8–10.

Here's a problem to try on your calculator or computer's software: What's the value of $\sin 10^{22}$? The correct answer is $-0.85220049\ldots$. The authors give the results for 30 different machines (few get it right). They also cite the argument-reduction algorithm (discovered in 1982) that makes correct results over the entire domain both possible and demandable. The authors (who work for Sun) deplore an AT&T standard that requires that for x sufficiently large, the software must return 0 for both $\cos x$ and $\sin x$ together with an error message of total loss of significance. But why settle for less than the correct answer?

Stroh, Michael, Breathing lessons, *Science News* 141 (9 May 1992) 314–315.

Where do particles, such as dust and asbestos—or even aerosol droplets from an inhaler—go once they enter a human lung? You can't experiment on people, and other animals have varying lung shape. S. Anjilvel (Duke University) created a mathematical model to answer this question, using the Navier-Stokes equations to calculate forces inside the lung. The model predicts where particles will get to (mucus, deep lung, or breathed back out) and where they will deposit (e.g., at places where airways branch). Says Anjilvel: "Biologists ... need these sophisticated mathematical tools, and mathematicians are in the position to provide them."

Cipra, Barry A., All theorems great and small, *SIAM News* 25(4) (July 1992) 28, 26.

"Researchers have made progress on a liar's version of Twenty Questions; on a problem of Erdös concerning the number of distinct distances between points in the plane; in the theory of self-avoiding walks; and on a fundamental question concerning the number of closed geodesics on distorted spheres."

Burrill, Gail, et al., *Data Analysis and Statistics Across the Curriculum*, NCTM, 1992; viii + 88 pp, $15 (P). ISBN 0–87353–329–1

This book was written as a resource for high-school teachers (e.g., in implementing the NCTM Curriculum and Evaluation Standards), but it will also prove valuable to the many instructors who teach introductory statistics at the college level. Particularly useful may be the 18 activities that are described (including ready-to-copy student handouts), the chapter on student projects, and the discussion of assessing statistical understanding. Soon students may arrive at college having already done all the exploratory and inferential statistics in this book (or another); will they need to take our introductory college course?

Steen, Lynn Arthur (ed.), *Heeding the Call for Change: Suggestions for Curricular Action*, MAA, 1992; x + 247 pp, $20 (P). ISBN 0–88385–079–6

This book responds to the MAA's *A Call for Change: Recommendations for the Mathematical Preparation of Teachers of Mathematics* (1991) and goes further in suggesting changes in how we teach *all* students in college mathematics courses. The papers and opinions here are based mostly on focus groups of experts; the subject areas include statistics, geometry, environmental mathematics, quantitative literacy, multiculturalism, assessment of student outcomes, and educational research. Reading these contributions is like taking part in conversations with leading authorities: stimulated, provoked, assured, you feel ready and inspired to act.

Aho, Alfred V., and Jeffrey D. Ullman, *Foundations of Computer Science*, Freeman, 1992; xiii + 765 pp, $56.95. ISBN 0–7167–8233–2

"This book was motivated by the desire ... to further the evolution of the core course in computer science." It features "three working methodologies or processes—theory, abstraction, design—as fundamental to all undergraduate programs in the discipline," and emphasizes the "key recurring concepts" of " conceptual and formal models, efficiency, and levels of abstraction." The content combines a first course in data structures (CS2) with a course in the discrete mathematics that the authors feel a computer user "really needs." Important unifying principles are design algebras (algebras as the principle for design of circuits, subsystems, and programs) and recursion. The result is a book whose flavor and content are up-to-date with curricular recommendations of the ACM and IEEE and distinctly different from the many clones for CS2. If you teach computer science, you need to experience this book.

Foley, James D., Andries van Dam, Steven K. Feiner, and John F. Hughes, *Computer Graphics: Principles and Practice*, 2nd ed., Addison-Wesley, 1990; xxiii + 1174 pp. ISBN 0-201-12110-7

Eight years later and twice the size, this is the second edition of a classic text in computer graphics. A 28-page appendix reviews Newton's method and the necessary linear algebra (vector spaces, affine spaces, linear transformations, eigenvalues and eigenvectors). A major change is abandonment of vector graphics (based on point- and line-plotting commands) in favor of raster graphics (based on pixels), a consequence of cheaper memory and the predominance of bit-mapped graphics.

Pickover, Clifford A., *Computers and the Imagination: Visual Adventures Beyond the Edge*, St. Martin's, 1991; xix + 424 pp, $29.95. ISBN 0-312-06131-5

This is one of the the most imaginative and visually interesting books of the year, featuring old engravings, fractals, photographs, and computer graphics of all kinds. Pickover's fertile imagination includes butterfly curves, growing your own font, twisted mirror worlds, monkey curves and spirals, the lute of Pythagoras, cakemorphic integers, earthworm algebra, and undulating pseudofareymorphic integers. (You probably don't need this book if these are all part of your everyday experience.) The main topics are simulation (algorithms are provided for the reader to implement), exploration, visualization (computer graphics has become indispensable), speculation (what would be the social and political impact of a soda-can-sized supercomputer? or of a personal computer in the year 1900?), invention (e.g., anti-dyslexic fonts, speech synthesis grenades), imagination, fiction, exercises, and resources. This is truly a book of *ideas*.

Ewing, John H., and F.W. Gehring (eds.), *Paul Halmos: Celebrating 50 Years of Mathematics*, Springer-Verlag, 1991; viii + 320 pp, $49. ISBN 0-387-97509-8

Just in case Halmos's *I Want to Be a Mathematician: An Automathography* (Springer-Verlag, 1985; MAA, 1988) didn't tell you all you wanted to know about him, here are interviews, photographs, reminiscences, his bibliography, and mathematical papers by friends (some related to his work, several clearly not). Young researchers may be encouraged in noting that Halmos's reputation is based on quality, not quantity: a couple of papers a year, and (in maturity) a book every three years or so. "Every single word that I publish I write at least six times."

Lucas, Édouard, *Récréations Mathématiques*, 2nd ed., 4 vols., and *L'Arithmétique Amusante*, Blanchard, 1979, 1992; xxiii + 965 pp (P) and 266 pp (P). ISBN 2-85367-121-6 [printed on vol. 1], 121-6 [quoted in vol. 1 as ISBN for vol. 2], 122-4, 123-2, and 177-1.

Originally published 1892–1895, these volumes by the famous nineteenth-century master of recreational mathematics are all finally available again. Here you will find river-crossing puzzles, Eulerian paths, mazes, the eight Queens problem, peg solitaire, the Chinese rings puzzle, the 15 puzzle, magic squares, dominos, mosaics, interdecomposability of areas, circular permutations, finger arithmetic, calculating machines (with Lucas's own invention, the Tower of Hanoi), a perpetual calendar, figurate numbers, and mathematical analysis of numerous other board games (including checkers to lose, Hamilton's Icosian Game, and the Mill). Aficionados will continue to wish for an English edition, with annotations and bibliography to bring the analyses up to date. (My thanks to Jeff Shallit.)

Watts, Robert G., and A. Terry Bahill, *Keep Your Eye on the Ball: The Science and Folklore of Baseball*, Freeman, 1990; x + 213 pp, $18.95. ISBN 0-7167-2104- X

Two engineers consider the physics of baseball: curve balls, fly balls, keeping your eye on the ball, and the optimal weight of a bat (moral: use an aluminum one if it's allowed).

NEWS AND LETTERS

LETTERS TO THE EDITOR

Dear Editor:

Readers of W.P. Cooke's "From Fences to Hyperboxes and Back Again" (December 1991) would probably prefer the following simpler proof:

By the arithmetic-geometric mean inequality:

$$C = \sum_{j=1}^{n} a_j x_j^{-1} \prod_{i=1}^{n} x_i \geq n\{\prod_{j=1}^{n} a_j x_j^{n-1}\}^{1/n},$$

so that

$$v_{max} = \{(C/n)^n \prod_{j=1}^{n} a_j^{-1}\}^{1/(n-1)}$$

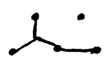
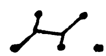
and with equality if and only if $a_j x_j^{-1}$ = constant for all j.

Also his Lagrange multiplier proof is incomplete in that sufficiency was not established, which usually seems to be the case in this type of proof.

M.S. Klamkin
University of Alberta
Canada

Dear Editor:

In the article "Steiner Minimal Trees on Chinese Checkerboards" (December 1991), Hwang and Du give a construction of heuristic Steiner Minimal Trees on triangular arrays T_n of lattice points. It might be interesting to note that for $n \equiv 0, 3 \pmod 4$ and $n > 3$, their construction (which they do not claim is optimal) can be improved on by replacing the configuration:

 by  ,

which replaces segments of length $1 + \sqrt{3} = 2.732...$ by segments of length $\sqrt{7} = 2.645...$

Anna Hughes
Beloit College, Class of 1993
Beloit, WI

Dear Editor:

Regarding my article "Assigning Driver's License Numbers" (February 1991), I have one correction and a few additions. Oren Patashnik and Rudy Beyl have kindly brought to my attention the fact that the state of Washington assigns the value 4 to an asterisk rather than 0 for the purpose of computing the check digit. The methods used by New York and Missouri is described in *UMAP Journal*, 13(1992) 37-42. Finally, the check digit for Wisconsin is determined by using the Washington scheme to assign the initial letter a numerical value and the South Dakota scheme is employed to calculate the check digit.

Only New Jersey remains to be determined.

Joseph A. Gallian
University of Minnesota
Duluth, MN

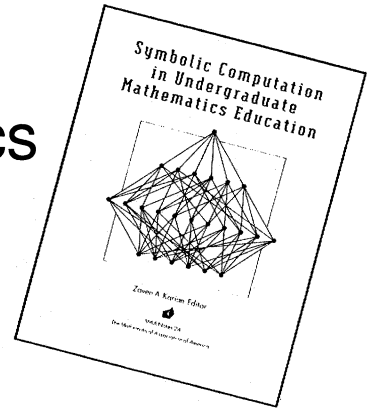
Dear Editor:

In the April 1992 issue of this *MAGAZINE*, Stephen Snover presents a nice Proof without Words showing that an alternating sum of squares is a triangular number. Readers may wish to compare Snover's contribution with a similar one that appeared in the December 1987 issue (p. 291) of this *MAGAZINE* by the late Dave Logothetti illustrating the same result.

Roger B. Nelsen
Lewis and Clark College
Portland, OR

Symbolic Computation in Undergraduate Mathematics Education

Zaven Karian, Editor



If you are interested in learning about how you can use the computer to help your students learn about important mathematical concepts this book needs to be on your shelf.

The availability of powerful symbolic computing systems on inexpensive micro computers is revolutionizing mathematics instruction in the nation's colleges and universities. This volume brings together many of the facets associated with the pedagogic uses of symbolic computation.

Part I consists of articles that deal with general issues of learning mathematics and the role of symbolic computation in that process. The articles in Part II describe the use of symbolic computation in teaching calculus. Some of the areas covered are the use of symbolic computation in a laboratory calculus course, the uses of Derive in the instruction of calculus, antidifferentiation and the definite integral, and the experiences and reflections of teachers who have used symbolic computation in calculus instruction.

Part III consists of papers on sophomore-level courses on linear algebra and differential equations. Some of the areas covered are the use of CAS in teaching linear algebra and calculus, the use of graphing calculators to enhance the teaching of linear algebra, the use of linear systems of differential equations using MAPLE, and the use of programmable graphics calculators in teaching a course on differential equations. The articles in Part IV describe what can be done in using symbolic computation in teaching combinatorics, probability and statistics courses. The articles and references in Part V will help you get started in using some of these ideas at your own institution.

200 pp., 1992, Paperbound
ISBN 0-88385-082-6

List: $22.00

Catalog Number NTE-24

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
(202) 387-5200 Fax (202) 265-2384

Membership Code ______

Name ____________

Address ____________

City ____________

State ___ Zip Code ____________

Qty.	Catalog Number	Price

Total $ ________

Payment ❑ Check ❑ VISA ❑ MASTERCARD

Credit Card No. ____________

Signature ____________ Exp. Date ________

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW
Washington, D.C. 20036