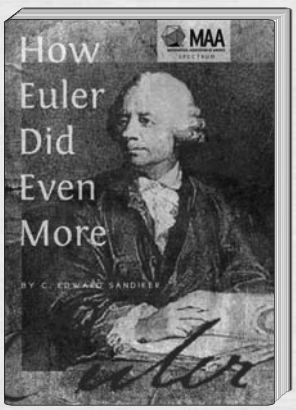
THE AMERICAN MATHEMATICAL
MONTHLY
MAA
VOLUME 122, NO. 1 JANUARY 2015

An Official Publication of the **Mathematical Association of America**

New from the MAA



How Euler Did Even More
by C. Edward Sandifer

"Read Euler, read Euler, he is master of us all," LaPlace exhorted us. And it is true, Euler writes with unerring grace and ease. He is exceptionally clear thinking and clear speaking. It is a joy and a pleasure to follow him. It is especially so with Ed Sandifer as your guide. Sandifer has been studying Euler for decades and is one of the world's leading experts on his work. This volume is the second collection of Sandifer's "How Euler Did It" columns. Each is a jewel of historical and mathematical exposition. The sum total of years of work and study of the most prolific mathematician of history, this volume will leave you marveling at Euler's clever inventiveness and Sandifer's wonderful ability to explicate and put it all in context.

Catalog Code: HEDM
ISBN: 978-0-88385-584-3
240 pp., Paperbound, 2014
List Price: $35.00
Member Price: $28.00
Series: Spectrum

To order, visit maa-store.hostedbywebstore.com or call 800-331-1622.

MAA100
MATHEMATICAL ASSOCIATION OF AMERICA
CELEBRATING A CENTURY OF ADVANCING MATHEMATICS

THE AMERICAN MATHEMATICAL
MONTHLY
MAA
VOLUME 122, NO. 1 JANUARY 2015

NOTICE TO AUTHORS

The MONTHLY publishes articles, as well as notes and other features, about mathematics and the profession. Its readers span a broad spectrum of mathematical interests, and include professional mathematicians as well as students of mathematics at all collegiate levels. Authors are invited to submit articles and notes that bring interesting mathematical ideas to a wide audience of MONTHLY readers.

The MONTHLY's readers expect a high standard of exposition; they expect articles to inform, stimulate, challenge, enlighten, and even entertain. MONTHLY articles are meant to be read, enjoyed, and discussed, rather than just archived. Articles may be expositions of old or new results, historical or biographical essays, speculations or definitive treatments, broad developments, or explorations of a single application. Novelty and generality are far less important than clarity of exposition and broad appeal. Appropriate figures, diagrams, and photographs are encouraged.

Notes are short, sharply focused, and possibly informal. They are often gems that provide a new proof of an old theorem, a novel presentation of a familiar theme, or a lively discussion of a single issue.

Submission of articles, notes, and filler pieces is required via the MONTHLY's Editorial Manager System. Initial submissions in pdf or LaTeX form can be sent to the Editor Scott Chapman at

www.editorialmanager.com/monthly

The Editorial Manager System will cue the author for all required information concerning the paper. The MONTHLY has instituted a double blind refereeing policy. Manuscripts which contain the author's names will be returned. Questions concerning submission of papers can be addressed to the Editor at monthly@shsu.edu. Authors who use LATEX can find our article/note template at **www.maa.org/monthly.html**. This template requires the style file maa-monthly.sty, which can also be downloaded from the same webpage. A formatting document for MONTHLY references can be found there too.

Letters to the Editor on any topic are invited. Comments, criticisms, and suggestions for making the MONTHLY more lively, entertaining, and informative can be forwarded to the Editor at **monthly@shsu.edu**.

The online MONTHLY archive at **www.jstor.org** is a valuable resource for both authors and readers; it may be searched online in a variety of ways for any specified keyword(s). MAA members whose institutions do not provide JSTOR access may obtain individual access for a modest annual fee; call 800-331-1622 for more information.

See the MONTHLY section of MAA Online for current information such as contents of issues and descriptive summaries of forthcoming articles:

www.maa.org/monthly.html

Proposed problems or solutions should be sent to:

DOUG HENSLEY, MONTHLY Problems
Department of Mathematics
Texas A&M University
3368 TAMU
College Station, TX 77843-3368

In lieu of duplicate hardcopy, authors may submit pdfs to **monthlyproblems@math.tamu.edu.**

Advertising correspondence should be sent to:

MAA Advertising
1529 Eighteenth St. NW
Washington DC 20036
Phone: (202) 319-8461
E-mail: **advertising@maa.org**

Further advertising information can be found online at **www.maa.org.**

Change of address, missing issue inquiries, and other subscription correspondence can be sent to:

maaservice@maa.org

or

The MAA Customer Service Center
P.O. Box 91112
Washington, DC 20090-1112
(800) 331-1622
(301) 617-7800

Recent copies of the MONTHLY are available for purchase through the MAA Service Center at the address above.

Microfilm Editions are available at: University Microfilms International, Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, NW, Washington, DC 20036 and Lancaster, PA, and copyrighted by the Mathematical Association of America (Incorporated), 2015, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. Permission to make copies of individual articles, in paper or electronic form, including posting on personal and class web pages, for educational and scientific use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear the following copyright notice: [Copyright 2015 Mathematical Association of America. All rights reserved.] Abstracting, with credit, is permitted. To copy otherwise, or to republish, requires specific permission of the MAA's Director of Publications and possibly a fee. Periodicals postage paid at Washington, DC, and additional mailing offices. **Postmaster:** Send address changes to the American Mathematical Monthly, Membership/Subscription Department, MAA, 1529 Eighteenth Street, NW, Washington, DC 20036-1385.

A Letter from the Editor: 2015 is the Centennial Year of the MAA

Scott T. Chapman

Happy New Year and welcome to 2015, the centennial year of the Mathematical Association of American. The *Monthly* has the unique distinction of being older than the MAA. Founded in January of 1894, the *Monthly* predated the MAA by more than 20 years. In fact, the entire founding of the MAA was predicated on the future of the *Monthly*. Readers can get a close up glimpse of this in our lead article this month "The First 100 Years of the MAA," by David Zitarelli. During the year, there will be various events to mark the centennial—watch for a Special Session at the 2015 MathFest in Washington, D.C. tentatively titled, "Generations of *Monthly* Gems," where selected speakers will highlight some of the great papers of the *Monthly's* past.

The New Year will also bring some change to the *Monthly*. In August of 2012, the MAA Board of Governors passed a resolution mandating a switch for all MAA journals from traditional refereeing of manuscripts to double blind refereeing (i.e., submitted manuscripts will not identify the names of the author(s)). At the time of the resolution, this practice was already in use by *The College Mathematics Journal*. The resolution goes into effect for the *Monthly* when I cease to handle new submissions (January 2016). While under no pretense to do so, I have decided to usher in the new refereeing system this year. Hence, as of January 1, 2015, all refereeing at the *Monthly* will be double blind. I hope, as the Editor-Elect of the *Monthly* takes office on January 1, 2016, that this lessens the burden on his or her transition to a full-time Editorship. Hopefully, by this time next year we will have some statistical data that reflects the impact of this new system on our operations.

In July of last year, Pearson Publishing, our long time typesetters, ceased operation. We thank Don DeLand, Anne Holmes, and Leslie Galen of Pearson for their countless hours of devoted service. I am sure that the *Monthly* would not be the journal it is today without their years of hard work. Shortly before Pearson closed, the MAA came to an agreement with Cenveo Corporation to begin typesetting all its journals. The October 2014 issue of the *Monthly* was the first typeset by Cenveo. As we have maintained our prior format, readers should notice little difference. I would like to particularly thank Bonnie Ponce, Assistant Managing Editor for MAA Publications, for her help in this transition, which forced us to change several of our long time editorial practices.

In closing, I note that the New Year has triggered two significant retirements. The first is Ivars Peterson, who has guided the helm of all MAA journals as Director of Publications since 2007. Ivars has led the MAA journals during a critical period where the boundaries of academic publishing seemed to be changing almost daily. His guidance and leadership will sorely be missed. The second is Roger Horn, who is stepping down from the Editorial Board of the *Monthly*. Roger served as Editor of the *Monthly* from 1997–2001 and has served the past 13 years as an Associate Editor. Roger's contributions to the *Monthly* cannot be overstated. If there were a *Monthly* Hall of Fame, then Roger would be in the inaugural class. Thanks Ivars and Roger for your many years of dedicated service!

http://dx.doi.org/10.4169/amer.math.monthly.122.01.3

The First 100 Years of the MAA

David E. Zitarelli

Abstract. Why was the MAA founded? What role has the Association played in American mathematics? What were its primary activities? We answer these questions in this overview of the MAA over its 100-year history from its founding in 1915. Along the way, we describe MAA sections, governance, meetings, prizes/awards, and headquarters. The account of MAA activities is divided into two periods, 1916–1955 and 1955–2014 and contains a discussion for the critical role played by the Committee on the Undergraduate Program in Mathematics in this division.

1. INTRODUCTION. This article presents an overview of the history of the Mathematical Association of America as part of the celebration of its centennial in 2015. It describes events this author regards as the most important over the century, but the account is certainly not exhaustive; for example, it makes little mention of competitions conducted under the aegis of the Association or of the expanded book publication program. Our account begins with the founding of the MAA and then describes its sections, governance, and meetings. Overarching activities are outlined in two distinct periods, 1916–1955 and 1955–2014, and I supply an explanation for the partition into disjoint stages. The article then discusses prizes and awards before ending with a brief mention of MAA headquarters.

2. FOUNDING. One of the most historic moments for mathematics in America occurred with the establishment of a national organization on the last two days of 1915. It is rather miraculous that the Mathematical Association of America (**MAA**) was founded amidst World War I, a year after Canada entered the fray as a Dominion of the British Empire and 16 months before the U.S. Congress declared war. It is important to note that the use of "America" in the title of the Association includes both Canada and the United States as Albert Bennett wrote upon the 50th anniversary of the MAA in 1965, "The phrase 'of America' was interpreted from the start to include Canada and indeed the North American continent" [**3**, p. 1]. Since that time, members living in the Caribbean areas belong to the Florida Section of the MAA.

The founding of the MAA took the reverse of the usual route whereby an organization is established first and creates its official journal later. For instance, the American Statistical Association was formed in 1839 but did not establish a publication for another 49 years. The American Mathematical Society (**AMS**) was quicker, being founded in 1888 but creating the *Bulletin* as its first periodical three years later. Yet the MAA's official journal, the *American Mathematical Monthly*, was initiated more than 20 years before the Association was founded. So in this case, a journal (the *Monthly*) spawned an association (the MAA).

After the MAA assumed the reins of the *Monthly*, its masthead read, "Founded in 1894 by Benjamin F. Finkel, published by him until 1913. From 1913 to 1916 it was owned and published by representatives of fourteen Universities and Colleges in the Middle West." Benjamin Franklin Finkel was a graduate of Ohio Northern University who had taught in secondary schools before being appointed professor of mathematics

http://dx.doi.org/10.4169/amer.math.monthly.122.01.4
MSC: Primary 01A60, Secondary 01A61; 01A74

and physics at Drury College in Missouri. Finkel pursued graduate studies during summers at the University of Chicago before earning his Ph.D. at the University of Pennsylvania.

Finkel's earlier classroom experience had made him keenly aware of poor instruction in elementary mathematics, and that inspired him to establish a journal "devoted solely to mathematics and suitable to the needs of teachers of mathematics in these schools" [**8**, p. 309]. He consulted with numerous high-school teachers and college professors, but few teachers responded favorably, whereas he received enthusiastic support from several notable professors. The first issue of the *Monthly* had appeared in January 1894 with the title *The American Mathematical Monthly: A Journal for Teachers of Mathematics in the Collegiate and Advanced Secondary Fields*. The task of physically producing a journal was not easy, especially mathematical typesetting, and it became a family affair, with Finkel carving most of the woodcuts himself while his wife Hannah Cokeley Finkel served as proofreader.

Finkel met Leonard Dickson while studying at the University of Chicago in the summer of 1895. Seven years later, Finkel invited Dickson to become an editor of the *Monthly* and called his acceptance "a red-letter day in the history of the *Monthly*" [**8**, p. 314]. The cover of the January 1903 issue reads, "Edited by B. F. Finkel and Leonard E. Dickson." By then, it was apparent that college and university instructors evinced a much greater interest in the *Monthly* than its intended audience, especially the problems department, which was central at the outset and has been a mainstay to this day.

A "second red-letter day in the history of the *Monthly*" [**8**, p. 314] occurred in 1907 when Herbert Ellsworth Slaught replaced Leonard Dickson as editor along with Finkel. Slaught was a graduate of Colgate College who taught at the Peddie School near Princeton before matriculating at the University of Chicago, where he obtained his Ph.D. in 1898 under E. H. Moore, four years after joining the mathematics faculty. Slaught was in charge of submissions during 1907 and 1908, an experience that heightened his awareness of the need for financial support for the *Monthly* beyond subsidies provided by Drury College and the University of Chicago. He soon obtained a matching subsidy from the University of Illinois with the help of a third editor, G. A. Miller. But in 1912, Benjamin Finkel confided to Slaught that the printer could no longer afford the low-cost services he had been providing. Undaunted, by the end of the year, Slaught, along with Florian Cajori and Earle Hedrick, arranged for 11 Midwestern universities and colleges to help defray the costs and to pass the *Monthly* legally from the private possession of Benjamin Finkel to the board of editors.

Moreover, during 1913–1915 Slaught gained the conviction that more had to be done for the average mathematics teacher within the field of collegiate mathematics. Accordingly, he conducted an informal discussion at the dinner of the April 1914 meeting of the Chicago Section of the AMS about the role of collegiate mathematics in America. It was felt that, on one end of the spectrum, secondary mathematics was being handled well by existing secondary associations throughout the country, while at the other end, research interests were being fortified by the AMS. Yet, as Slaught wrote, "the great intermediate field of collegiate mathematics … so far has had no organized attention" [**18**, p. 251]. Karen Parshall describes this division of labor as "the stratification of the American mathematical community" [**16**].

In late 1914, the resolute Slaught appealed to the Council of the AMS to appoint a committee to "consider the general relation of the Society to the promotion of teaching, especially in the collegiate field" [**12**, p. 20]. Over the next year, two critical elements nudged closer together: the *Monthly* and a movement to emphasize collegiate mathematics. The impetus was the search for a source of dedicated support for the *Monthly*,

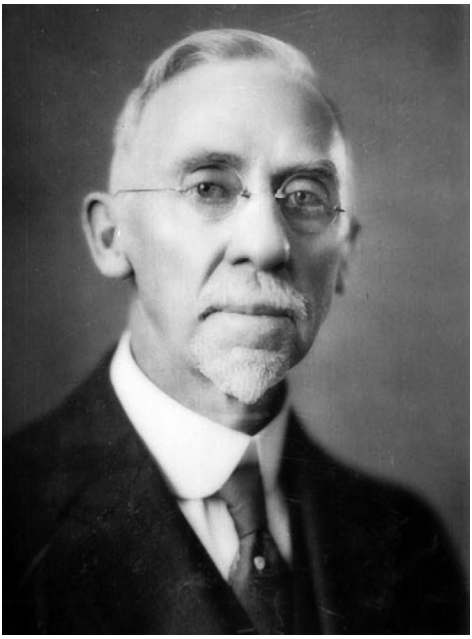


Figure 1. MAA president 1919, Herbert E. Slaught.

but Slaught had conceived a grander idea of founding a national body as well. Thereupon, the AMS council appointed the requested committee, which voted three-to-two that the AMS should take neither control nor responsibility for publishing the *Monthly* but resolved [**2**, p. 79; **12**, p. 20]:

> It is deemed unwise for the American Mathematical Society to enter into the activities of the special field now covered by the *American Mathematical Monthly*; but the Council desires to express its realization of the importance of the work in this field and its value to mathematical science, and to say that should an organization be formed to deal specifically with this work, **the Society would entertain toward such an organization only feelings of hearty good will and encouragement**. [Emphasis added.]

This resolution was endorsed by an overwhelming majority of the council, which then adopted it.

Events moved quickly after that decision. Herbert Slaught and a loyal band of supporters gave wide publicity to the idea of forming the kind of body the AMS resolved to encourage with good will. In June 1915, he mailed letters to mathematicians throughout the United States and Canada soliciting feedback on this idea, enclosing a reply post card. It met with unbridled support, but also sprinkled opposition. Finally, Slaught circulated a form letter seeking to identify those who favored such an organization, resulting in a call to an organizational meeting signed by 450 mathematicians representing every state in the United States and province in Canada.

Where and when should the organizational meeting be held? The obvious choice, it would seem, would be at the annual AMS meeting that December 27–28. However, the AMS met at Columbia University in New York City, whereas the majority of support for the new body came from the Midwest. Instead, the organizational meeting was held at Ohio State University on December 30–31, 1915 in conjunction with the annual meetings of Section A of the American Association for the Advancement of Science (**AAAS**) and the Chicago Section of the AMS. Ironically, the AAAS meeting had been

scheduled for Toronto by organizer J. C. Fields, but the outbreak of WWI cancelled those plans and the AAAS moved to Columbus instead. Otherwise, the MAA might have held its organizational meeting in Canada.

Two of David Hilbert's American students played decisive roles in the meeting—Earle Hedrick (who presided) and Will Cairns (who served as secretary *pro tem*). Overall, 104 delegates attended, including ten women. Cairns reported, "After three hours of patient and painstaking deliberation, all mooted questions were settled except the name of the new organization" [**4**, p. 3]. A committee of three was appointed to select a name from among the 18 that were submitted. The next day, they voted independently, only to discover a "remarkable unanimity of purpose" [**1**, p. 29], as all three favored the Mathematical Association of America. Therefore, the centennial of the MAA will occur on December 31, 2015.

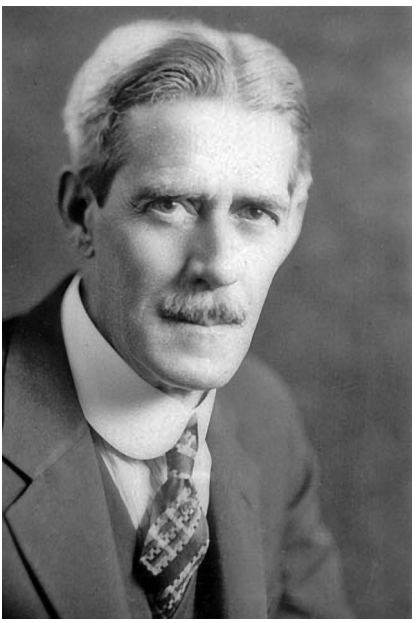


Figure 2. MAA president 1916, Earle R. Hedrick.

The response to the formation of the new organization was overwhelming. Charter membership was closed on April 1, 1916, at 1045. (While that number might seem small today, it represented 1/100,000 of the U.S. population at the time.) By contrast, AMS membership that year was 732. Because the first names of women were recorded, I counted 132 (12.6%) females; women accounted for 29 of the 131 with Ph.D. degrees (22.1%). It is harder to distinguish African Americans; the only charter member I know was Dudley Weldon Woodard. With its Midwestern origins, one might assume that the majority of charter members would hail from there, yet the four leading states were New York (120), Ohio (72), Massachusetts (70), and Pennsylvania (67). A mere 15 came from Canada. It was telling that only 145 of the 1045 were high-school teachers, confirming once again the wisdom of Benjamin Finkel's emphasis on collegiate-level mathematics.

Founding the Association obviated the need for university subsidies that had kept the *Monthly* afloat the previous three years. Thus, Slaught's idea to form an organization, with members' dues paying for the journal, was brilliant. *Herbert Slaught was*

truly the father of the MAA. To reinforce this designation, the title page for the January 1938 issue of the *Monthly* asserted:

> In appreciation of his exceptional services to the Mathematical Association of America and to this *Monthly*, this volume is dedicated to the late Herbert Ellsworth Slaught, who at the time of his death was Honorary President of the Association and had served continuously as an editor of the *Monthly* for thirty years.

The second day of the organizational meeting continued with passage of a constitution that had been tentatively prepared beforehand. In addition, a nominating committee presented a slate of officers: Earle Hedrick, president; E. V. Huntington and G. A. Miller, vice-presidents; and W. D. Cairns, secretary. The MAA's indebtedness to Germany was immediate—Hedrick and Cairns had obtained doctorates at Göttingen, Huntington at Strassburg, and Miller had studied with Sophus Lie at Leipzig after receiving a Ph.D. from Cumberland University (Tennessee).

An appointed 12-person executive council selected Alexander Ziwet and Karl Swartzel to negotiate with the owners of the *Monthly* to secure it as the official journal of the Association. As a result, the *Monthly* was formally transferred to the MAA in time for the appearance of the January 1916 issue. This action underlines the fact that it was the journal that had spawned the Association! The organizational meeting ended with an illustrated address titled "The story of algebra" by L. C. Karpinski.

In 1965, Brown University emeritus professor A. A. Bennett lyrically described the work of those mathematicians involved in establishing the MAA [**3**, p. 1]:

> Our Association was founded under especially auspicious circumstances. The many favorable factors were not accidents, nor miracles, nor achieved through serendipity. Some were the end results of a chain of events, not always desired, not always with the eventual outcome in view. But in large part they were secured through wise planning, tactful compromise, cajoling of the apathetic, courageous facing of pessimists in high places, and unremitting work.

The MAA developed in several directions at once, perhaps the most influential being the establishment of MAA sections.

3. SECTIONS. In his November 1915 *Monthly* article "The teaching of mathematics," Herbert Slaught stated that the object of the journal was "to *stimulate activity* on the part of college teachers … that may lead to *production*" [**19**, p. 291]. But he quickly added that "the formation of smaller groups … will provide a far-reaching stimulus to individual activity" and he hoped "to see the college teachers of mathematics organized in every state, or even in some smaller groups."

Those smaller groups had a model to follow: the AMS. Its first section (Chicago) was established in 1896, eight years after its founding. Two sections followed: San Francisco (1902) and Southwestern (1906). How long did it take the MAA to form sections? Minutes! No sooner had the constitution been approved than three states submitted formal applications to become sections—Kansas, Missouri, and Ohio.

Which of the three state organizations became the first MAA section? Organizationally savvy, the Ohio group created a special committee at the conclusion of the first day of the two-day meeting to prepare its own constitution. Consequently, Ohio beat Missouri by a few minutes in the heated race; Kansas placed third. Therefore, the Council of the MAA (the Board of Governors since 1938) acted expediently and granted Ohio's application as the first section on March 1, 1916. But the honor of holding the first official meeting reverted to Kansas, which met in March. The Ohio Section held its initial meeting a month later for two days, a tradition that has contin-

ued every spring since then (except the war year 1945). The Missouri Section waited until November to hold its initial gathering.

Four more MAA sections were formed during 1916, three from Midwestern states (Iowa, Indiana, and Minnesota) and the first one with a nonstate geographical boundary (Maryland–Virginia–District of Columbia). Ten years later, Slaught singled out the 17th section—Philadelphia—as the first to bear the name of a city instead of a state or union of regions because he had feared "the seeming apathy or lethargy" in the Atlantic states [**25**, p. 7]. Unlike Midwestern mathematicians, those in New England felt adequately represented by the Association of Teachers of Mathematics in New England until 1955, when the Northeastern Section was formed to also include the Canadian provinces of New Brunswick, Newfoundland, Nova Scotia, and Prince Edward Island.

There were 28 sections when the MAA celebrated its semicentennial in 1965. Only one has been added since that time, the Intermountain Section, founded in 1975 when the Rocky Mountain Section was partitioned.

4. OFFICERS AND GOVERNANCE. Since its founding, the MAA has had a core of four officers: president, vice president, secretary, and treasurer. Its online site lists all presidents with photographs and short biographies:

http://www.maa.org/about-maa/governance/maa-presidents

During the first ten years, the presidency was an honorary office held for one year. However, the term limit was increased to the present two years starting in 1927. Dorothy Bernstein was the first woman president, elected for 1979–1980. Due to increasing membership, today's officers include, in addition to the core: president-elect, first vice president, second vice president, associate secretary, and associate treasurer.

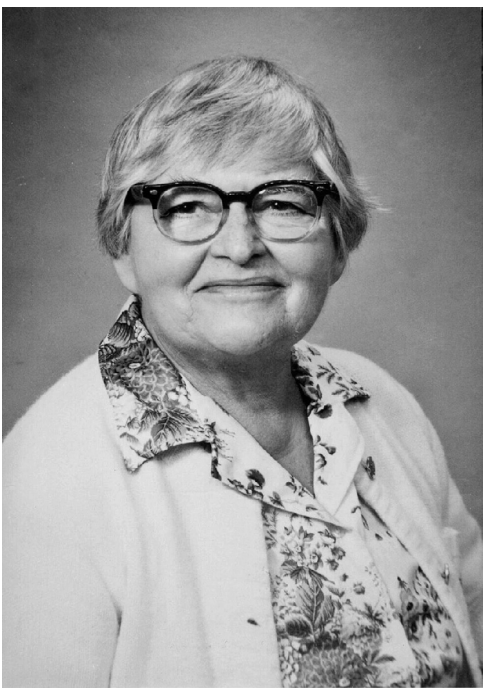


Figure 3. MAA president 1979–1980, Dorothy Bernstein.

By 1938, many MAA leaders had become concerned about the tenuous relationship between the Association and its sections, so the trustees appointed the Committee to Review the Activities of the Association. This committee reported its findings in the *Monthly* two years later, whereupon the trustees accepted the report and discharged the committee with appreciation. Its five recommendations provided the most extensive analysis of the structure and operations of the MAA since its founding in 1915. One of those recommendations replaced the trustees with the now familiar Board of Governors to administer and control all MAA scholarly and scientific activities. To exercise its fiduciary responsibility, the Board of Governors votes on dues and on reports from the treasurer, Budget Committee, and Audit Committee. The Board also approves the editors and editorial boards of the *American Mathematical Monthly*, *Mathematics Magazine*, *The College Mathematics Journal*, and *MAA Focus*.

The board was initially composed of six national officers, six governors elected at large, and 14 governors elected by region. The term for governors at large was three years, so two were elected annually. However, the designation of regional governors soon proved to be unnecessarily cumbersome, and so in 1945, the board began phasing them out in favor of sectional governors serving three-year terms. Since that time, the Executive Committee was formed to act on behalf of the governors on matters that arise between board meetings. This committee has also made recommendations on the management, policies, and activities of the MAA.

5. MEETINGS. There are three kinds of mathematicians—those who can count and those who can't. The assignment of numbers to initial MAA annual meetings seems to bear out this joke.

The December 1915 organizational meeting was regarded as the first annual meeting because the next one was called second in the *Monthly* report. The 1917 annual meeting was called the third. Yet the 1918 annual meeting was also referred to as the third. Subsequent meetings were numbered after this one. The 1942 meeting was cancelled at the request of the Office of Defense Transport, the only time in MAA history that this annual affair did not take place, so the 40th occurred in December 1956. However, after that meeting the board voted "to hold the Annual Meeting normally scheduled for December 1958 during the latter part of January 1959" [**9**, p. 213]. Up to that point, most annual meetings had been scheduled between Christmas and New Year's Day (except three held in November during WWII), but they have all been held in January since the 41st in 1958. This explains why the 98th annual meeting will take place in January 2015.

The philosophy of including Canada in the title of the MAA was reinforced at the sixth annual meeting held in Toronto in 1921 in conjunction with the AMS, AAAS, and the APS (American Physical Society). This suggests that the titles of all four organizations implicitly included Canada.

The MAA held its first summer meeting September 1–2, 1916, at MIT, sandwiched between the organizational and second annual meetings. ("Annual meeting" refers to the December/January gathering, as opposed to "summer meeting," though summer meetings are held annually as well.) Attendees wore badges with identifying names and institutions for the first time, a custom followed to this day. The nine invited lectures were delivered in MIT's spanking new buildings while Harvard Yard dormitories provided accommodations. At the welcoming dinner on Friday night (September 1), a cablegram with fraternal greetings was received from Mittag-Leffler in Sweden. Except for three years during WWII, summer meetings were held after the U.S. Labor Day until 1952, when they were switched to the week before the national holiday. Twenty-five years later, these gatherings were scheduled for earlier in August because

many universities and colleges began their fall semesters the week before the three-day holiday. Moreover, a mandatory group photograph was shot through 1948. The largest meeting of the MAA (through 2014) was held in January 2012 in Boston with 7199 registrants.

The attendance of 750 persons in 1958 encouraged both the MAA and AMS to meet later in January from that time forward. Moreover, the MAA began to enlarge its program dramatically. Longtime MAA secretary H. M. Gehman reported, "The reasons for the expanded programs of meetings is due to the increase in the membership of the Association, the broadening of interests of the membership, and the increase in the activities in which the M.A.A. is engaged" [**10**, p. 579].

A historic change for MAA national meetings took place in 1996 when the AMS voted to disband its summer gatherings. The MAA decided to continue alone, adopting the name "MathFest" starting in 1997, and has sponsored this meeting every summer since then.

The MAA has sponsored two meetings a year up to the present time with only a few cancellations. We mentioned the only annual meeting cancelled (in 1942), but seven summer meetings were not held. Four of these were cancelled whenever an International Congress of Mathematicians was held in North America (1924 in Toronto, 1950 in Cambridge, 1974 in Vancouver, and 1986 in Berkeley) while the other three were in 1938 for the AMS semicentennial celebration, 1945 for WWII, and 1992 for the Canadian Mathematical Society.

Today, most sessions and lodgings of annual meetings are held in hotels, often in conjunction with a convention center, and not on a college campus. The first time the MAA did not hold a winter meeting on a college campus was 1929, when the site was the Hotel Fort Des Moines, but that did not happen again until Chicago in 1960. Only a few annual winter meetings have met at universities since then.

There have been at least two controversial locations of annual meetings. Negotiations for sites take place years beforehand, and those for the January 1970 meeting began five years earlier. Yet in late October 1969, just ten weeks before the affair was scheduled to begin, officials in Miami (Florida) informed meeting organizers that they could not supply the number of rooms agreed upon earlier. Fortunately, Henry L. Alder (MAA president 1977–1978) informed organizers about the spanking new convention center in San Antonio, whereupon the location was hastily changed to that Texas city (and its renowned River Walk), which has now hosted more annual meetings than any other site.

Another controversial annual meeting was held in Las Vegas in 1972. The combination of smoke-filled rooms and an "ambiance of gambling [that some found] intrusive and oppressive (and a handful found too tempting)" caused MAA and AMS leaders to resolve that "the Society not meet again in Las Vegas in this century" [**17**, p. 36]. Nonetheless, the AMS Western Section has held meetings there, and one is scheduled for 2015.

The locations of some section meetings were controversial because of discrimination against African Americans, particularly in the Southeastern Section of the MAA. As far as is known, no African American attended a meeting in this five-state section up to 1951, when Fisk University chair Lee Lorch and three African American colleagues attended the gathering in Nashville. However, they were prohibited from registering for the banquet to hear a talk by MAA president Saunders Mac Lane. Shortly thereafter, the Board of Governors responded to a letter from Fisk University faculty, not by changing its bylaws, as the group had requested, but by affirming its intention to conduct the affairs of the MAA without discrimination. Nonetheless, no African American mathematicians attended Southeastern meetings again until 1960.

However, when a group from Atlanta University was refused rooms at the meeting site, the Wade Hampton Hotel in Columbia (South Carolina), they left in protest. It was not until the late 1960s and early 1970s that African Americans felt comfortable enough to attend Southeastern meetings on a regular basis. Since that time, minority mathematicians have participated in sectional governance and have presented invited addresses. In 1995 the report, "A history of minority participation in the Southeastern Section," concluded [**7**, p. 10]:

> The Southeastern Section has been the birthplace of protest against discrimination on the basis of race. Confrontations here have helped to move the MAA forward in the elimination of barriers in the mathematics community, not only for African-Americans, but for women, gays and lesbians, and other minorities. Today the Southeastern Section stands as a leader and a model in its determination to encourage and promote the full participation of all of its members.

6. ACTIVITIES 1916–1955. Once the MAA had been established in 1915, its charter members launched several projects, forming six committees in 1916 alone. The Committee on Relations with the *Annals of Mathematics*, chaired by E. H. Moore, entered into an agreement with Princeton, the journal's publisher, whereby the MAA provided an annual subvention in exchange for the publication of expository articles and subscription rates at half price for MAA members. The MAA entered into a similar agreement with the *Duke Journal* when it was established in 1935, but both subventions were discontinued at the end of 1942 due to lack of resources.

The most significant initial project was carried out by the National Committee on Mathematics Requirements (**NCMR**), formed in March 1916 "to investigate the whole field of mathematical education from the secondary school through the college and to make recommendations looking toward a desirable reorganization of courses and the improvement of teaching" [**23**, p. 5]. NCMR issued a preliminary report at the MAA's first summer meeting that September and a final report seven years later—a detailed, 650-page document titled *The Reorganization of Mathematics in Secondary Education*.

The NCMR report was influential in establishing the basic contour of the high-school curriculum for several decades, but jousting over the aims in teaching mathematics continues to this day. However, no major change in practice occurred through the 1930s due mainly to the Great Depression and the inertia of maintaining traditional educational practices. One important byproduct of this work was the founding of another professional organization of mathematicians in 1920, the National Council of Teachers of Mathematics (**NCTM**). Like the MAA, the NCTM adopted an existing publication as its official journal, *The Mathematics Teacher*.

The impact of the NCMR was generally positive, but not the work of the Committee on Standard Departments, charged with formulating standards for undergraduate departments. As we know from NCTM documents formulated in the 1990s, the term "standards" can be controversial, and it was no different back then, as the committee made little progress and was allowed to lapse within two years.

In contrast with this disappointment, the establishment of mathematics clubs at the undergraduate level was a roaring success. The *Monthly* initiated a column "Undergraduate Mathematics Clubs" in 1918 that continued under different names until 1954.

The printing costs problem raised its ugly head again in 1920, when the MAA was forced to increase dues by a dollar from the initial \$3 per year. However, the resulting \$4 dues remained in place until 1958, causing Lester R. Ford (MAA president 1947–1948) to joke that the three most important constants in mathematics were π, e, and 4 [**11**, p. 105].

To address rising costs, the MAA was incorporated in 1920 under the statutes of the state of Illinois, an important move that allowed the Association to receive donations and bequests; the AMS followed suit three years later. The MAA took advantage of this status right away when Herbert Slaught announced a sizable gift from Mary Carus, the editor of Open Court Publishing Co. The purpose of her gift was to fund a series of books to publish mathematical exposition at nominal cost. The first Carus Mathematical Monograph was *Calculus of Variations* by G. A. Bliss (1925) and the second *Functions of a Complex Variable* by D. R. Curtiss (1926). By 2014, 31 books had appeared in this popular series.

In his retiring presidential address published in 1932, John Wesley Young called for the MAA to (1) provide funds to send national leaders to speak at sectional meetings in remote locales, (2) launch a second periodical with particular appeal to undergraduate students, (3) initiate a Committee on Publicity, and (4) form a competitive national examination like the one given in Hungary [**24**]. All four proposed initiatives have now become part and parcel of MAA activities, but none was instituted at once.

The MAA's second journal grew out of the effort of S. T. Sanders of Louisiana State University (LSU) to encourage high school teachers to join the Association. In 1926, he began publishing a monthly pamphlet and the next year expanded it into a magazine. However, financial support from LSU was terminated in 1942 due to university budget constraints. Deficits mounted alarmingly! The MAA provided subsidies, but even that dried up in 1945, whereupon the journal abruptly ceased publication. Fortunately, in 1947, Glenn James assumed sponsorship, resumed publication, and shortened the title to *Mathematics Magazine* from the name *National Mathematics Magazine* that had been used since 1934. But by 1959, deteriorating eyesight caused him to negotiate, and the December 1960 issue revealed the complete transfer to the MAA. In 1974, the MAA initiated its third periodical, the *College Mathematics Journal*, which had been published by Prentice-Hall as the *Two-Year College Mathematics Journal* the previous four years.

During the interwar years, the MAA sponsored several studies of undergraduate mathematics courses, especially at the freshman and sophomore levels. At the 1922 summer meeting, J. W. Young proposed a general one-year course for freshmen that would meet the needs of those pursuing upper-level mathematics as well as those not continuing. Like today, there was little agreement on the contents of such an integrated course. Five years later, an MAA committee headed by A. A. Bennett supported traditional instruction during the first two years but suggested that a certain amount of historical and philosophical background could provide additional mathematical concepts. The committee also drew up a suggested list of readings, a precursor to today's "Basic Library List."

The Great Depression exacted a distressing toll on the profession, but the MAA addressed the employment problem directly. In 1932, the board discussed the training of Ph.D. students for teaching at junior colleges or high schools due to the scarcity of available professorships. The next year, those trustees asked president Arnold Dresden to appoint the Commission on the Training and Utilization of Advanced Students in Mathematics. Under the leadership of Northwestern University chair E. J. Moulton, the commission issued a report describing the unemployment situation facing the 180 mathematics Ph.D.s seeking positions for 1934–1935. Only 14 were unemployed, but many others held "makeshift" positions [**15**, p. 143]. Yet the commission forecasted correctly that demand would soon exceed the supply of Ph.D.s, so no further action was taken and the situation improved markedly in later years.

The other big activity taken in response to external demands of the profession took place when the MAA acted swiftly after war broke out in Europe in September 1939.

Within a month, the board of trustees authorized MAA president W. B. Carver (the chair at Cornell) to consult with AMS president G. C. Evans (the chair at UC Berkeley) about appropriate measures regarding national defense. The result was the joint War Preparedness Committee that presented its report at the summer 1940 meeting held at Dartmouth College [**14**]. The report listed three recommendations: (1) competent secondary students should take the maximum amount of mathematics courses available; (2) colleges should offer courses in mechanics, probability, surveying, navigation, and other essentials of military science; and (3) universities should offer courses in applied mathematics at the graduate level. This committee also compiled and maintained a register of vacancies and availability of mathematicians for service throughout the war.

7. ACTIVITIES 1955–2014. What is the distinctive role of the MAA? The founders viewed the pedagogy of collegiate mathematics as its distinguishing characteristic, yet over its first 40 years, the Association mainly held meetings, elected officers, selected hour speakers, and published a journal, all of which duplicated AMS actions. The critical event that separated these two national organizations occurred during the presidencies of Edward J. McShane and William L. Duren, Jr., 1953–1956. These two leaders had been friends as undergraduates at Tulane and graduate students at Chicago, where both earned Ph.D.s in 1930. Duren returned to Tulane after graduating from Chicago and taught there until 1955. Duren then joined McShane at Virginia, where McShane had been since 1935.

Figure 4. MAA president 1953–1954, Edward J. McShane.

Both McShane and Duren had established solid reputations in research by 1955, which was typical of most MAA leaders. But their experiences in teaching propelled them into roles that altered the character of the MAA. Before assuming the MAA presidency in January 1953, McShane had settled upon the idea that the cornerstone of his

term would be to dramatically improve undergraduate instruction in mathematics. His first action was to establish the Committee on the Undergraduate Mathematical Program (**CUP**), where by "program" he meant not only the curriculum but also faculty and students as well. He appointed Duren to chair CUP.

The idea of organizing a national committee to deal with the undergraduate program seems commonplace today, but in retrospect, this was an unusual move for the MAA and would be the first in a series of broad strokes that would distinguish it from the AMS. This was not the first MAA committee to deal with curricular issues, as noted above. However, even the massive NCMR report from 1923 dealt only with secondary education.

Duren was the perfect choice to head CUP. As chair at Tulane, he had established a graduate program in 1947 with grants from the Office of Naval Research administered by Mina Rees. But the program had a low success rate, which Duren attributed to poor undergraduate training. His idea was to replace the traditional freshman college algebra course by calculus or by a general analysis course including calculus. CUP acted with dispatch and filed a report to the board at the 1953 summer meeting. The major recommendations that the governors adopted at once were to (1) concentrate on first-year courses for able students, (2) articulate with physicists and engineers and cultivate contacts in the social sciences, (3) avoid overemphasizing abstract topics, and (4) enlist college instructors in the process.

Figure 5. MAA president 1955–1956, William L. Duren, Jr.

Two future MAA presidents were instrumental in these activities: Albert Tucker (president 1961–1962) and G. Baley Price (president 1957–1958). Duren had met Tucker during a sabbatical year at the Institute for Advanced Study in the mid-1930s and credited him with getting the idea of replacing the college algebra course with calculus accepted: "Tucker gave it the prestige of Princeton, which was essential to its

general acceptance" [**6**, p. 3]. Baley Price, as chair at Kansas, obtained a small grant and devoted space from his university to support CUP members to meet during the next summer to begin to carry out details. Up to this time, it had been difficult for a national group to operate in this manner.

As a result, in August 1954, CUP issued a revolutionary report urging the adoption of one universal freshman course. What set CUP apart from its predecessors was that it not only issued a set of recommendations for the course, but it also produced source material for what it called the Universal Course. This material was tested at Tulane and several other institutions the next year.

CUP reached a crossroads by 1958 when it realized that although its work had been restricted to mathematics in the freshman year, only the course designed for students in the social sciences (mostly under the direction of John Kemeny at Dartmouth) had found success. But the calculus part was generally unsuccessful, and the overall program for mathematics majors remained unchanged from a half century earlier. CUP members felt strongly that the time was ripe for an expanded group to investigate the matter, resulting in a conference that November to examine its work over the prior five years and to assist in formulating plans and policies for its future work.

Two of the reports from that conference described progress that had been made at the precollege level on what came to be called the New Math. One report was by Max Beberman, the head of the University of Illinois Committee on School Mathematics, which had been formed in 1951, to develop a four-year program of high-school mathematics based on conceptual understanding as well as manipulative skills. With the help of Bruce Meserve at Vermont, the group produced mimeographed notes that were distributed to high-school teachers throughout Illinois. Those notes were then revised based on feedback and redistributed the following year. The initial aim was to be experimental, but by the 1960s, this aspect got lost because those mimeographed notes were converted into the *High School Mathematics* textbook series published by Heath starting in 1964.

The other report anticipating the New Math was given by Edward G. Begle, who had just been appointed as director of the School Mathematics Study Group (**SMSG**), an eight-person group created by the AMS. The launch of Sputnik in the fall of 1957 helped SMSG gain substantial support from NSF. However, the AMS distanced itself from the project's further activities and neither the *Bulletin* nor the official book on the history of the Society 1938–1988 (edited by Everett Pitcher) contains any mention whatsoever of SMSG. Nonetheless, efficient writing groups produced books that were adopted worldwide and developed ways of testing their success. Ultimately, however, parents and teachers joined in a massive public backlash against the New Math by the end of the 1960s, and consequently, SMSG fell out of favor.

Five other reports from the November 1958 CUP conference indicate other types of activities that were of prime interest at the time:

1. Films and television for mathematical instruction,
2. The role of numerical analysis in the undergraduate program,
3. Undergraduate statistics in a mathematics department,
4. The mathematical training of social scientists, and
5. The relation of mathematics to physics instruction.

For our purposes, however, the most important report was given by W. L. Duren, who reviewed CUP accomplishments over the preceding six years. He noted that although CUP had made no specific curricular suggestions, its members felt strongly that they had completed their mission, yet they recommended that a much broader study be

carried out by an expanded group. The name of the new group changed slightly, the Committee on the Undergraduate Program in Mathematics (**CUPM**), but it included most of the original members. R. Creighton Buck was appointed chair. CUPM set about expanding its investigation into the entire undergraduate program, and five years later, the committee published its most important document. Over that period, CUPM issued several vital reports that laid the foundations for the critical one in 1963.

One of CUPM's first official acts was to establish a panel on teacher training under John Kemeny with a charge to prepare a set of recommendations of minimum standards for the training of teachers on all levels. Another initiative was to establish a consultants bureau to aid colleges and universities in upgrading and revising their present undergraduate mathematics offerings or with planning new curricula. CUPM also conducted a massive survey to learn about course offerings at smaller institutions; its findings were very revealing and paint a telling picture of the mathematics undergraduate program *circa* 1960. In addition, CUPM formed the Panel on Physical Sciences and Engineering to investigate the mathematics curriculum as it related to those clients.

Arguably the most important group that CUPM assembled was the Panel on Pregraduate Training (**PPT**), initially chaired by Berkeley's John C. Moore. PPT began in 1960 by constructing an idealized program suitable for honors students in mathematics. Two years later, the panel issued a report in the *Monthly* listing its recommendations for an honors program in mathematics that had received the approval of the full CUPM beforehand [**13**]. PPT recommended a modicum of the algebra of vector spaces as well as the introduction of appropriate geometric and topological concepts into the 12-hour calculus sequence. Beyond that, the panel endorsed a one-semester course on linear algebra at the sophomore level. Regarding upper-level courses, PPT listed five (or six) courses that every college should offer its majors: two in real analysis; one in each of abstract algebra, complex analysis, and geometry-topology; and either probability or mathematical physics.

Of the four panels that performed a major portion of CUPM work, PPT became the most influential. The critical CUPM report presented the PPT's set of recommendations in 1963. The next year, CUPM executive director A. B. Willcox wrote a rather extensive, though whimsical, summary for the *Monthly*. The crux of the recommendations was to bridge the gap between undergraduate instruction and contemporary mathematics by "an idealized program which, while keeping in touch with reality, would also help set the pace for curricular improvements for some years to come" [**22**, p. 1120].

Alan Tucker recently described how, in spite of all of CUPM's work and its vetting process, the curriculum proposed in the 1963 report was soon seen to be overly ambitious [**21**]. To compensate, two years later, CUPM issued a revision recommending a watered-down version. The full set of recommendations in the 1965 report was sent to all mathematics departments, with complimentary copies available for MAA members. CUPM provided course outlines as samples for all lower-division and most upper-division courses. In an attempt to gain grassroots support, the committee asked MAA sections to evaluate the proposed curriculum and publish their findings in the *Monthly*. Many did, and almost every issue contained sectional reports on the CUPM recommendations over the next few years.

Nonetheless, the actions taken by CUPM in its reports from 1963 and 1965 were monumental and allowed William Duren to answer the question, What was the *raison d'être* of the MAA? When the MAA celebrated its semicentennial, he concluded (with brutal honesty), "I finally decided that [the] MAA existed to give comfort and status to college mathematicians .... This role of the MAA continues today [1967], but no longer as its only role" [**5**, p. 24].

The vetting process of the 1965 CUPM report by MAA sections and numerous colleges and universities over the next several years showed that it too was overly ambitious. Consequently, the committee proposed dramatic changes when it issued a new set of recommendations in 1972. This time, the MAA issued a two-volume, 700-page publication that included a 64-page extensive revision of the earlier report called "Commentary on 'A general curriculum in mathematics for colleges'." It also contained a 32-page basic library list.

The gist of the "Commentary" was a curriculum of 12 courses aimed at what CUPM regarded as the three major problem areas: (1) the evolving nature of mathematics, (2) the service functions of mathematics departments, and (3) prerequisites for entry into the program and requirements for graduation. Regarding (1), the Commentary suggested using the broader term "mathematical sciences" to account for the subject's growing interconnections with other human endeavors. Was calculus still the bedrock for the mathematical sciences? CUPM nodded in the affirmative. The report also recommended multiple curricula to account for (2). For (3), although CUPM did not go beyond recommending that the mathematics curriculum provide suitable points of entry for all students, the MAA ultimately became involved in constructing its own placement tests for use by colleges.

The curriculum for mathematics majors that CUPM recommended in its 1972 report proved to be poised at the right level and was flexible enough to accommodate future changes. It has basically withstood the test of time over the past 42 years, with slight alterations since then to address ongoing changes in the nature of mathematics. For instance, the report recommended two linear algebra courses, one essentially dealing with $\mathbb{R}^n$ and the other proof theoretic (dealing with vector spaces over fields, triangular and Jordan forms of matrices, dual spaces and tensor products, bilinear forms, and inner product spaces). CUPM also advised that every college offer courses in abstract algebra and applied mathematics, though the latter proved to be problematic for many institutions. One other notable recommendation was that probability and statistics should be offered over two semesters and not combined into one.

The flexibility built into the 1972 report allowed the mathematics major to seamlessly adjust to two developments since then. One was the "Introduction to Proofs" course aimed at easing the transition from computational to conceptual mathematics. The other was the reduction to one sophomore-level course in linear algebra that continues to have wide variation. Alan Tucker has singled out the period 1955–1974 as the Golden Age of Mathematics Majors, and CUPM played an important role in shaping the curriculum during that period [**21**].

The 1972 report was not the last one issued by CUPM. A successor was issued in 1981, that recommended a multitrack structure to account for various alternatives within the mathematical sciences. Later activities brought CUPM into collaboration with NCTM in recommending sometimes controversial standards. Even now, this influential committee is preparing an updated report in 2015.

The emphasis here has been on CUPM's recommendations on the undergraduate program for mathematics majors, but from its founding in 1958, the committee also formed subcommittees on teacher training, applied mathematics, and statistics whose findings were instrumental in the lists of recommendations. Later, CUPM added panels on two-year colleges, computer science, and minority participation to address the evolving nature of college mathematics and those who teach and study it. Most of these committees are no longer linked to CUPM.

My history of the MAA EPADEL Section [**25**] shows that, up to the 1950s, its leaders were research mathematicians and its annual meetings generally featured lectures

on cutting-edge advances. After that time, the section began to expand horizons into various pursuits under the direction of a new generation of college professors. That account reads, "From 1956 to 1978 the character of the section changed from one devoted almost exclusively to the development of mathematics to one that sponsored a variety of activities on pedagogical and curricular themes" [**25**, p. 137]. These new activities included high-school contests, a newsletter, a panel on industrial opportunities, sessions of undergraduate speakers, competitions, films, presentations and workshops on curriculum and pedagogy, and special interest groups.

The national MAA has been involved in all of these activities, and more, since that time and up to the present. One of the most successful was the result of an MAA task force formed in 1988 to address the issue of communities underrepresented in mathematics. Two years later, the MAA established **SUMMA** (Strengthening Underrepresented Minority Mathematics Achievement) to increase the representation of minorities and improve the education of minorities in mathematics. The initial fulltime staff consisted of William Hawkins and Florence Fasanelli. During 1991–1997, SUMMA sponsored five conferences for project directors through its Consortium of Intervention Programs, two national conferences devoted to the issue of attracting minorities into teaching mathematics (with proceedings published by the MAA), and a survey of minority graduate students in the mathematical sciences. In 1994, the MAA began sponsoring the David Blackwell Lecture at its annual winter meeting, named in honor of this distinguished African American mathematician; it has been sponsored solely by the National Association of Mathematicians at MathFests since 2006.

Similarly, the Association of Women in Mathematics (**AWM**) was established to improve the status of women in mathematics, from changing attitudes about girls' ability to learn mathematics as early as elementary school to ending discrimination against professional women in mathematics. Although the MAA was not directly involved in the founding of AWM, in a speech at an MAA meeting in 1990, Judy Green noted:

> The [AWM] was formed in January 1971 with a goal of encouraging the participation of women in mathematics and, at the summer meeting that year, the MAA sponsored a panel entitled "Women in Mathematics." In January 1974, the Board of Governors approved a recommendation "that the MAA participate in a joint committee with AMS in an investigation of the status of women in Mathematics." That Joint Committee still exists and has been expanded to include representatives of the other mathematical organizations. … The MAA Committee on the Participation of Women was … formed [in] 1987.

The AWM inaugurated a series of lectures at the 1996 MathFest to enhance the attraction of women and minorities into scientific careers. These lectures became a joint effort of the AWM and the MAA eight years later, when the name was changed to Etta Z. Falconer Lectures.

The Blackwell and Falconer lectures are but two of several that the MAA sponsors each year to honor a mathematician. The earliest were Hedrick Lectures, named after the first president when established in 1952; these three-lecture series have been held at summer meetings and MathFests. Two lecture series have been created in the last 16 years. The Leitzel Lectures were established in 1998 for the improvement of mathematical sciences education to honor James R. C. Leitzel for his efforts in improving that field. And the Porter Public Lectures, established in 2010 and sponsored jointly by the MAA, AMS, and SIAM, deal with a mathematical topic accessible to the broader community; they honor former MAA treasurer and now retired University of Pennsylvania mathematician Gerald Porter and his wife Judith, a retired professor of sociology at Bryn Mawr College.

Another successful MAA activity, launched in 1994, is **Project NExT** (New Experiences in Teaching), a professional development program for new and recent Ph.D.s in the mathematical sciences that introduces them to senior established mathematicians, which had been very hard until then. The program helps to ease the graduate-student-to-faculty-member transition by addressing all aspects of an academic career: teaching, scholarship, and professional activities. Almost 1500 Project NExT fellows have been chosen to date. At national meetings, they wear different colored dots on their badges to help them identify each other by year; the first ones wear red dots, and a recent *Monthly* author described himself as "a 'brown-dot' Project NExT fellow" [**20**, p. 915]. Many fellows have become national leaders; for example, 1996 fellow (blue dot) Francis Su is MAA president-elect for 2014. The program was initially funded by the Exxon Education Foundation (now the ExxonMobil Foundation). The biggest current funder is the Mary P. Dolciani-Halloran Foundation.

In 1999, an MAA task force headed by Ed Dubinsky and Ann Watkins reviewed the Association's status heading into the 21st century. The most successful of the task force's five priority action recommendations seems to have been the formation of special interest groups—**SIGMAA**s. Designed to take advantage of affordable online communication, the first specially focused group was formed in January 2000—SIGMAA RUME (Research in Undergraduate Mathematics Education). Two more SIGMAAs were formed over the next two years, whereupon the task force was disbanded and replaced by the standing committee on SIGMAAs. As of 2014, there were a dozen active SIGMAAs, with another in the works.

8. PRIZES AND AWARDS. In 1925, MAA president Julian L. Coolidge donated money to establish the MAA's first award, the Chauvenet Prize, for an outstanding expository article by an MAA member. The first recipient was G. A. Bliss. By the time the second Chauvenet Prize was given to T. H. Hildebrandt in 1929, the funding for the award had been supplemented by donations from the next two MAA presidents, Dunham Jackson (1926) and Walter B. Ford (1927). The prize was awarded every three years from 1929 until 1963, when it began being awarded annually.

The MAA has created eight other writing awards since then. The first was the Lester R. Ford Award, established in 1965 for expository articles published in either the *Monthly* or *Mathematics Magazine*. Six awards were given that year. In 1976, the Allendoerfer Award was created for articles in *Mathematics Magazine*. From that time onward, Ford Awards were restricted to the *Monthly*, with four given each year. In 2012, the Board of Governors designated this the Halmos–Ford Award.

The other writing prizes are (1) the Pólya Award, established in 1976 for expository articles published in *The College Mathematics Journal*; (2) the Beckenbach Book Prize, funded in 1986 as the successor to the MAA Book Prize, which had been created four years earlier; (3) the Hasse Prize, funded in 1986 by an anonymous donor for expository papers appearing in an Association publication; (4) the Evans Award, established by the Board of Governors in 1992 and first awarded in 1996, for articles in *Math Horizons* accessible to undergraduates; (5) the Robbins Prize, established in 2005 for outstanding papers in algebra, combinatorics, or discrete mathematics; and (6) the Euler Book Prize, established in 2005 with a gift from Virginia and Paul Halmos.

The MAA did not create any awards after the Chauvenet Prize until 1961, when the Board of Governors established the Award for Distinguished Service to the Association. It was intended to be the MAA's most prestigious honor for service; the first recipient (in 1962) was Mina Rees. The endowed successor to this award is the Gung and Hu Award for Distinguished Service to Mathematics, which was first presented in 1990 (to Leon Henkin).

The MAA has created four other awards for service since the late 1970s. The first was the Certificate of Merit, given at irregular intervals for special work or service associated with the mathematical community. The first recipient was Henry M. Cox in 1977. Six years later, the Board of Governors established a Certificate for Meritorious Service for service generally to an MAA section. The first certificates were presented in 1984. Four years later, the Joint Policy Board for Mathematics created a Communications Award to reward and encourage communicators for informing the public about mathematical ideas. (The JPBM is a collaborative effort of the MAA with the AMS, SIAM, and the ASA.) Finally, the Dolciani Award was established in 2012 for mathematicians who make distinguished contributions to the mathematical education of K–16 students in the United States or Canada.

In addition to awards for writing and service, the MAA began to create prizes for teaching about 25 years ago. The first was the Award for Distinguished College or University Teaching of Mathematics established in 1991 to honor annually at most three extraordinarily successful teachers whose influence extended beyond their own institutions. Two years later, it was renamed in honor of Deborah and Franklin Tepper Haimo. In 2003, the Alder Award for Distinguished Teaching by a Beginning College or University Mathematics Faculty Member was created for undergraduate mathematics teachers. A recent award combines teaching and research—the Selden Prize for Research in Undergraduate Mathematics Education was established in 2004 for teachers with significant records of published research in undergraduate mathematics education.

In 1926, Elizabeth Putnam created a trust fund to encourage team competition in college studies. Upon her death eight years later, the Putnam Competition assumed its present form and was administered by the MAA, which established an award for undergraduate students when the examination was first held in 1938. While that award is named for her husband, William Lowell Putnam, the Elizabeth Lowell Putnam Prize was established in 1992 for women with the highest scores. Since 1995, the Frank and Brennie Morgan Prize has been awarded jointly by the MAA, AMS, and SIAM to undergraduates for outstanding original work.

9. HEADQUARTERS. The MAA lived a vagabond existence over its first 60 years, with headquarters located at the home institution of its secretaries. Those officers passed along the massive records from one to another up to 1978, when a central organizational location was established. That year, the MAA purchased a three-building complex at 1529 18th St. NW in Washington, DC, that has served as the Association's headquarters since then. The acquisition of the properties was aided by two anonymous members who pledged substantial amounts of money toward the project.

In 2002, a $3 million donation from Paul and Virginia Halmos restored one of the three buildings, the Carriage House, which is located behind the two buildings housing MAA headquarters and the Washington, DC, offices of the AMS. The Carriage House serves as a mathematical sciences conference center. The MAA used the gift to renovate the interior completely. The Carriage House was built in 1892, 11 years before the building housing MAA headquarters.

10. CONCLUSION. The MAA was founded in late 1915 by an exuberant group dedicated to the interests of collegiate mathematics. H. E. Slaught (the father of the Association) and E. R. Hedrick (the first president) were at the forefront of this movement. With a large membership from the beginning, the Association engaged in many activities from the founding up to the mid-1950s. Most of those activities were similar to AMS actions, especially meetings at the national and sectional level. The second

half of the decade saw the MAA distinguish itself from the AMS under the leadership of E. J. McShane and W. L. Duren, Jr. The formation of CUPM was, and remains, critical for defining the undergraduate mathematics major.

This article also discussed MAA sections, prizes, awards, and headquarters. Many other aspects of the Association will be described in a forthcoming volume produced by the MAA Centennial Committee.

ACKNOWLEDGMENTS. The author hereby expresses his gratitude to four anonymous referees and to the following mathematicians for providing invaluable documentation and constructive criticism during the preparation of this paper: Judy Green, Bill Hawkins, Victor Katz, Ken Ross, Chris Stevens, and Jim Tattersall.

REFERENCES

1. Anon, Notes on the Columbus meeting, *Amer. Math. Monthly* **23** (1916) 29–30.
2. R. C. Archibald, *A Semicentennial History of the American Mathematical Society 1888–1938*. American Mathematical Society, New York, 1938, http://dx.doi.org/10.1086/347621.
3. A. A. Bennett, Brief history of the Mathematical Association of America before World War II, *Amer. Math. Monthly* **74** no. 1 Part II (1967; Fiftieth Anniversary Issue) 1–11, http://dx.doi.org/10.2307/2314864.
4. W. D. Cairns, The Mathematical Association of America, *Amer. Math. Monthly* **23** (1916) 1–6, http://dx.doi.org/10.2307/2972132.
5. W. L. Duren, Jr., CUPM, the history of an idea, *Amer. Math. Monthly* **74** no. 1 Part II (1967; Fiftieth Anniversary Issue) 23–37, http://dx.doi.org/10.2307/2314866.
6. ———, A career as a scientific generalist based in mathematics, http://www.wldurenjrmemorial.net/docs/ACareer.2.pdf.
7. E. T. Falconer, H. J. Walton, J. E. Wilkins, Jr., A. A. Shabazz, S. T. Bozeman, A history of minority participation in the Southeastern Section, April 1995. This report, a supplement to "Threescore and ten: A history of the Southeastern Section of the Mathematical Association of America 1922–1992," http://sections.maa.org/southeastern/minority/minority.html.
8. B. F. Finkel, The human aspect in the early history of the *American Mathematical Monthly, Amer. Math. Monthly* **38** (1931) 305–320, http://dx.doi.org/10.2307/2301822.
9. H. M. Gehman, The fortieth annual meeting of the Association, *Amer. Math. Monthly* **64** (1957) 212–215.
10. ———, The Washington conference, *Amer. Math. Monthly* **65** (1958) 575–586.
11. ———, Financial history, in *The Mathematical Association of America: Its First Fifty Years*. Edited by K. O. May. Mathematical Association of America, Washington, DC, 1972. 104–110.
12. P. S. Jones, Historical background and founding of the Association, in *The Mathematical Association of America: Its First Fifty Years*. Edited by K. O. May. Mathematical Association of America, Washington, DC, 1972. 1–23.
13. J. C. Moore, *et al.*, Preliminary recommendations for an honors program, *Amer. Math. Monthly* **69** (1962) 976–979, http://dx.doi.org/10.2307/2313192.
14. M. Morse, Report of the War Preparedness Committee, *Amer. Math. Monthly* **47** (1940) 500–502.
15. E. J. Moulton, The unemployment situation for Ph.D.'s in mathematics, *Amer. Math. Monthly* **42** (1935) 143–144, http://dx.doi.org/10.2307/2300635.
16. K. H. Parshall, The stratification of the American mathematical community: The Mathematical Association of America and the American Mathematical Society, 1915–1925. (forthcoming)
17. E. Pitcher, *A History of the Second Fifty-Years: American Mathematical Society, 1939–1988*. American Mathematical Society, Providence, RI, 1988.
18. H. E. Slaught, The promotion of collegiate mathematics, *Amer. Math. Monthly* **22** (1915) 251–253, http://dx.doi.org/10.2307/2971850.
19. ———, The teaching of mathematics, *Amer. Math. Monthly* **22** (1915) 289–292.
20. W. Traves, From Pascal's theorem to d-constructible curves, *Amer. Math. Monthly* **120** (2013) 901–915, http://dx.doi.org/10.4169/amer.math.monthly.120.10.901.
21. A. Tucker, The history of the undergraduate program in mathematics in the United States, *Amer. Math. Monthly* **120** (2013) 689–705, http://dx.doi.org/10.4169/amer.math.monthly.120.08.689.
22. A. B. Willcox, Pregraduate training in mathematics—A report of a CUPM panel, *Amer. Math. Monthly* **71** (1964) 1117–1129, http://dx.doi.org/10.2307/2311415.
23. J. W. Young, The work of the National Committee on Mathematics Requirements, *Mathematics Teacher* **14** (1921) 5–15.

24. ———, Functions of the Mathematical Association of America, *Amer. Math. Monthly* **39** (1932) 6–15, http://dx.doi.org/10.2307/2301454.
25. D. E. Zitarelli, *EPADEL: A Semisesquicentennial History, 1926–2000*. Raymond-Reese Book Co., Elkins Park, PA, 2001. Available upon request from the author and online at http://www.personal.psu.edu/ecb5/EPaDel/Zittarelli/EPL0_Intro.html.

DAVID E. ZITARELLI retired as professor of mathematics at Temple University in December 2012 after 42 years. He has devoted the last 25 years to the study of the history of mathematics, with an emphasis on the United States and Canada. He hopes to write a tome on that topic while books are still being printed. He and his wife Anita moved from Philadelphia to Minneapolis to be closer to grandchildren; their own children, Paul and Nicole, studied mathematics at Harvard (BS, applied math) and Cornell (BS, MS, operations research and information engineering). Hopefully, this tradition will continue with the three offspring (so far).
3441 St. Louis Ave., Minneapolis, MN 55416
zit@temple.edu

Simultaneous Proof of the First Fundamental Theorem of Calculus and Integrability of Continuous Functions

The proof of the First Fundamental Theorem of Calculus (FTC1) for continuous functions is ubiquitous: given a continuous function $f : [a, b] \to \mathbb{R}$, the ϵ-δ definition of continuity at x and a few properties of the integral allow us to conclude that $\frac{d}{dx}\int_a^x f = f(x)$ for $x \in [a, b]$ (read one-sided, as appropriate, at a and b). However, this argument assumes that continuous functions have been proven to be Riemann-integrable—another ubiquitous result whose proof typically involves an appeal to uniform continuity.

This need not be the case.* Darboux's equivalent definition of the integral of a bounded function f defines $\overline{\int_a^b} f$, the *upper integral* of f over $[a, b]$, as the infimum of all upper sums of f over the interval and $\underline{\int_a^b} f$, the *lower integral*, as the supremum of all such lower sums. Darboux calls f integrable over $[a, b]$ just when the values of $\overline{\int_a^b} f$ and $\underline{\int_a^b} f$ (which are guaranteed to exist) are equal, with $\int_a^b f$ defined as their shared value—this subtle difference gives us just enough to prove the integrability of continuous functions in the process of proving FTC1.

The properties of the integral needed in the usual proof of FTC1 are as follows (assuming integrability as necessary): (i) if $f \le g$ on $[a, b]$, then $\int_a^b f \le \int_a^b g$; (ii) for any constant $k \in \mathbb{R}$, $\int_a^b k = (b - a)k$; and (iii) if $a \le c \le b$, then $\int_a^b f = \int_a^c f + \int_c^b f$—the analogues of these three results are readily proven for upper and lower Darboux integrals without any additional (and here unnecessary) hypotheses of integrability.

Thus, the standard proof of FTC1 works without modification for upper and lower integrals: if f is continuous on $[a, b]$, then the derivatives of the functions $L(x) = \underline{\int_a^x} f$ and $U(x) = \overline{\int_a^x} f$ on $[a, b]$ are each f. From this, the mean value theorem tells us that $U(x) - L(x) = C$ for some $C \in \mathbb{R}$, and as $L(a) = \underline{\int_a^a} f = 0 = \overline{\int_a^a} f = U(a)$, we have $C = 0$. Our two conclusions immediately follow: first, L and U are identical functions on $[a, b]$, so for any $a \le x \le b$, we have that $\underline{\int_a^x} f = \overline{\int_a^x} f$—by definition, f is integrable on $[a, x]$ (and thus on all of $[a, b]$ as a special case). Second, we have by definition that $\int_a^x f = L(x)$, so $\frac{d}{dx}\int_a^x f = L'(x) = f(x)$, i.e., the first fundamental theorem of calculus.

—Submitted by Frank Swenton, Middlebury College

http://dx.doi.org/10.4169/amer.math.monthly.122.01.23
MSC: Primary 26A06, Secondary 26A42
*M. Spivak, *Calculus*. Fourth edition. Publish or Perish, Houston, TX, 2002. 295–296.

Napoleon Polygons

Titu Andreescu, Vladimir Georgiev, and Oleg Mushkarov

Abstract. An n-gon is called Napoleon if the centers of the regular n- gons erected outwardly on its sides are vertices of a regular n-gon. In this paper we obtain a new geometric characterization of Napoleon n-gons and give a new proof of the well-known theorem of Barlotti–Greber ([**1**], [**4**]) that an n-gon is Napoleon if and only if it is affine-regular. Moreover, we generalize this theorem by obtaining an analytic characterization of the n-gons leading to a regular n-gon after iterating the above construction k times.

1. INTRODUCTION. A popular topic in plane geometry is to study configurations obtained by constructing polygons on the sides of a given polygon. The most classical result in this direction is the so-called Napoleon's theorem which states that if equilateral triangles are erected outwardly (inwardly) on the sides of an arbitrary triangle, then their centers are vertices of an equilateral triangle. We refer the reader to [**2**] for the history of this theorem and its connection to Napoleon. There are various interesting generalizations of Napoleon's theorem (see, e.g., [**3**] and the literature cited there) of which we mention were obtained first by Barlotti [**1**] in 1955 and then by Greber [**4**] in 1980. It says that if regular n-gons are erected outwardly (inwardly) on the sides of an n-gon P, then their centers are vertices of a regular n-gon if and only if P is affine-regular, i.e., it is the image of a regular n-gon under an affine transformation of the plane. We call the polygons having this property Napoleon polygons.

In this paper we obtain a new geometric characterization of Napoleon polygons. Namely, we prove in Theorem 1 that any such n-gon is obtained by fixing two consecutive vertices of a regular n-gon and translating the remaining $(n-2)$ vertices by collinear vectors with lengths whose ratios we compute explicitly. As an application, we give a new proof of the theorem of Barlotti–Greber (Theorem 2). Moreover, we examine the polygons obtained by iterating the above construction. Given an n-gon P denote by $P^{(1)}$ the n-gon whose vertices are the centers of the regular n-gons erected outwardly on its sides. Then we define recursively the sequence $P^{(k)}$ of n-gons by

$$P^{(0)} = P,\ P^{(k+1)} = (P^{(k)})^{(1)} \quad \text{for } k \geq 0.$$

We say that a polygon P is k-step Napoleon if the polygon $P^{(k)}$ is regular. For example, the Barlotti–Greber theorem says that a polygon is 1-step Napoleon if and only if it is affine-regular. In Theorem 3 we generalize this result by obtaining an analytic characterization of the k-step Napoleon polygons for all $k \geq 1$.

2. NAPOLEON POLYGONS. In what follows we denote a point on the plane and the complex number it represents by the same symbol. Also, we always assume that all polygons under consideration are simple and nondegenerate plane polygons.

Given an n-gon P with vertices $z_1, z_2, \ldots, z_n$ (as usual all subscripts are taken modulo n) we denote by $P^{(1)}$ the n-gon whose vertices $z_1^{(1)}, z_2^{(1)}, \ldots, z_n^{(1)}$ are the centers of the regular n-gons erected outwardly on its sides $z_1z_2, z_2z_3, \ldots, z_nz_1$, respectively.

http://dx.doi.org/10.4169/amer.math.monthly.122.01.24
MSC: Primary 51M04, Secondary 51M20; 51N10

Definition. We say that a polygon P is Napoleon if the polygon $P^{(1)}$ is regular.

In this section we give a new proof of the Barlotti–Greber theorem mentioned in the Introduction. To do this we first prove the following analytic characterization of Napoleon polygons.

Theorem 1. *Let P be an n-gon with vertices $z_1, z_2, \ldots, z_n$ and let $z_1^0, z_2^0, \ldots, z_n^0$ be the vertices of the regular n-gon erected inwardly on the side z_1z_2 of P. Then P is a Napoleon n-gon if and only if*

$$z_k = z_k^0 + p_k \cdot u, \tag{1}$$

where u is a complex number and

$$p_k = \frac{\sin\frac{(k-2)\pi}{n}\sin\frac{(k-1)\pi}{n}}{\sin\frac{\pi}{n}\sin\frac{2\pi}{n}} \tag{2}$$

for all $1 \le k \le n$.

Proof. Set $\omega = e^{i(2\pi/n)}$. Since $z_k^{(1)}$ is the center of the regular n-gon erected outwardly on the side z_kz_{k+1} of P we have

$$z_k - z_k^{(1)} = \omega(z_{k+1} - z_k^{(1)})$$

and we get

$$z_k^{(1)} = \frac{z_k - \omega \cdot z_{k+1}}{1-\omega}, \quad 1 \le k \le n. \tag{3}$$

On the other hand, the n-gon $P^{(1)}$ is regular if and only if

$$z_{k+1}^{(1)} - z_k^{(1)} = \omega^{k-1}(z_2^{(1)} - z_1^{(1)}), \quad 1 \le k \le n. \tag{4}$$

Hence it follows from (3) and (4) that P is a Napoleon n-gon if and only if its vertices satisfy the following recursive relation:

$$\omega \cdot z_{k+2} - (1+\omega)z_{k+1} + z_k = \omega^{k-1}(\omega \cdot z_3 - (1+\omega)z_2 + z_1) \quad \text{for } 1 \le k \le n. \tag{5}$$

We now set $z_k = z_k^0 + u_k$, $1 \le k \le n$ and notice that $u_1 = u_2 = 0$. Set also $u_3 = u$. The polygon $z_0^0, z_1^0, \ldots, z_n^0$ is regular, thus the z_k^0 satisfy the relation (5) and plugging z_k in (5) gives the following relation for u_k:

$$\omega \cdot u_{k+2} - (1+\omega)u_{k+1} + u_k = \omega^k \cdot u \quad \text{for } 1 \le k \le n. \tag{6}$$

If $u = 0$, then $u_1 = u_2 = \cdots = u_n = 0$. Hence we may set $u_k = p_k \cdot u$, where $p_1 = p_2 = 0$, $p_3 = 1$ and (6) implies that the p_k satisfy the recurrence relation

$$\omega \cdot p_{k+2} - (1+\omega)p_{k+1} + p_k = \omega^k \quad \text{for } 1 \le k \le n. \tag{7}$$

Now setting $q_k = p_{k+1} - p_k$ we obtain from (7) that

$$q_k - \omega \cdot q_{k+1} = -\omega^k \quad \text{for } 1 \le k \le n.$$

It follows by induction on k that

$$q_k = \frac{1-\omega^{2k-2}}{(1-\omega^2)\omega^{k-2}} \quad \text{for } 1 \le k \le n. \tag{8}$$

Now taking into account that

$$p_k = q_2 + q_3 + \cdots + q_{k-1} \quad \text{for } 3 \le k \le n$$

and using (8) we get

$$p_k = \sum_{s=2}^{k-1} \frac{\omega^{-s+2} - \omega^s}{1-\omega^2} = \frac{(1-\omega^{k-2})(1-\omega^{k-1})}{\omega^{k-3}(1-\omega)(1-\omega^2)} \tag{9}$$

for all $1 \le k \le n$. Finally, to obtain (2) it is enough to apply the formula

$$1 - \omega^s = -2i \sin \frac{s\pi}{n} \omega^{\frac{s}{2}}$$

in the above identity. ∎

The above theorem gives a nice geometric description of the Napoleon n-gons. Namely, it shows that each of them can be obtained by fixing two consecutive vertices of a regular n-gon and translating the remaining $n-2$ vertices by collinear vectors with lengths in ratio $p_3 : p_4 : \cdots : p_n$, where p_k is given by (2).

Now we can prove the Barlotti–Greber theorem by using Theorem 1.

Theorem 2. *(Barlotti–Greber) A polygon is Napoleon if and only if it is affine-regular.*

Proof. Note first that by using complex numbers every affine transformation of the (complex) plane has the form $w = az + b\overline{z} + c$, where a, b, c are complex numbers. Hence an n-gon P with vertices $z_1, z_2, \ldots, z_n$ is affine-regular if and only if there are complex numbers a, b, c such that

$$z_k = a\omega^k + b\overline{\omega}^k + c \quad \text{for } 1 \le k \le n, \tag{10}$$

where $\omega = e^{i(2\pi/n)}$. One can see easily that P is a regular n-gon if and only if there are complex numbers a, c such that

$$z_k = a\omega^k + c, \quad 1 \le k \le n. \tag{11}$$

Let P now be an n-gon with vertices $z_1, z_2, \ldots, z_n$ and let $z_1^0, z_2^0, \ldots, z_n^0$ be the vertices of the regular n-gon erected inwardly on the side z_1z_2 of P. Then there are complex numbers a, c such that

$$z_k^0 = a\omega^k + c \quad \text{for } 1 \le k \le n. \tag{12}$$

Hence by (10), (12), and (1) in Theorem 1, it follows that to prove the theorem it is enough to show that there are complex numbers α, β, γ such that

$$a\omega^k + c + p_k u = \alpha\omega^k + \beta\overline{\omega}^k + \gamma \tag{13}$$

for all $1 \le k \le n$. Plugging the formula for p_k given in (9) and $\overline{\omega} = \frac{1}{\omega}$ in (13), and then clearing the denominators, we can write both sides of the resulting equality as quadratic functions with respect to ω^k. Now comparing the coefficients leads to

$$\alpha = a + \frac{u}{(1-\omega)(1-\omega^2)}, \quad \beta = \frac{\omega^3 \cdot u}{(1-\omega)(1-\omega^2)} \quad \text{and} \quad \gamma = c - \frac{\omega \cdot u}{(1-\omega)^2}.$$

Hence α, β, and γ are uniquely determined by a, c, and u, and vice versa. The theorem is proved. ∎

Remark 1. Theorems 1 and 2 are true also if regular n-gons are erected inwardly on the sides of the given n-gon.

3. A GENERALIZATION OF BARLOTTI–GREBER THEOREM. Given an n-gon P we can iterate the construction of the n-gon $P^{(1)}$ k times to obtain an n-gon denoted by $P^{(k)}$. More precisely, this sequence of polygons can be defined recursively as follows:

$$P^{(0)} = P, \quad P^{(s+1)} = (P^{(s)})^{(1)} \quad \text{for } s \ge 0.$$

Definition. A polygon P is said to be k-step Napoleon if the polygon $P^{(k)}$ is regular.

For instance, a polygon is 0-step Napoleon if and only if it is regular and 1-step Napoleon if it is affine-regular. The next theorem gives an analytic characterization of k-step Napoleon polygons for every $k \ge 0$. Note that in its statement we assume that the degree of 0 polynomial is -1, not $-\infty$.

Theorem 3. *An n-gon with vertices $z_1, z_2, \ldots, z_n$ is k-step Napoleon if and only if there are complex numbers a, c and a degree $k-1$ polynomial $b_{k-1}(x)$ with complex coefficients such that*

$$z_m = a\omega^m + b_{k-1}(m)\overline{\omega}^m + c \quad \textit{for } 1 \le m \le n. \tag{14}$$

Here $b_{-1} = 0$ and b_0 is a constant.

Proof. We proceed by induction on k. For $k = 0$ and $k = 1$ the statement follows by equations (10) and (11) which characterize the regular and affine-regular polygons, respectively. Suppose it is true for some k and let P be a $(k+1)$-step Napoleon n-gon. This means that $P^{(1)}$ is a k-step Napoleon polygon and it follows by the inductive assumption that

$$z_m^{(1)} = a\omega^m + b_{k-1}(m)\overline{\omega}^m + c \quad \text{for } 1 \le m \le n, \tag{15}$$

where a, c are complex numbers and $b_{k-1}(x)$ is a degree $k-1$ polynomial with complex coefficients. Similarly to equation (3) we have

$$z_m^{(1)} = \frac{z_m - \omega \cdot z_{m+1}}{1-\omega} \quad \text{for } 1 \le m \le n$$

and we obtain from (15) that

$$z_m - \omega z_{m+1} = a(1-\omega)\omega^m + (1-\omega)b_{k-1}(m)\overline{\omega}^m + c(1-\omega) \tag{16}$$

for $1 \le m \le n$. (Recall that $z_{n+1} = z_1$.) Denote the right-hand side of (16) by A_m. Then it follows by induction on m that

$$z_m = \sum_{s=m}^{n} \omega^{s-m} A_s + \omega^{n-m+1} z_1.$$

Now plugging the expression for A_s in the above identity and summing up we find that

$$z_m = \frac{a\omega^m(1-\omega^{2(n-m+1)})}{1+\omega} + (1-\omega)\overline{\omega}^m \sum_{s=m}^{n} b_{k-1}(s) + c(1-\omega^{n-m+1}) + \omega^{n-m+1} z_1.$$

Notice that $\omega^n = 1$ and $\sum_{s=m}^{n} b_{k-1}(s)$ is a polynomial of degree k on m. Hence we can write z_m in the form

$$z_m = A\omega^m + b_k(m)\overline{\omega}^m + c,$$

where

$$A = \frac{a}{1+\omega}, \quad b_k(m) = \left(z_1 - c - \frac{a}{1+\omega}\right)\omega + (1-\omega)\sum_{s=m}^{n} b_{k-1}(s).$$

Conversely, suppose that P is an n-gon whose vertices are given by (14). Then using (3) it follows easily by induction on s that for all $1 \le s \le k$ we have

$$z_m^{(s)} = a(1+\omega)^s\omega^m + d_{k-s-1}(m)\overline{\omega}^m + c \quad \text{for } 1 \le m \le n,$$

where $d_{k-s-1}(x)$ is a polynomial of degree $k-s-1$. Hence

$$z_m^{(k)} = a(1+\omega)^k\omega^m + c, \quad 1 \le m \le n$$

and therefore $P^{(k)}$ is a regular polygon. Thus the theorem is proved. ■

Finally, let us note that Theorem 3 holds also true if we construct regular n-gons inwardly on the sides of the given n-gon P. In this case one has to switch the roles of ω and $\overline{\omega}$ in the formula for the vertices of P.

ACKNOWLEDGMENT. The authors are grateful to the referee for the remarks and suggestions to improve the first version of this work. The second and third authors are partially supported by the Comenius project "Dynamat" Project No. 510028-LLP-1-2010-1-IT-COMENIUS-CMP.

REFERENCES

1. A. Barlotti, Una Proprietá degli n-agoni che si ottengono Transformando in una Affinata un n-agono Regolare, *Boll. Unione Mat. Ital.* **10** (1955) 96–98.
2. P. Davis, *Mathematical Encounters of the Second Kind*. Birkhäuser, Boston, 1987, http://dx.doi.org/10.1007/978-1-4612-2462-4.
3. S. B. Gray, Generalizing the Peter-Douglas-Neuman theorem on n-gons, *Amer. Math. Monthly* **110** (2003) 210–227, http://dx.doi.org/10.2307/3647935.
4. L. Greber, Napoleon's theorem and the parallelogram inequality for affine-regular polygons, *Amer. Math. Monthly* **87** (1980) 644–648, http://dx.doi.org/10.2307/2320952.

TITU ANDREESCU was born in 1956 in Timisoara, Romania. He received his Ph.D. in mathematics from University of West Timisoara in 2003. He is an associate professor at The University of Texas at Dallas. Titu is the author, co-author, or editor of over 40 books. He is the founder of AwesomeMath and XYZ Press. Titu's main areas of research interests are Quadratic Diophantine Equations and mathematics education, with emphasis on teaching gifted students and mathematics competitions.

School of Natural Sciences and Mathematics, The University of Texas at Dallas, 800 W. Campbell Road, Richardson, TX 75080, USA
titu.andreescu@utdallas.edu

VLADIMIR GEORGIEV was born in 1956 in Sofia, Bulgaria. He received his Ph.D. in mathematics from University of Sofia in 1983. In the period 1990–1991 he was a Humboldt fellow at University of Bonn. He was in the Department of Mathematical Physics at the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences until 1997. After this he was a full professor at University of L'Aquila, Italy and since 2001 is a full professor at University of Pisa, Italy. His main research interests are partial differential equations and problems of mathematical physics.

Department of Mathematics, University of Pisa, Largo Bruno Pontecorvo 5 Pisa, 56127 Italy
georgiev@dm.unipi.it

OLEG MUSHKAROV was born in 1951 in Blagoevgrad, Bulgaria. He received his Ph.D. in mathematics from University of Sofia in 1979. Since then he has been in the Department of Complex Analysis at the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, where he is now a full professor. His main research interests are in the areas of complex analysis, complex differential geometry, and twistor theory.

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str. Bl. 8, 1113 Sofia, Bulgaria
muskarov@math.bas.bg

Zeta of Two Equals Pi Squared O'er Six

That zeta of two equals pi squared o'er six
Has been shown to be so by a boatload of tricks.
First Euler exploited a product for sine,
Although absolute rigor was not in his line.

Then followed a horde from Calabi to me
Who nailed down our proofs from step A to step Z.
With proud demonstrations, some plain ones, some subtlish,
We papered the walls and then hastened to publish.

It's not that I argue one proof is enough
For a formula once thought exceedingly tough,
But why do we need many more than a dozen,
A new proof by each Tom, Dick 'n Harry's cousin?

There's a proof without calculus based on de Moivre.
There are proofs with more sizzle than beefsteak au poivre.
There are proofs that require no more than Calc I.
There are proofs that use everything under the sun.

Now let's hear it for having a little restraint—
An existence-of-God proof this theorem ain't—
And should our sheer numbers in Guiness' book place us,
Know Pythagoras' followers there would erase us.

—Submitted by Ralph Krause, Washington DC

Historical Remark on Ramanujan's Tau Function

Kenneth S. Williams

Abstract. It is shown that Ramanujan could have proved a special case of his conjecture that his tau function is multiplicative without any recourse to modularity results.

1. INTRODUCTION. In his path-breaking paper on arithmetic functions published in 1916, Ramanujan [**6**, eq. (92)] introduced the function $\tau(n)$ that in his honor is now called the Ramanujan tau function. This function is defined for all positive integers n by

$$q\prod_{m=1}^{\infty}(1-q^m)^{24}=\sum_{n=1}^{\infty}\tau(n)q^n,\quad q\in\mathbb{C},\quad |q|<1. \tag{1}$$

Ramanujan calculated the first 30 values of $\tau(n)$ [**6**, Table V] and observed that $\tau(n)$ appeared to be multiplicative [**6**, eq. (103)], that is,

$$\tau(n_1n_2)=\tau(n_1)\tau(n_2),\quad n_1,n_2\in\mathbb{N},\quad \gcd(n_1,n_2)=1. \tag{2}$$

This was proved by Mordell [**5**] shortly afterward using modular techniques, which were unknown to Ramanujan. A modern proof of (2) is given, for example, in [**4**, p. 298]. The author is not aware of any proof of (2) that does not appeal to the theory of modular forms.

It is known from the theory of modular forms for all primes p and all positive integers n that the following property of $\tau(n)$ holds, namely,

$$\tau(pn)=\tau(p)\tau(n)-p^{11}\tau(n/p), \tag{3}$$

where $\tau(n/p)=0$ if p does not divide n [**4**, p. 298]. Moreover, a simple induction argument using (3) gives the multiplicativity property (2) of $\tau(n)$; see, for example, [**4**, Cor. 5.6, p. 298].

The purpose of this historical note is to show that Ramanujan had all the tools necessary to prove the special case of (3) when $p=2$, namely,

$$\tau(2n)=\tau(2)\tau(n)-2^{11}\tau(n/2),\quad n\in\mathbb{N}, \tag{4}$$

from which the multiplicativity property

$$\tau(2^kN)=\tau(2^k)\tau(N),\quad \text{for } k\in\mathbb{N}\cup\{0\},\ N\in\mathbb{N},\ \text{and } N \text{ odd}, \tag{5}$$

follows.

http://dx.doi.org/10.4169/amer.math.monthly.122.01.30
MSC: Primary 11A05, Secondary 11F20; 11F27; 01A60; 11F11

2. PROOF OF (4) IN THE SPIRIT OF RAMANUJAN. Ramanujan defined a general theta function [**3**, Definition 1.2.1, p. 6] and studied its properties. An important special case of his function is the theta function $\varphi(q)$ given by

$$\varphi(q) := \sum_{n=-\infty}^{\infty} q^{n^2}, \quad \text{for } q \in \mathbb{C} \text{ and } |q| < 1. \tag{6}$$

Ramanujan knew many properties of $\varphi(q)$, including the two simple identities

$$\varphi^2(q^2) = \frac{1}{2}(\varphi^2(q) + \varphi^2(-q)) \quad \text{and} \quad \varphi^2(-q^2) = \varphi(q)\varphi(-q); \tag{7}$$

see [**3**, pp. 15, 72]. Ramanujan also used extensively three Eisenstein series, which he denoted by $P(q)$, $Q(q)$, and $R(q)$ [**6**, eq. (25)], the latter two of which are

$$Q(q) := 1 + 240 \sum_{n=1}^{\infty} \frac{n^3 q^n}{1 - q^n}, \quad q \in \mathbb{C}, \quad |q| < 1, \tag{8}$$

and

$$R(q) := 1 - 504 \sum_{n=1}^{\infty} \frac{n^5 q^n}{1 - q^n}, \quad q \in \mathbb{C}, \quad |q| < 1. \tag{9}$$

Ramanujan's pioneering work established the relationship between Eisenstein series and theta functions. He was well aware of results of the type

$$Q(q) = (1 + 14x + x^2)z^4 \tag{10}$$

and

$$R(q) = (1 - 33x - 33x^2 + x^3)z^6, \tag{11}$$

[**3**, Theorems 5.4.11 and 5.4.12], where [**3**, p. 120]

$$x = x(q) := 1 - \frac{\varphi^4(-q)}{\varphi^4(q)} \quad \text{and} \quad z = z(q) := \varphi^2(q), \tag{12}$$

so that

$$Q(q) = 16\varphi^8(q) - 16\varphi^4(q)\varphi^4(-q) + \varphi^8(-q) \tag{13}$$

and

$$R(q) = -64\varphi^{12}(q) + 96\varphi^8(q)\varphi^4(-q) - 30\varphi^4(q)\varphi^8(-q) - \varphi^{12}(-q). \tag{14}$$

In [**6**, eq. (44)], Ramanujan proved, using only an elementary argument, the fundamental relation

$$1728q \prod_{m=1}^{\infty} (1 - q^m)^{24} = Q^3(q) - R^2(q). \tag{15}$$

The following relation follows from (1) and (13)–(15):

$$16\sum_{n=1}^{\infty}\tau(n)q^n = \varphi^8(q)\varphi^{16}(-q) - \varphi^4(q)\varphi^{20}(-q). \tag{16}$$

Replacing q by $-q$ in (16) and adding the resulting equation to (16) gives

$$32\sum_{n=1}^{\infty}\tau(2n)q^{2n} = -\varphi^{20}(q)\varphi^4(-q) + \varphi^{16}(q)\varphi^8(-q) + \varphi^8(q)\varphi^{16}(-q) - \varphi^4(q)\varphi^{20}(-q). \tag{17}$$

Replacing q by q^2 in (16) and making use of (7) gives

$$256\sum_{n=1}^{\infty}\tau(n)q^{2n} = \varphi^{16}(q)\varphi^8(-q) - 2\varphi^{12}(q)\varphi^{12}(-q) + \varphi^8(q)\varphi^{16}(-q). \tag{18}$$

Replacing q by q^2 in (18) and appealing to (7) gives

$$65536\sum_{n=1}^{\infty}\tau(n)q^{4n} = \varphi^{20}(q)\varphi^4(-q) - 4\varphi^{16}(q)\varphi^8(-q) + 6\varphi^{12}(q)\varphi^{12}(-q) - 4\varphi^8(q)\varphi^{16}(-q) + \varphi^4(q)\varphi^{20}(-q). \tag{19}$$

Then (17)–(19) give

$$\sum_{n=1}^{\infty}\tau(2n)q^{2n} + 24\sum_{n=1}^{\infty}\tau(n)q^{2n} + 2048\sum_{n=1}^{\infty}\tau(n)q^{4n} = 0 \tag{20}$$

so that

$$\tau(2n) + 24\tau(n) + 2048\tau(n/2) = 0, \quad n \in \mathbb{N}, \tag{21}$$

from which (4) follows as $\tau(2) = -24$ and $2^{11} = 2048$. ■

It would be very interesting to know if Ramanujan had a proof along these lines for the special case (5) of his conjecture.

3. PROOF OF (3) FOR $p = 3$. The question naturally arises, "Can (3) be proved for primes $p \neq 2$ in a similar manner to the elementary proof given in Section 2 for $p = 2$?" However, this seems to be quite difficult. We carry out the proof for the prime $p = 3$ and at the end of the proof summarize the difficulties involved in giving such a proof for an arbitrary prime $p > 3$.

As $\tau(3) = 252$ and $3^{11} = 177147$, we must prove analogously to (20) that

$$\sum_{n=1}^{\infty}\tau(3n)q^{3n} - 252\sum_{n=1}^{\infty}\tau(n)q^{3n} + 177147\sum_{n=1}^{\infty}\tau(n)q^{9n} = 0. \tag{22}$$

We determine the sum of each of the three infinite series in (22) individually in terms of the function φ. First, from (16), we have

$$16\sum_{n=1}^{\infty}\tau(n)q^{3n}=\varphi^8(q^3)\varphi^{16}(-q^3)-\varphi^4(q^3)\varphi^{20}(-q^3) \tag{23}$$

and

$$16\sum_{n=1}^{\infty}\tau(n)q^{9n}=\varphi^8(q^9)\varphi^{16}(-q^9)-\varphi^4(q^9)\varphi^{20}(-q^9). \tag{24}$$

Now we turn to $\sum_{n=1}^{\infty}\tau(3n)q^{3n}$. We let $\omega=\exp(2\pi i/3)$ and note that

$$1+\omega^n+\omega^{2n}=\begin{cases}3 & \text{if } n\equiv 0 \pmod 3,\\ 0 & \text{if } n\not\equiv 0 \pmod 3.\end{cases} \tag{25}$$

By (25), we have

$$\begin{aligned}48\sum_{n=1}^{\infty}\tau(3n)q^{3n}&=16\sum_{n=1}^{\infty}\tau(n)(1+\omega^n+\omega^{2n})q^n\\&=16\sum_{n=1}^{\infty}\tau(n)q^n+16\sum_{n=1}^{\infty}\tau(n)(\omega q)^n+16\sum_{n=1}^{\infty}\tau(n)(\omega^2 q)^n,\end{aligned}$$

so that by (16),

$$\begin{aligned}48\sum_{n=1}^{\infty}\tau(3n)q^{3n}&=\varphi^8(q)\varphi^{16}(-q)-\varphi^4(q)\varphi^{20}(-q)\\&\quad+\varphi^8(\omega q)\varphi^{16}(-\omega q)-\varphi^4(\omega q)\varphi^{20}(-\omega q)\\&\quad+\varphi^8(\omega^2 q)\varphi^{16}(-\omega^2 q)-\varphi^4(\omega^2 q)\varphi^{20}(-\omega^2 q).\end{aligned} \tag{26}$$

Next, we consider $\varphi(\omega q)$. From (6), we deduce that

$$\varphi(\omega q)=\sum_{n=-\infty}^{\infty}\omega^{n^2}q^{n^2}=\sum_{\substack{n=-\infty\\ n\equiv 0 \pmod 3}}^{\infty}q^{n^2}+\omega\sum_{\substack{n=-\infty\\ n\not\equiv 0 \pmod 3}}^{\infty}q^{n^2},$$

as $n^2\equiv 1 \pmod 3$ for $n\not\equiv 0 \pmod 3$. Hence,

$$\varphi(\omega q)=\varphi(q^9)+\omega(\varphi(q)-\varphi(q^9)). \tag{27}$$

Similarly, we have

$$\varphi(-\omega q)=\varphi(-q^9)+\omega(\varphi(-q)-\varphi(-q^9)), \tag{28}$$

$$\varphi(\omega^2 q)=\varphi(q^9)+\omega^2(\varphi(q)-\varphi(q^9)), \tag{29}$$

and

$$\varphi(-\omega^2 q)=\varphi(-q^9)+\omega^2(\varphi(-q)-\varphi(-q^9)). \tag{30}$$

Using (27)–(30) in (26), we deduce

$$\begin{aligned}
&48\sum_{n=1}^{\infty}\tau(3n)q^{3n}\\
&\quad=\varphi^8(q)\varphi^{16}(-q)-\varphi^4(q)\varphi^{20}(-q)\\
&\qquad+(\varphi(q^9)+\omega(\varphi(q)-\varphi(q^9)))^8(\varphi(-q^9)+\omega(\varphi(-q)-\varphi(-q^9)))^{16}\\
&\qquad-(\varphi(q^9)+\omega(\varphi(q)-\varphi(q^9)))^4(\varphi(-q^9)+\omega(\varphi(-q)-\varphi(-q^9)))^{20}\\
&\qquad+(\varphi(q^9)+\omega^2(\varphi(q)-\varphi(q^9)))^8(\varphi(-q^9)+\omega^2(\varphi(-q)-\varphi(-q^9)))^{16}\\
&\qquad-(\varphi(q^9)+\omega^2(\varphi(q)-\varphi(q^9)))^4(\varphi(-q^9)+\omega^2(\varphi(-q)-\varphi(-q^9)))^{20}. \qquad (31)
\end{aligned}$$

Multiplying (22) by 48, and appealing to (31), (23), and (24), we must prove the following identity relating $\varphi(q)$, $\varphi(-q)$, $\varphi(q^3)$, $\varphi(-q^3)$, $\varphi(q^9)$, and $\varphi(-q^9)$, namely,

$$\begin{aligned}
&\varphi^8(q)\varphi^{16}(-q)-\varphi^4(q)\varphi^{20}(-q)\\
&\qquad+(\varphi(q^9)+\omega(\varphi(q)-\varphi(q^9)))^8(\varphi(-q^9)+\omega(\varphi(-q)-\varphi(-q^9)))^{16}\\
&\qquad-(\varphi(q^9)+\omega(\varphi(q)-\varphi(q^9)))^4(\varphi(-q^9)+\omega(\varphi(-q)-\varphi(-q^9)))^{20}\\
&\qquad+(\varphi(q^9)+\omega^2(\varphi(q)-\varphi(q^9)))^8(\varphi(-q^9)+\omega^2(\varphi(-q)-\varphi(-q^9)))^{16}\\
&\qquad-(\varphi(q^9)+\omega^2(\varphi(q)-\varphi(q^9)))^4(\varphi(-q^9)+\omega^2(\varphi(-q)-\varphi(-q^9)))^{20}\\
&\qquad-756(\varphi^8(q^3)\varphi^{16}(-q^3)-\varphi^4(q^3)\varphi^{20}(-q^3))\\
&\qquad+531441(\varphi^8(q^9)\varphi^{16}(-q^9)-\varphi^4(q^9)\varphi^{20}(-q^9))=0. \qquad (32)
\end{aligned}$$

The most elementary way of proving the identity (32) known to the author is to use the (p,k)-parametrizations of $\varphi(q)$, $\varphi(q^3)$, $\varphi(q^9)$, $\varphi(-q)$, $\varphi(-q^3)$, and $\varphi(-q^9)$ due to Alaca, Alaca, and Williams [**2**]. (We emphasize that here p is a function of q and is not being used to denote a prime.) We note that all of these parametrizations have been proved without the use of modular forms. As in [**2**, p. 178], we set

$$p=p(q):=\frac{\varphi^2(q)-\varphi^2(q^3)}{2\varphi^2(q^3)},\quad k=k(q):=\frac{\varphi^3(q^3)}{\varphi(q)}, \qquad (33)$$

so that

$$\varphi(q)=(1+2p)^{3/4}k^{1/2} \qquad (34)$$

and

$$\varphi(q^3)=(1+2p)^{1/4}k^{1/2}. \qquad (35)$$

A. Alaca [**1**, Theorem 2.2, p. 156] has shown that

$$\varphi(q^9)=\frac{1}{3}(1+2p)^{3/4}k^{1/2}+\frac{2^{2/3}}{3}(1+2p)^{1/12}(1-p)^{1/3}(2+p)^{1/3}k^{1/2}. \qquad (36)$$

The "change of sign" principle [**2**, Theorem 11, p. 180] asserts that

$$p(-q)=\frac{-p}{1+p}\quad\text{and}\quad k(-q)=(1+p)^2k. \qquad (37)$$

Changing q to $-q$ in (34)–(36) and appealing to (37), we obtain

$$\varphi(-q) = (1-p)^{3/4}(1+p)^{1/4}k^{1/2}, \tag{38}$$

$$\varphi(-q^3) = (1-p)^{1/4}(1+p)^{3/4}k^{1/2}, \tag{39}$$

and

$$\varphi(-q^9) = \frac{1}{3}(1-p)^{3/4}(1+p)^{1/4}k^{1/2} + \frac{2^{2/3}}{3}(1-p)^{1/12}(1+2p)^{1/3}(2+p)^{1/3}(1+p)^{1/4}k^{1/2}. \tag{40}$$

Using MAPLE to substitute (34)–(36) and (38)–(40) into (32) and to simplify the resulting expression, we find that it is equal to 0, thereby establishing the identity (32) and proving (3) in the case $p = 3$. ■

In attempting to extend this elementary argument to an arbitrary prime $p > 3$, three obstacles become apparent. First, we do not know at the outset the value of $\tau(p)$ to use in the analogue of (22). Secondly, we need to determine $\varphi(\omega q)$, where $\omega = \exp(2\pi i/p)$, analogously to (27). Finally, a parametrization of $\varphi(q)$, $\varphi(q^p)$, $\varphi(q^{p^2})$, $\varphi(-q)$, $\varphi(-q^p)$, and $\varphi(-q^{p^2})$ would be helpful in order to verify the identity analogous to (32) for a prime $p > 3$.

Ramanujan's tau function is nearly a century old. It is hoped that this historical note will encourage the reader to learn more of its interesting properties and its place in the theory of modular forms.

ACKNOWLEDGMENT. The author thanks the two anonymous referees for their useful suggestions.

REFERENCES

1. A. Alaca, Representations by quaternary quadratic forms whose coefficients are 1, 3 and 9, *Acta Arith.* **136** (2009) 151–166.
2. A. Alaca, Ş. Alaca, K. S. Williams, On the two-dimensional theta functions of the Borweins, *Acta Arith.* **124** (2006) 177–195.
3. B. C. Berndt, *Number Theory in the Spirit of Ramanujan.* American Mathematical Society, Providence, RI, 2006.
4. H. M. Farkas, I. Kra, *Theta Constants, Riemann Surfaces and the Modular Group.* American Mathematical Society, Providence, RI, 2001.
5. L. J. Mordell, On Mr. Ramanujan's empirical expansions of modular functions, *Proc. Cambridge Philos. Soc.* **19** (1917) 117–124.
6. S. Ramanujan, On certain arithmetical functions, *Trans. Cambridge Philos. Soc.* **22** (1916) 159–184. [*Collected Papers of Srinivasa Ramanujan*, AMS Chelsea Publishing, American Mathematical Society, Providence, RI, 2000, pp. 136–162.]

KENNETH S. WILLIAMS received his B.Sc. from the University of Birmingham, England in 1962 and his M.A. and Ph.D. degrees from the University of Toronto in 1963 and 1965, respectively. In 1979, he was awarded a D.Sc. by the University of Birmingham for his research in the theory of numbers. He retired from Carleton University in 2002 after 36 years of service. In retirement, he enjoys time with his family, gardening, birding with his wife, and visiting with his young granddaughter Isabelle.
School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada K1S 5B6
kwilliam@connect.carleton.ca

Somewhat Stochastic Matrices

Branko Ćurgus and Robert I. Jewett

Abstract. The standard theorem for stochastic matrices with positive entries is generalized to matrices with no sign restriction on the entries. The condition that column sums be equal to 1 is kept, but the positivity condition is replaced by a condition on the distances between columns.

1. INTRODUCTION. The notion of a Markov chain is ubiquitous in linear algebra and probability books. In linear algebra, a Markov chain is a sequence $\{\mathbf{x}_k\}$ of vectors defined recursively by a specified vector $\mathbf{x}_0$, a square matrix P, and the recursion $\mathbf{x}_k = P\mathbf{x}_{k-1}$ for $k = 1, 2, \ldots$. That is, $\mathbf{x}_k = P^k\mathbf{x}_0$. Natural probabilistic restrictions are imposed on $\mathbf{x}_0$ and P. It is assumed that $\mathbf{x}_0$ is a *probability vector*, that is, its entries are nonnegative and add up to 1. It is assumed that P is a *stochastic matrix*, that is, it is a square matrix whose columns are probability vectors. The original version of the main theorem about Markov chains appears in Markov's paper [**2**]. In the language of linear algebra, it reads:

> *Suppose that P is a stochastic matrix with all positive entries. Then there exists a unique probability vector $\mathbf{q}$ such that $P\mathbf{q} = \mathbf{q}$. If $\{\mathbf{x}_k\}$ is a Markov chain determined by P, then it converges to $\mathbf{q}$.*

More generally, the same conclusion holds for a stochastic matrix P for which P^s has all positive entries for some positive integer s. All elementary linear algebra textbooks that we examined state this theorem. None give a complete proof. Partial proofs or intuitive explanations of the theorem's validity are always based on knowledge about the matrix's eigenvalues and eigenvectors. This argument becomes sophisticated when the matrix is not diagonalizable.

What these proofs leave obscure is a certain contractive property of a stochastic matrix already observed by Markov. Of course, this contractive property is explored in research papers and some advanced books. However, the relative simplicity of the underlying idea gets lost in the technical details of an advanced setting. We feel that this contractive property deserves to be popularized. We use it here to provide a direct proof of a theorem, which is more general than the one stated above.

We consider real square matrices A whose columns add up to 1. Such a matrix we call a *somewhat stochastic* matrix. The probabilistic condition that all entries be nonnegative is dropped. Instead of assuming that all entries of A^s are positive, we make an assumption about distances between the columns of A^s. This assumption leads to a contractive property of a matrix that yields convergence. This and other definitions are given next.

2. DEFINITIONS. All numbers in this article are real, except in Example 3. All matrices are square and will be denoted by uppercase letters. All vectors, except for

http://dx.doi.org/10.4169/amer.math.monthly.122.01.36
MSC: Primary 15B51, Secondary 15B48; 60J10

$\mathbb{1}$, are column vectors and will be denoted by bold lowercase letters. All entries of the row vector $\mathbb{1}$ are equal to 1,

$$\mathbb{1} = [1 \;\; 1 \;\cdots\; 1].$$

This row vector helps to express conditions that were already mentioned and will appear repeatedly. The equation $\mathbb{1}A = \mathbb{1}$ says that all column sums of A are equal to 1. And $\mathbb{1}\mathbf{x} = 1$ says that the entries of a column vector $\mathbf{x}$ sum to 1, while $\mathbb{1}\mathbf{x} = 0$ says that the entries of a column vector $\mathbf{x}$ sum to 0. We use the standard notation $\mathbb{N}$ for the set of positive integers and $\mathbb{R}$ for the set of real numbers.

For a vector $\mathbf{x}$ with entries $x_1, \ldots, x_n$, we set

$$\|\mathbf{x}\| := \sum_{j=1}^{n} |x_j|.$$

Notice that the distance $\|\mathbf{x} - \mathbf{y}\|$ between vectors $\mathbf{x}$ and $\mathbf{y}$ associated with this norm is the n-dimensional version of the Manhattan distance.

Consider an $n \times n$ matrix A with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and entries a_{ij}. For the purpose of the next two definitions, we think of the columns of a matrix as points in $\mathbb{R}^n$. In this way, the concept of a diameter of a set is applied to a matrix as follows:

$$\operatorname{diam} A = \max_{1 \le i,j \le n} \|\mathbf{a}_i - \mathbf{a}_j\|.$$

Next, we define

$$\operatorname{var} A = \frac{1}{2} \operatorname{diam} A = \max_{1 \le i,j \le n} \frac{1}{2} \sum_{l=1}^{n} |a_{li} - a_{lj}|.$$

We call this quantity the *column variation* of a matrix A. The idea of using that quantity is due to Markov [**2**, Section 5]. In [**2**], for fixed i, j the quantity $\frac{1}{2}\sum_{l=1}^{n} |a_{li} - a_{lj}|$ is not given explicitly as half the sum of the absolute values of the real numbers $a_{li} - a_{lj}$ but rather as the sum of the positive terms in this list. Since Markov considered only stochastic matrices, for which the sum of all terms in this list is 0, the quantity he used coincides with the variation. For more on Markov's work, see [**4**]. The column and row variation appear in research literature under various names; see [**1**, Section 3.3].

Recalling the original theorem about a Markov chain stated in our first paragraph, we will show that the inequality

$$\|P^k\mathbf{x}_0 - \mathbf{q}\| \le (\operatorname{var} P)^k \|\mathbf{x}_0 - \mathbf{q}\| \tag{1}$$

holds for all k. Furthermore, for a stochastic matrix P with all positive entries, it turns out that $\operatorname{var} P < 1$. This strict contractive property of a stochastic matrix with all positive entries implies convergence of the Markov chain $\{P^k\mathbf{x}_0\}$.

3. THE COLUMN VARIATION OF A MATRIX. The first step toward a proof of inequality (1) is the proposition that follows. To repeat, our results do not require entries to be nonnegative. Not only that, but in this proposition A could be a rectangular matrix with complex entries. However, the assumption that the entries of the vector $\mathbf{y}$ are real is essential, as is shown in Example 3.

Proposition. *Let A be an $n \times n$ real matrix and let $\mathbf{y}$ be an $n \times 1$ vector with real entries such that $\mathbb{1}\mathbf{y} = 0$. Then*

$$\|A\mathbf{y}\| \leq (\operatorname{var} A)\|\mathbf{y}\|. \tag{2}$$

Proof. We will use the common notation for the positive and negative part of a real number t: $t^+ = \max\{t, 0\}$ and $t^- = \max\{-t, 0\}$. Clearly, $t^+, t^- \geq 0$, and $t = t^+ - t^-$ and $|t| = t^+ + t^-$.

Let A be an $n \times n$ matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and let $\mathbf{y} \in \mathbb{R}^n$ be such that $\mathbb{1}\mathbf{y} = 0$.

The inequality (2) is obvious if $\mathbf{y} = \mathbf{0}$. Assume that $\mathbf{y} \neq \mathbf{0}$. Set $\mathbf{z} = \big(2/\|\mathbf{y}\|\big)\mathbf{y}$. Then (2) is equivalent to

$$\|A\mathbf{z}\| \leq (\operatorname{var} A)\|\mathbf{z}\| \quad \text{with} \quad \|\mathbf{z}\| = 2. \tag{3}$$

Clearly $\mathbb{1}\mathbf{z} = 0$. Let $z_1, \ldots, z_n$ be the entries of $\mathbf{z}$. Then we have

$$2 = \|\mathbf{z}\| = \sum_{j=1}^{n} |z_j| = \sum_{j=1}^{n} \big(z_j^+ + z_j^-\big) = \sum_{j=1}^{n} z_j^+ + \sum_{j=1}^{n} z_j^-$$

and

$$0 = \sum_{j=1}^{n} z_j = \sum_{j=1}^{n} \big(z_j^+ - z_j^-\big) = \sum_{j=1}^{n} z_j^+ - \sum_{j=1}^{n} z_j^-.$$

From the last two displayed relations, we deduce that

$$\sum_{j=1}^{n} z_k^+ = \sum_{j=1}^{n} z_k^- = 1. \tag{4}$$

Using again the notation introduced at the beginning of the proof, we get

$$A\mathbf{z} = \sum_{j=1}^{n} z_j \mathbf{a}_j = \sum_{j=1}^{n} z_j^+ \mathbf{a}_j - \sum_{i=1}^{n} z_i^- \mathbf{a}_i. \tag{5}$$

Since $A\mathbf{z}$ is represented in (5) as a difference of two convex combinations of the columns of A, the inequality $\|A\mathbf{z}\| \leq \operatorname{diam} A$ follows from the geometrically clear fact that a set has the same diameter as its convex hull. However, we use (4) and (5) to continue with an algebraic argument:

$$\begin{aligned} A\mathbf{z} &= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} z_i^- \right) z_j^+ \mathbf{a}_j - \sum_{i=1}^{n} \left(\sum_{j=1}^{n} z_j^+ \right) z_i^- \mathbf{a}_i \\ &= \sum_{j=1}^{n} \sum_{i=1}^{n} z_j^+ z_i^- (\mathbf{a}_j - \mathbf{a}_i). \end{aligned}$$

Consequently,

$$\begin{aligned}
\|A\mathbf{z}\| &\le \sum_{j=1}^{n}\sum_{i=1}^{n} z_j^+ z_i^- \|\mathbf{a}_j - \mathbf{a}_i\| && \text{(by the triangle inequality and } z_j^+, z_j^- \ge 0\text{)}\\
&\le (\operatorname{diam} A)\sum_{k=1}^{n} z_k^+ \sum_{j=1}^{n} z_j^- && \text{(by definition of diam } A\text{)}\\
&= 2(\operatorname{var} A) && \text{(by (4) and definition of var } A\text{)}\\
&= (\operatorname{var} A)\|\mathbf{z}\| && \text{(since } \|\mathbf{z}\| = 2\text{)}.
\end{aligned}$$

This completes the proof of (3) and the theorem is proved. ∎

4. POWERS OF SOMEWHAT STOCHASTIC MATRICES. Let A be a somewhat stochastic matrix, that is, A is square and real and $\mathbb{1}A = \mathbb{1}$. The equalities

$$\mathbb{1}A^k = (\mathbb{1}A)A^{k-1} = \mathbb{1}A^{k-1} = \cdots = (\mathbb{1}A)A = \mathbb{1}A = \mathbb{1}$$

show that any power A^k of A is somewhat stochastic. Furthermore, if $\mathbb{1}\mathbf{y} = 0$, then $\mathbb{1}A^k\mathbf{y} = \mathbb{1}\mathbf{y} = 0$ for all positive integers k. This property of a somewhat stochastic matrix A allows us to repeatedly apply the proposition to powers of A. Assuming that $\mathbb{1}\mathbf{y} = 0$ and, for the sake of simplicity, setting $c = \operatorname{var} A$, we have

$$\|A^k\mathbf{y}\| = \|A(A^{k-1}\mathbf{y})\| \le c\|A^{k-1}\mathbf{y}\| \le \cdots \le c^{k-1}\|A\mathbf{y}\| \le c^k\|\mathbf{y}\|. \tag{6}$$

Now we are ready to state and prove the main result.

Theorem. *Let A be an $n \times n$ somewhat stochastic matrix. Assume that there exists $s \in \mathbb{N}$ such that* $\operatorname{var}(A^s) < 1$. *Then*

(a) *there exists a unique $\mathbf{q} \in \mathbb{R}^n$ such that $A\mathbf{q} = \mathbf{q}$ and $\mathbb{1}\mathbf{q} = 1$,*

(b) *if $\mathbf{x}$ is such that $\mathbb{1}\mathbf{x} = 1$, then the sequence $\{A^k\mathbf{x}\}$ converges to $\mathbf{q}$ as k tends to $+\infty$.*

Proof. The assumption that $\mathbb{1}A = \mathbb{1}$ means that 1 is an eigenvalue of $A^\top$, the transpose of A. Since $A^\top$ and A have the same eigenvalues, there exists a real nonzero vector $\mathbf{v}$ such that $A\mathbf{v} = \mathbf{v}$.

Let s be a positive integer such that $\operatorname{var}(A^s) < 1$ and set $c = \operatorname{var}(A^s)$. Clearly, $A^s\mathbf{v} = \mathbf{v}$. If $\mathbb{1}\mathbf{v} = 0$, then the proposition yields

$$\|\mathbf{v}\| = \|A^s\mathbf{v}\| \le c\|\mathbf{v}\| < \|\mathbf{v}\|.$$

This is a contradiction. Therefore, $\mathbb{1}\mathbf{v} \ne 0$. Setting $\mathbf{q} = (\mathbb{1}\mathbf{v})^{-1}\mathbf{v}$ provides a vector whose existence is claimed in (a). To verify uniqueness, let $\mathbf{p}$ be another such vector. Then $\mathbb{1}(\mathbf{p} - \mathbf{q}) = 0$, $A^s(\mathbf{p} - \mathbf{q}) = \mathbf{p} - \mathbf{q}$, and, by the proposition,

$$\|\mathbf{p} - \mathbf{q}\| = \|A^s(\mathbf{p} - \mathbf{q})\| \le c\|\mathbf{p} - \mathbf{q}\|.$$

Consequently, $\mathbf{p} - \mathbf{q} = 0$ since $0 \le c < 1$.

Let $k \in \mathbb{N}$ be such that $k > s$ and assume that $\mathbb{1}\mathbf{y} = 0$. By the division algorithm, there exist unique integers j and r such that $k = sj + r$ and $r \in \{0, \ldots, s-1\}$. Here $j > (k/s) - 1 > 0$. Now we apply (6) to the matrix A^s and vector $A^r\mathbf{y}$. We obtain

$$\|A^k\mathbf{y}\| = \|(A^s)^j A^r\mathbf{y}\| \le c^j \|A^r\mathbf{y}\|.$$

Consequently, for all $k > s$, we have

$$\|A^k\mathbf{y}\| \le c^{(k/s)-1} \max_{0 \le r < s} \|A^r\mathbf{y}\|. \tag{7}$$

Let $\mathbf{x} \in \mathbb{R}^n$ be such that $\mathbb{1}\mathbf{x} = 1$. Then $\mathbb{1}(\mathbf{x} - \mathbf{q}) = 0$. Substituting $\mathbf{y} = \mathbf{x} - \mathbf{q}$ in (7) yields

$$\left\|A^k\mathbf{x} - \mathbf{q}\right\| = \left\|A^k(\mathbf{x} - \mathbf{q})\right\| \le c^{(k/s)-1} \max_{1 \le r < s} \|A^r(\mathbf{x} - \mathbf{q})\|.$$

Now, since $0 \le c < 1$, we get $A^k\mathbf{x} \to \mathbf{q}$ as $k \to +\infty$. This proves (b) and completes the proof. ∎

The standard theorem about Markov chains is obtained as a corollary to our theorem.

Corollary. *Let P be an $n \times n$ stochastic matrix. Assume that there exists $s \in \mathbb{N}$ such that all entries of P^s are positive. Then*

(a) *there exists a unique probability vector $\mathbf{q} \in \mathbb{R}^n$ such that $P\mathbf{q} = \mathbf{q}$,*

(b) *if $\mathbf{x}$ is a probability vector, then the sequence $\{P^k\mathbf{x}\}$ converges to $\mathbf{q}$ as k tends to $+\infty$.*

Proof. To apply the theorem, we will prove that $\operatorname{var}(P^s) < 1$. For $i, j \in \{1, \ldots, n\}$, denote by b_{ij} the entries of P^s which, by assumption, are positive. Next, notice that for positive numbers a and b we have $|a - b| < a + b$. Therefore, for arbitrary i, j we have

$$\sum_{l=1}^{n} |b_{li} - b_{lj}| < \sum_{l=1}^{n} (b_{li} + b_{lj}) = 2.$$

This proves that the distance between arbitrary columns of P^s is less then 2. Consequently, $\operatorname{diam}(P^s) < 2$, and hence, $\operatorname{var}(P^s) < 1$. Now we apply the theorem for convergence. The proofs of the remaining claims are standard. ∎

The theorem can be restated in terms of the powers of A. This follows from the following equivalency.

Let A be an arbitrary square matrix and let $\mathbf{q}$ be a vector such that $\mathbb{1}\mathbf{q} = 1$. Denote by Q the square matrix, each of whose columns is equal to $\mathbf{q}$, that is, $Q = \mathbf{q}\mathbb{1}$. Then the following two statements are equivalent:

(i) If $\mathbf{x}$ is such that $\mathbb{1}\mathbf{x} = 1$, then the sequence $\{A^k\mathbf{x}\}$ converges to $\mathbf{q}$ as k tends to $+\infty$;

(ii) the powers A^k tend to Q as k tends to $+\infty$.

Assume (i) and let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the vectors of the standard basis of $\mathbb{R}^n$. Then the jth column of A^k is $A^k\mathbf{e}_j$. By (i), $A^k\mathbf{e}_j$ converges to $\mathbf{q}$ as k tends to $+\infty$. This proves (ii).

Now assume (ii) and let $\mathbf{x}$ be a vector with $\mathbb{1}\mathbf{x} = 1$. Then $A^k\mathbf{x}$ converges to $Q\mathbf{x} = (\mathbf{q}\mathbb{1})\mathbf{x} = \mathbf{q}(\mathbb{1}\mathbf{x}) = \mathbf{q}$. This proves (i).

In fact, Q is a projection onto the span of $\mathbf{q}$. To see this, calculate $Q^2 = (\mathbf{q}\mathbb{1})(\mathbf{q}\mathbb{1}) = \mathbf{q}(\mathbb{1}\mathbf{q})\mathbb{1} = \mathbf{q}1\mathbb{1} = \mathbf{q}\mathbb{1} = Q$ and $Q\mathbf{x} = \mathbf{q}(\mathbb{1}\mathbf{x}) = (\mathbb{1}\mathbf{x})\mathbf{q}$ for any $\mathbf{x} \in \mathbb{R}^n$.

5. EXAMPLES. We conclude the article with three examples.

Example 1. The matrix

$$A = \frac{1}{5}\begin{bmatrix} 0 & 2 & -4 \\ -1 & -1 & 0 \\ 6 & 4 & 9 \end{bmatrix}$$

is somewhat stochastic. The largest distance between two columns is between the second and the third, and it equals $12/5$. Therefore, $\operatorname{var} A = 6/5 > 1$. But

$$A^2 = \frac{1}{25}\begin{bmatrix} -26 & -18 & -36 \\ 1 & -1 & 4 \\ 50 & 44 & 57 \end{bmatrix}$$

and $\operatorname{var}(A^2) = 18/25 < 1$. Hence, the theorem applies, and $\mathbf{q} = \frac{1}{3}\begin{bmatrix} -2 & 1 & 8 \end{bmatrix}^\top$.

Example 2. Consider the following three kinds of stochastic matrices:

$$A = \begin{bmatrix} 1 & + & + \\ 0 & + & + \\ 0 & 0 & + \end{bmatrix}, \quad B = \begin{bmatrix} + & + & 0 \\ + & 0 & + \\ 0 & + & + \end{bmatrix}, \quad \text{and} \quad C = \begin{bmatrix} 0 & + & 0 \\ 0 & 0 & 1 \\ 1 & + & 0 \end{bmatrix}.$$

Here we use $+$ for positive numbers.

Since A is upper triangular, all its powers are upper triangular, so no power of A has all positive entries. Thus, the standard theorem does not apply. However, directly from the definition it follows that $\operatorname{var} A < 1$, so the theorem applies. Also, $\mathbf{q} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^\top$.

The matrix B is not positive but $\operatorname{var} B < 1$, so our theorem applies. Also, the standard theorem applies here as well since B^2 is positive.

The first five powers of C are

$$\begin{bmatrix} 0 & + & 0 \\ 0 & 0 & 1 \\ 1 & + & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & + \\ 1 & + & 0 \\ 0 & + & + \end{bmatrix}, \begin{bmatrix} + & + & 0 \\ 0 & + & + \\ + & + & + \end{bmatrix}, \begin{bmatrix} 0 & + & + \\ + & + & + \\ + & + & + \end{bmatrix}, \text{ and } \begin{bmatrix} + & + & + \\ + & + & + \\ + & + & + \end{bmatrix}.$$

The variation of the first two matrices is 1, while $\operatorname{var}(C^3) < 1$. The first positive power of C is C^5.

Example 3. In this example, we consider matrices with complex entries. Let $\omega = (-1 + i\sqrt{3})/2$. Then $1, \omega$, and $\overline{\omega}$ are the cube roots of unity. So, $1 + \omega + \overline{\omega} = 0$, $\overline{\omega}\omega = 1$, $\omega^2 = \overline{\omega}$, and $\overline{\omega}^2 = \omega$.

Consider one vector and two matrices:

$$\mathbf{v} = \begin{bmatrix} 1 \\ \omega \\ \overline{\omega} \end{bmatrix}, \quad A = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \text{and} \quad B = \frac{1}{3}\begin{bmatrix} 1 & \overline{\omega} & \omega \\ \omega & 1 & \overline{\omega} \\ \overline{\omega} & \omega & 1 \end{bmatrix}.$$

We calculate $A\mathbf{v} = \mathbf{0}$, $B\mathbf{v} = \mathbf{v}$, and $\operatorname{var} B = \sqrt{3}/2$. Since $\|B\mathbf{v}\| > \sqrt{3}/2\|\mathbf{v}\|$, the matrix B and the vector $\mathbf{v}$ provide an example showing that the conclusion of the proposition may not hold with complex entries.

A linear combination of A and B shows that the restriction to real numbers cannot be dropped in the theorem. Let γ be a complex number and set $C = A + \gamma B$. Since $A^2 = A$, $AB = BA = 0$ and $B^2 = B$, we have $C^k = A + \gamma^k B$.

The matrix A is stochastic with variation 0, while $\mathbb{1}B = 0$ and $\operatorname{var} B = \sqrt{3}/2$. Hence, $\mathbb{1}C = \mathbb{1}$, that is, C is somewhat stochastic with complex entries. Also,

$$\operatorname{var} C = \operatorname{var}(\gamma B) = |\gamma|\sqrt{3}/2.$$

Therefore, if $1 < |\gamma| < 2/\sqrt{3}$, then $\operatorname{var} C < 1$, but the sequence $\{C^k\}$ diverges, as we can see from the formula for C^k.

Finally, we mention that the vector $\mathbf{v}$ together with the vectors $\mathbf{u} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\top$ and $\mathbf{w} = \begin{bmatrix} 1 & \overline{\omega} & \omega \end{bmatrix}^\top$ form an orthogonal basis for the complex inner product space $\mathbb{C}^3$, that A is the orthogonal projection onto the span of $\mathbf{u}$, and that B is the orthogonal projection onto the span of $\mathbf{v}$.

REFERENCES

1. I. Ipsen, T. Selee, Ergodicity coefficients defined by vector norms. *SIAM J. Matrix Anal. Appl.* **32** (2011) 153–200.
2. A. A. Markov, Extension of the law of large numbers to dependent quantities (in Russian), Izvestiia Fiz.-Matem. Obsch. Kazan Univ., (2nd Ser.), **15** (1906) 135–156; also in [**3**, pp. 339–361]; English translation [**5**, Section 11].
3. A. A. Markov, *Selected Works. Theory of Numbers. Theory of Probability.* (Russian) Izdat. Akad. Nauk SSSR, Leningrad, 1951.
4. E. Seneta, *Markov and the Creation of Markov Chains*. Edited by A. N. Langville and W. J. Stewart. MAM 2006. Markov Anniversary Meeting, Boson Books, Raleigh, NC, 2006, 1–20.
5. *Probability and Statistics. Russian Papers.* Edited by O. B. Sheynin. Selected and translated by Oscar Sheynin. NG Verlag, Berlin, 2004, also item 5 http://www.sheynin.de/download.html.

BRANKO ĆURGUS received his Ph.D. in mathematics in 1985 from the University of Sarajevo in the former Yugoslavia. His Ph.D. research was done under the advisement of Prof. Heinz Langer at the Technical University Dresden in the former German Democratic Republic. Since 1987, he has been enjoying life, teaching, and researching mathematics at Western Washington University in Bellingham, Washington.
Department of Mathematics, Western Washington University, Bellingham, Washington 98225, USA
curgus@wwu.edu

ROBERT I. JEWETT has done research in combinatorics and analysis. He taught mathematics for many years at Western Washington University. He is now retired and lives in Bellingham.
Department of Mathematics, Western Washington University, Bellingham, Washington 98225, USA

NOTES

Edited by **Sergei Tabachnikov**

Stereographic Trigonometric Identities

Michael Hardy

Abstract. We show that trigonometric identities arising from the most well known alternative to the arc-length parametrization of the circle share some of the same elaborate nature as the more familiar identities involving sines, tangents, etc.

Some trigonometric identities, for example $\sin^2\alpha + \cos^2\alpha = 1$, do not depend on the fact that the circle is parametrized by arc length and would be equally valid with other parametrizations. For others, such as $\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$, the parametrization is essential. There exists an immense variety of what we know as trigonometric identities. Can anything comparable exist for any other parametrizations of the circle? This paper is a step toward an affirmative answer to that question.

We will work with the most well-known alternative parametrization, the one used in the tangent half-angle substitution used for antidifferentiating rational functions of sine and cosine. The substitution is

$$\begin{aligned} \tan\frac{\alpha}{2} &= y, \\ \cos\alpha &= \frac{1-y^2}{1+y^2}, \\ \sin\alpha &= \frac{2y}{1+y^2}, \quad \text{and} \\ d\alpha &= \frac{2\,dy}{1+y^2}. \end{aligned}$$

Geometrically, the relationship between y and the point $(c(y), s(y))$ is as follows. Draw a line passing through the point $(-1, 0)$ on the unit circle $x^2 + y^2 = 1$ and the point $(0, y)$ on the y-axis. That line intersects the unit circle at $(c(y), s(y))$. If we regard the coordinate y as lying in the space $\mathbb{R} \cup \{\infty\}$, with a single ∞ at both ends of the line rather than $\pm\infty$ at opposite ends, then the line is topologically a circle, and the mapping $y \mapsto (c(y), s(y))$ is a homeomorphism, taking 0 to $(1, 0)$, $\pm\pi/2$ to $(0, \pm 1)$, and ∞ to $(-1, 0)$.

Thus, the mapping $(0, y) \mapsto (c(y), s(y))$ is the inverse of a stereographic projection.

http://dx.doi.org/10.4169/amer.math.monthly.122.01.43
MSC: Primary 33B10

Definition (Stereographic trigonometric functions). The **stereographic sine** $\operatorname{ss} y$ and **stereographic cosine** $\operatorname{cs} y$ are continuous functions of $y \in \mathbb{R} \cup \{\infty\}$ given by

$$\operatorname{ss} y = \frac{2y}{1+y^2}, \quad \text{and}$$
$$\operatorname{cs} y = \frac{1-y^2}{1+y^2}.$$

As far as the author knows, these definitions are new. An anonymous referee has pointed out that the use in sperical trigonometry of stereographic projections in one higher dimension—involving the 2-spere in 3-space—is the topic of a chapter in a recent book [**1**].

Now we can prove some stereographic trigonometric identities. Our first ones are routine and probably known to everybody.

Proposition 1. *If* $y \in \mathbb{R} \cup \{\infty\}$, *then*

$$\operatorname{cs}(-y) = \operatorname{cs}(y),$$
$$\operatorname{ss}(-y) = -\operatorname{ss}(y),$$
$$\operatorname{cs}(1/y) = -\operatorname{cs}(y),$$
$$\operatorname{ss}(1/y) = \operatorname{ss}(y).$$

To facilitate our statement of some somewhat more involved identities, we first introduce some notation.

Definition. For variables $y_1, y_2, y_3, \ldots$, let $e_{k,n}(\operatorname{cs} y)$ be the kth-degree elementary symmetric polynomial in the first n variables $\operatorname{cs} y_1, \ldots, \operatorname{cs} y_n$. In particular $e_{0,n}(\operatorname{cs} y)$ is an empty product equal to 1 and $e_{k,n}(\operatorname{cs} y)$ is an empty sum equal to 0 if $k > n$. Similarly, for variables $c_1, c_2, c_3, \ldots$, we let $e_{k,n}(c)$ be the kth-degree elementary symmetric polynomial in $c_1, \ldots, c_n$ and similarly with x in place of c. For $\alpha_1, \alpha_2, \alpha_2, \ldots$, we let $e_{k,n}(\tan \alpha)$ be the kth-degree elementary symmetric polynomial in $\tan \alpha_1, \ldots, \tan \alpha_n$.

Proposition 2. *Let* $y_1, \ldots, y_n \in \mathbb{R} \cup \{\infty\}$.

$$\operatorname{cs}(y_1 \cdots y_n) = \frac{e_{1,n}(\operatorname{cs} y) + e_{3,n}(\operatorname{cs} y) + e_{5,n}(\operatorname{cs} y) + \cdots}{e_{0,n}(\operatorname{cs} y) + e_{2,n}(\operatorname{cs} y) + e_{4,n}(\operatorname{cs} y) + \cdots} \tag{1}$$
$$\operatorname{ss}(y_1 \cdots y_n) = \frac{\operatorname{ss} y_1 \cdots \operatorname{ss} y_n}{e_{0,n}(\operatorname{cs} y) + e_{2,n}(\operatorname{cs} y) + e_{4,n}(\operatorname{cs} y) + \cdots} \tag{2}$$

A proof will appear below.

A Pythagorean identity follows:

$$\Big(e_{1,n}(\operatorname{cs} y) + e_{3,n}(\operatorname{cs} y) + e_{5,n}(\operatorname{cs} y) + \cdots\Big)^2 + (\operatorname{ss} y_1)^2 \cdots (\operatorname{ss} y_n)^2$$
$$= \Big(e_{0,n}(\operatorname{cs} y) + e_{2,n}(\operatorname{cs} y) + e_{4,n}(\operatorname{cs} y) + \cdots\Big)^2.$$

In the case $n = 1$, this says $(\operatorname{cs} y)^2 + (\operatorname{ss} y)^2 = 1$. This Pythagorean identity holds just as much for the conventional sine and cosine as for the stereographic sine and cosine, but the author is not aware of any prior appearance of it. Indeed, if $e_k(c)$ is the elementary symmetric polynomial in any complex numbers $c_1, \ldots, c_n$, then

$$\Big(e_{1,n}(c) + e_{3,n}(c) + e_{5,n}(c) + \cdots\Big)^2 + (1 - c_1^2) \cdots (1 - c_n^2)$$
$$= \Big(e_{0,n}(c) + e_{2,n}(c) + e_{4,n}(c) + \cdots\Big)^2.$$

Parallels exist between Proposition 2 and some identities involving conventional trigonometric functions.

Proposition 3.

$$\tan(\alpha_1 + \cdots + \alpha_n) = \frac{e_{1,n}(\tan\alpha) - e_{3,n}(\tan\alpha) + e_{5,n}(\tan\alpha) - \cdots}{e_{0,n}(\tan\alpha) - e_{2,n}(\tan\alpha) + e_{4,n}(\tan\alpha) - \cdots},$$
$$\sec(\alpha_1 + \cdots + \alpha_n) = \frac{\sec\alpha_1 \cdots \sec\alpha_n}{e_{0,n}(\tan\alpha) - e_{2,n}(\tan\alpha) + e_{4,n}(\tan\alpha) - \cdots}.$$

The signs alternate in Proposition 3 but not in Proposition 2.

These also entail a Pythagorean identity:

$$\Big(e_{0,n}(\tan\alpha) - e_{2,n}(\tan\alpha) + e_{4,n}(\tan\alpha) - \cdots\Big)^2$$
$$+ \Big(e_{1,n}(\tan\alpha) - e_{3,n}(\tan\alpha) + e_{5,n}(\tan\alpha) - \cdots\Big)^2 = \sec^2\alpha_1 \cdots \sec^2\alpha_n.$$

In the case $n = 1$, this says $1 + \tan^2\alpha = \sec^2\alpha$. This identity also does not depend on the operation of addition of angles or the choice of parametrization, but this identity entirely escaped the author's notice until everything above was done.

One way to prove Proposition 3 is by two straightforward mathematical inductions.

Proof of the finite case of Proposition 2. For the present proof, assume $y_1, \ldots, y_n \notin \{\infty\}$. We need to prove (1) and (2). Notice that

$$y^2 = \frac{1 - \operatorname{cs} y}{1 + \operatorname{cs} y} \quad \text{and} \quad y(1 + \operatorname{cs} y) = \operatorname{ss} y.$$

Then

$$\operatorname{cs}(y_1 \cdots y_n) = \frac{1 - y_1^2 \cdots y_n^2}{1 + y_1^2 \cdots y_n^2} = \frac{1 - \left(\frac{1 - \operatorname{cs} y_1}{1 + \operatorname{cs} y_1}\right) \cdots \left(\frac{1 - \operatorname{cs} y_n}{1 + \operatorname{cs} y_n}\right)}{1 + \left(\frac{1 - \operatorname{cs} y_1}{1 + \operatorname{cs} y_1}\right) \cdots \left(\frac{1 - \operatorname{cs} y_n}{1 + \operatorname{cs} y_n}\right)}$$
$$= \frac{(1 + \operatorname{cs} y_1) \cdots (1 + \operatorname{cs} y_n) - (1 - \operatorname{cs} y_1) \cdots (1 - \operatorname{cs} y_n)}{(1 + \operatorname{cs} y_1) \cdots (1 + \operatorname{cs} y_n) + (1 - \operatorname{cs} y_1) \cdots (1 - \operatorname{cs} y_n)}$$
$$= \frac{e_{1,n}(\operatorname{cs} y) + e_{3,n}(\operatorname{cs} y) + e_{5,n}(\operatorname{cs} y) + \cdots}{e_{0,n}(\operatorname{cs} y) + e_{2,n}(\operatorname{cs} y) + e_{4,n}(\operatorname{cs} y) + \cdots}.$$

This proves (1). Next, we have

$$\begin{aligned}\operatorname{ss}(y_1\cdots y_n) &= \frac{2y_1\cdots y_n}{1+y_1^2\cdots y_n^2} = \frac{2y_1\cdots y_n}{1+\left(\dfrac{1-\operatorname{cs} y_1}{1+\operatorname{cs} y_1}\right)\cdots\left(\dfrac{1-\operatorname{cs} y_n}{1+\operatorname{cs} y_n}\right)}\\ &= \frac{2y_1\cdots y_n(1+\operatorname{cs} y_1)\cdots(1+\operatorname{cs} y_n)}{(1+\operatorname{cs} y_1)\cdots(1+\operatorname{cs} y_n)+(1-\operatorname{cs} y_1)\cdots(1-\operatorname{cs} y_n)}\\ &= \frac{2\operatorname{ss} y_1\cdots\operatorname{ss} y_n}{2(e_{0,n}(\operatorname{cs} y)+e_{2,n}(\operatorname{cs} y)+e_{4,n}(\operatorname{cs} y)+\cdots)}.\end{aligned}$$

This proves (2). ∎

The following definition will be used to state a terser and perhaps more enlightening version of Proposition 2 and to prove Proposition 2 in the case where one or more $y_1, \ldots, y_n$ is ∞.

Definition. Let

$$x_1 \circ x_2 = \frac{x_1+x_2}{1+x_1x_2}.$$

Lemma. *The operation "$\circ$" is associative and*

$$x_1 \circ \cdots \circ x_n = \frac{e_{1,n}(x)+e_{3,n}(x)+e_{5,n}(x)+\cdots}{e_{0,n}(x)+e_{2,n}(x)+e_{4,n}(x)+\cdots}. \tag{3}$$

The numbers ± 1 are absorbing elements for this operation: $1 \circ x = 1$ for all values of x except -1 and $(-1) \circ x = -1$ for all values of x except 1. The number 0 is an identity element, i.e., $0 \circ x = x$. The inverse of x is $-x$, i.e., $x \circ (-x) = 0$.

Proof. We have

$$(x_1 \circ x_2) \circ x_3 = \frac{\left(\dfrac{x_1+x_2}{1+x_1x_2}\right)+x_3}{1+\left(\dfrac{x_1+x_2}{1+x_1x_2}\right)x_3} = \frac{x_1+x_2+x_3+x_1x_2x_3}{1+x_1x_2+x_1x_3+x_2x_3}.$$

This is symmetric in x_1, x_2, x_3 and associativity follows. Now proceed by induction on n:

$$\begin{aligned}&(x_1 \circ \cdots \circ x_n) \circ x_{n+1}\\ &= \frac{\left(\dfrac{e_{1,n}(x)+e_{3,n}(x)+e_{5,n}(x)+\cdots}{e_{0,n}(x)+e_{2,n}(x)+e_{4,n}(x)+\cdots}\right)+x_{n+1}}{1+\left(\dfrac{e_{1,n}(x)+e_{3,n}(x)+e_{5,n}(x)+\cdots}{e_{0,n}(x)+e_{2,n}(x)+e_{4,n}(x)+\cdots}\right)x_{n+1}}\\ &= \frac{\Big(e_{1,n}(x)+e_{3,n}(x)+e_{5,n}(x)+\cdots\Big)+\Big(e_{0,n}(x)+e_{2,n}(x)+e_{4,n}(x)+\cdots\Big)x_{n+1}}{\Big(e_{0,n}(x)+e_{2,n}(x)+e_{4,n}(x)+\cdots\Big)+\Big(e_{1,n}(x)+e_{3,n}(x)+e_{5,n}(x)+\cdots\Big)x_{n+1}}\\ &= \frac{e_{1,n+1}(x)+e_{3,n+1}(x)+e_{5,n+1}(x)+\cdots}{e_{0,n+1}(x)+e_{2,n+1}(x)+e_{4,n+1}(x)+\cdots}.\end{aligned}$$

The proofs of the other assertions in the lemma are trivial. ∎

Proposition 2 (Terse version.). *If* $y_1, \ldots, y_n \in \mathbb{R} \cup \{\infty\}$, *then*

$$\operatorname{cs}(y_1 \cdots y_n) = \operatorname{cs}(y_1) \circ \cdots \circ \operatorname{cs}(y_n).$$

This means that cs restricted to $[0, \infty]$ with multiplication is an isomorphism onto $[-1, 1]$ with "$\circ$".

Proof of the "infinite" case of Proposition 2. If $y_1 = \infty$, then

$$\operatorname{cs}(y_1 y_2) = \operatorname{cs}(\infty \cdot y_2) = -1 = \frac{-1 + \operatorname{cs} y_2}{1 + (-1) \operatorname{cs} y_2} = \frac{\operatorname{cs} y_1 + \operatorname{cs} y_2}{1 + \operatorname{cs} y_1 \operatorname{cs} y_2}.$$

We then use associativity. In a similar manner,

$$\operatorname{ss}(y_1 y_2) = \operatorname{ss}(\infty \cdot y_2) = 0 = \frac{0 \cdot y_2}{1 + (-1) \cdot \operatorname{cs} y_2} = \frac{\operatorname{ss} y_1 \operatorname{ss} y_2}{1 + \operatorname{cs} y_1 \cdot \operatorname{cs} y_2}.$$

■

Exercise. Let $\operatorname{ts} = \operatorname{ss} / \operatorname{cs}$ be the "stereographic tangent" function. Show that

$$\operatorname{ts}(y^n) = \frac{2(\operatorname{ss} y)^n}{(1 + \operatorname{cs} y)^n - (1 - \operatorname{cs} y)^n}.$$

REFERENCES

1. G. Van Brummelen, *Heavenly Mathematics: The Forgotten Art of Spherical Trigonometry*. Princeton Univ. Press. Princeton, NJ 2013.

MICHAEL HARDY received his Ph.D. in statistics with a minor in mathematics from the University of Minnesota in 1997 after completing nearly all coursework for the Ph.D. in mathematics and then switching departments. He has taught at a number of postsecondary institutions, among them the University of North Carolina, the Woods Hole Oceanographic Institution, and MIT.
Department of Mathematics and Statistics, St. Cloud State University, St. Cloud, MN 56301
drmichaelhardy@gmail.com

Infinitely Many Primes in the Arithmetic Progression $kn - 1$

Xianzu Lin

Abstract. In this paper we give a simple and elementary proof of the infinitude of primes in the arithmetic progression $kn - 1, n > 0$.

1. INTRODUCTION. Dirichlet theorem about primes in the arithmetic progression states that, if k, l are coprime positive integers, then there are infinitely many primes in the arithmetic progression $kn + l$, for $n > 0$. Dirichlet first proved this theorem by mean of L-functions and analysis. For some elementary but complicated proofs, see [**5, 6**]. On the other hand, there are many simple proofs of some special cases. For example, in [**2**] we can find proofs for the special cases $4n + 3$ and $6n + 5$, and in [**1**] Bateman and Low proved the case $24n + l$. All these proofs, called Euclidean proofs, are generalizations of Euclid's proof of the infinitude of primes (circa 300 BC). We say that a prime p is a prime divisor of a polynomial $f \in \mathbb{Z}[x]$ if $p|f(n)$ for some $n \in \mathbb{Z}$. The main point in an Euclidean proof is to find explicitly a polynomial with infinitely many prime divisors belonging to the arithmetic progression $kn + l$. For example, in the case of $4n + 3$, $4x^2 + 3$ is the required polynomial. Using Euclid's approach, Schur first proved that if $l^2 \equiv 1(mod\ k)$, then there exist infinitely many primes in the arithmetic progression $kn + l$. But his proof requires the assumption of existence of at least one prime p congruent to $l(mod\ k)$ satisfying $p > \varphi(k)/2$, where $\varphi(k)$ is the number of positive integers not greater than and prime to k. Later Murty [**3**] showed that the condition $l^2 \equiv 1(mod\ k)$ is also necessary for an Euclidean proof.

Theorem 1.1. *(Schur, Murty) There exists a Euclidean proof for the arithmetic progression $kn + l$ if and only if $l^2 \equiv 1(mod\ k)$.*

Theorem 1.1 tells us to what extent can we apply Euclidean proofs. But their proofs are not elementary. For example, the proof in [**3**] used Galois extension theory and assumed that there is at least one prime in the arithmetic progression $kn + l$. In [**8**] the case $kn + 1$ was proved using an elementary method. In this paper, we give a similar proof for the case $kn - 1$.

Theorem 1.2. *If k is a positive integer, then there are infinitely many primes in the arithmetic progression $kn - 1$, for $n > 0$.*

2. THE PROOF OF THEOREM 1.2. For each positive integer k, define polynomials $f_k(x), g_k(x) \in \mathbb{Z}[x]$ by

$$(1 + ix)^k = f_k(x) + g_k(x)i,$$

where $i = \sqrt{-1}$. It is easy to see that $g_{2k}(x)$ is of degree $2k - 1$. As $\mathbb{Z}[i][x]$ is a unique factorization domain, we deduce that the greatest common divisor $(f_k, g_k) = 1$ in $\mathbb{Q}[x]$.

http://dx.doi.org/10.4169/amer.math.monthly.122.01.48
MSC: Primary 11A41

Lemma 2.1. *Let d be a proper divisor of k, then $g_k(x) = g_d(x)g_{d,k}(x)$ for some $g_{d,k}(x) \in \mathbb{Z}[x]$, and $(g_d, g_{d,k}) = 1$. Moreover, there exists $s(x), t(x) \in \mathbb{Z}[x]$ and a positive integer $n_{d,k}$ such that $s(x)g_d(x) + t(x)g_{d,k}(x) = n_{d,k}$.*

Proof. Let $k = dm$. By definition

$$(f_d(x) + g_d(x)i)^m = f_k(x) + g_k(x)i.$$

Hence

$$g_k(x) = mf_d^{m-1}(x)g_d(x) + g_d^2(x)h(x)$$

for some $h(x) \in \mathbb{Z}[x]$. Set

$$g_{d,k}(x) = mf_d^{m-1}(x) + g_d(x)h(x).$$

Then

$$g_k(x) = g_d(x)g_{d,k}(x)$$

and

$$(g_d, g_{d,k}) = \left(g_d, mf_d^{m-1} + g_d h\right) = \left(g_d, mf_d^{m-1}\right).$$

As $(f_d, g_d) = 1$, we have $(g_d, g_{d,k}) = 1$. This proves the first assertion. The second assertion follows directly from the first. ∎

The roots of $g_{2k}(x)$ are

$$\tan\frac{a\pi}{2k}, \quad \text{for } a = -k+1, -k+2, \ldots, k-1.$$

When a is prime to $2k$, $\tan\frac{a\pi}{2k}$ is not a root of $g_d(x)$ for any proper divisor d of $2k$. Set

$$\Phi_k(x) = b_k \prod_{(a,2k)=1} \left(x - \tan\frac{a\pi}{2k}\right),$$

where b_k is an positive integer. Then it is easy to see that $\Phi_k(x)$ is a greatest common divisor of $\{g_{d,2k}(x)\}$, where d runs over all proper divisors of $2k$. We choose b_k such that $\Phi_k(x)$ is primitive. Hence $g_{d,2k}(x)$ is divisible by $\Phi_k(x)$ for any proper divisor d of $2k$.

Lemma 2.2. *For k large enough, there is a prime p of the form $4a + 3$ and an integer n such that $p > n_{k,2k}$, and $p|\Phi_k(n)$.*

Proof. For k large enough, we can choose a positive integer m between the largest root and the second largest root of $\Phi_k(x)$. Then $\Phi_k(m) = h < 0$. Replacing x by $x + m$, we get

$$\Phi_k(x+m) = a_r x^r + a_{r-1}x^{r-1} + \cdots - a_0,$$

where $a_r > 0$ and $a_0 = -h > 0$. Now suppose $b > 2\max(-h, n_{k,2k})$, then

$$\Phi_k(b! + m) = a_r(b!)^r + a_{r-1}(b!)^{r-1} + \cdots - a_0 = a_0\left(a_r\frac{(b!)^r}{a_0} + a_{r-1}\frac{(b!)^{r-1}}{a_0} + \cdots - 1\right).$$

One checks that

$$a_r \frac{(b!)^r}{a_0} + a_{r-1} \frac{(b!)^{r-1}}{a_0} + \cdots - 1$$

is a positive integer of the form $4a + 3$. Hence it has a prime divisor p of the same form. Obviously $p > n_{k,2k}$. ■

We are in a position now to prove Theorem 1.2. We can replace k by $2k$ and we only have to find one prime p of the form $2kn - 1$, for replacing $2k$ by $2pk$ we can find another prime and so on. In order to apply Lemma 2.2, we assume that k is large enough (multiplying k by a large integer).

Choose a prime p and an integer n as in Lemma 2.2. Then $p|g_{d,2k}(n)$ for any proper divisor d of $2k$. By Lemma 2.1, we see that $g_d(n)$ cannot be divisible by p (otherwise we would have $p|n_{k,2k}$). Consider $1 + ni$ as an element in the finite field $\mathbb{F}_p[i]$ (that is why we require p to be of the form $4n + 3$, otherwise $\mathbb{F}_p[i]$ will be $\mathbb{F}_p$). As the units of $\mathbb{F}_p[i]$ form a cyclic group of order $p^2 - 1$, and $\mathbb{F}_p^*$ is a subgroup of order $p - 1$, the above statement implies that $2k$ is the least positive integer satisfying $(1 + ni)^{2k} \in \mathbb{F}^*$. It follows that $2k$ must be a divisor of $p + 1$, i.e., p is of the form $2kn - 1$.

Remark 2.3. It was pointed out by the referee that the same result has been proved in [**4**]. In fact, the proof in [**4**] used $(x + i)^k$ instead of $(1 + ix)^k$. But these two proofs differ in some respects.

Remark 2.4. By Theorem 1.1, even in the simplest cases $5n \pm 2$, a Euclidean proof does not exist. But, as the proof in [**8**], our proof is not Euclidean, for infinitely many polynomials $\Phi_k(x)$, for $k \to \infty$, are involved for the search of large primes of the form $kn - 1$. We wonder whether there exists similar proofs for the cases $5n \pm 2$.

ACKNOWLEDGMENT. This paper is supported by the National Natural Science Foundation of China, Tian Yuan Special Foundation (No. 11226083). The author would like to express his deep gratitude to the referee for so many comments and suggestions about this paper.

REFERENCES

1. P. T. Bateman, M. E. Low, Prime numbers in arithmetic progressions with difference 24, *Amer. Math. Monthly* **72** (1965) 139–143, http://dx.doi.org/10.2307/2310975.
2. G. H. Hardy, E. M. Wright, *An Introduction to the Theory of Numbers*. Fourth edition. Clarendon Press, Oxford, 1960, https://archive.org/details/AnIntroductionToTheTheoryOfNumbers-4thEd-G.h.HardyE.m.Wright.
3. M. Ram Murty, N. Thain, Primes in certain arithmetic progressions, *Funct. Approx. Comment. Math.* **35** (2006) 249–259, http://dx.doi.org/10.7169/facm/1229442627.
4. T. Nagell, *Introduction to Number Theory*. Second reprint edition. Chelsea Publishing, Providence, RI, 2001.
5. A. Selberg, An elementary proof of Dirichlet's theorem about primes in an arithmetic progression, *Ann. Math.* **50** no. 2 (1949) 297–304.
6. H. N. Shapiro, On primes in arithmetic progression. II., *Ann. Math.* **52** no. 2 (1950) 231–243.
7. I. Schur, Über die Existenz unendlich vieler Primzahlen in einiger speziellen arithmetischen Progressionen, *S-B Berlin. Math. Ges.* **11** (1912) 40–50.
8. E. Wendt, Elementarer Beweis des Satzes, dass in jeder unbegrenzten arithmetischen Progression $my + 1$ unendlich viele Primzahlen vorkommen, *J. Reine Angew. Math.* **115** (1895) 85–88.

College of Mathematics and Computer Science, Fujian Normal University,
Fuzhou 350108, China
linxianzu@126.com

Deranged Matchings: Enumeration by Integration!!

Most readers are aware that the number d_n of derangements of an n-set is about $n!/e$; we consider objects perhaps less familiar but no less interesting. A *deranged matching* in a complete graph K_{2n}, relative to a fixed perfect matching M, is a perfect matching in the complement $K_{2n} - M$. Using a theorem of Godsil [**1**], we novelly redetermine the number D_n of deranged matchings in K_{2n}. Godsil's formula expresses the number $\Xi(G)$ of perfect matchings in a graph G as an integral involving the 'matching polynomial' $\alpha_{\overline{G}}(x)$ of G's complement $\overline{G}$: viz. $\Xi(G) = \int_{-\infty}^{\infty} \alpha_{\overline{G}}(x) e^{-x^2/2}\, dx/\sqrt{2\pi}$. If $\overline{G}$ has order m and $\mu_{\overline{G}}(k)$ counts the k-edge matchings in $\overline{G}$, then $\alpha_{\overline{G}}(x) := \sum_{k=0}^{\lfloor m/2 \rfloor} (-1)^k \mu_{\overline{G}}(k) x^{m-2k}$. With $G = K_{2n} - M$, we have $\mu_{\overline{G}}(k) = \binom{n}{k}$ because each choice of k edges from the n-edge matching M is a k-edge matching. This yields

$$D_n = \Xi(G) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2n-2k} e^{-x^2/2}\, dx \right].$$

The bracketed factor—a Gaussian moment—is $(2n - 2k - 1)!!$, so we arrive at

$$D_n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (2n - 2k - 1)!! \tag{1}$$

To sate any curiosity about the asymptotics of D_n, rewrite (1) as

$$\frac{D_n}{(2n-1)!!} = \sum_{k=0}^{n} (-1)^k \frac{(n)_k}{k!((2n-1))_k}, \tag{2}$$

where $((\cdot))_k$ denotes the falling double factorial. The k^{th} term in (2) is $[(-1)^k/k!] \cdot \prod_{i=1}^{k} (1 - (i-1)/n)/(2 - (2i-1)/n)$, with limit $(-1)^k/k!2^k$ as $n \to \infty$. This implies that $D_n/(2n-1)!! \to 1/\sqrt{e}$, giving the limiting proportion of the perfect matchings of K_{2n} that are deranged. This beautiful analogue of the derangements phenomenon (that $d_n/n! \to 1/e$) was conjectured in 2000 (by J.N. Brawner) and originally proved in 2001 (by B.H. Margolius).

REFERENCES

1. C.D. Godsil, Hermite polynomials and a duality relation for matching polynomials, *Combinatorica* **1** no. 3 (1981) 257–262.

—Submitted by P. Mark Kayll, Missoula, Montana

This work was partially supported by a grant from the Simons Foundation (#279367 to Mark Kayll).

http://dx.doi.org/10.4169/amer.math.monthly.122.01.51
MSC: Primary 05C30, Secondary 05C70; 05C31

A Point of Tangency Between Combinatorics and Differential Geometry

Francis C. Motta, Patrick D. Shipman, and Bethany Springer

Abstract. Edges of de Bruijn graphs, whose labeled vertices are arranged in sequential order on a circle, envelop epicycloids.

1. DE BRUIJN SEQUENCES AND GRAPHS.

The binary word

$$1011100010 \tag{1}$$

contains each of the eight binary words 000, 001, 010, 011, 100, 101, 110, and 111 of length 3 as a subword exactly once. Such a word is called a *de Bruijn sequence* since de Bruijn [**1**] (and others [**2**]) proved that for any natural numbers k and n, there is a sequence on k symbols (of length $k^n + n - 1$) whose k^n length-n subwords are exactly the k^n words of length n on k symbols.

De Bruijn's proof relies on finding paths in *de Bruijn* (k, n) *graphs*, which are directed graphs whose vertices represent the subwords of length n on k symbols and whose edges connect potential consecutive subwords in a sequence on k symbols. For example, the vertices in the de Bruijn $(2, 2)$ graph of Figure 1(a) are labeled by the four possible length-two words on the two symbols 0, 1. In a binary sequence, the subword 10 may be followed by either the subword 00 (to form the subword 100) or the subword 01 (to form the subword 101). These possibilities are recorded in the graph by the directed edges from vertex 10 to vertex 00 (labeled 100) and from vertex 10 to vertex 01 (labeled 101).

A path in a graph that includes each vertex exactly once is called a *Hamiltonian path*, and a path that includes each edge exactly once is called an *Eulerian path*. One forms the de Bruijn sequence 10011, which contains all words of length two exactly once, by following the Hamiltonian path $10 \to 00 \to 01 \to 11$ in the de Bruijn $(2, 2)$ graph. By focusing on edges rather than vertices, the same graph may be used to form a de Bruijn sequence that contains all subwords of length 3 exactly once. For example, we form the de Bruijn sequence (1) by following in Figure 1(a) the Eulerian path $10 \overset{101}{\to} 01 \overset{011}{\to} 11 \overset{111}{\to} 11 \overset{110}{\to} 10 \overset{100}{\to} 00 \overset{000}{\to} 00 \overset{001}{\to} 01 \overset{010}{\to} 10$. Alternatively, sequence (1) arises from following the Hamiltonian path $101 \to 011 \to 111 \to 110 \to 100 \to 000 \to 001 \to 010$ in the de Bruijn (2,3) graph of Figure 1(c).

We denote a (sub)word of length n on the k symbols $0, 1, 2, \ldots, k-1$ by $w_j^{(k,n)}$ if it is the base-k representation of the integer j. For example, $w_2^{(2,2)} = 10$ followed by $w_1^{(2,2)} = 01$ are the first two subwords of (1) of length 2 and form $w_5^{(2,3)} = 101$, the first subword of length 3. In a word on k symbols, a subword $w_j^{(k,n)} = v_1 v_2 \ldots v_n$ ($v_i \in \{0, 1, \ldots, k-1\}$) may be followed by the subwords $w_{\gamma(j,r)}^{(k,n)} = v_2 \ldots v_n r$, where

http://dx.doi.org/10.4169/amer.math.monthly.122.01.52
MSC: Primary 35A04, Secondary 05C20; 05C48

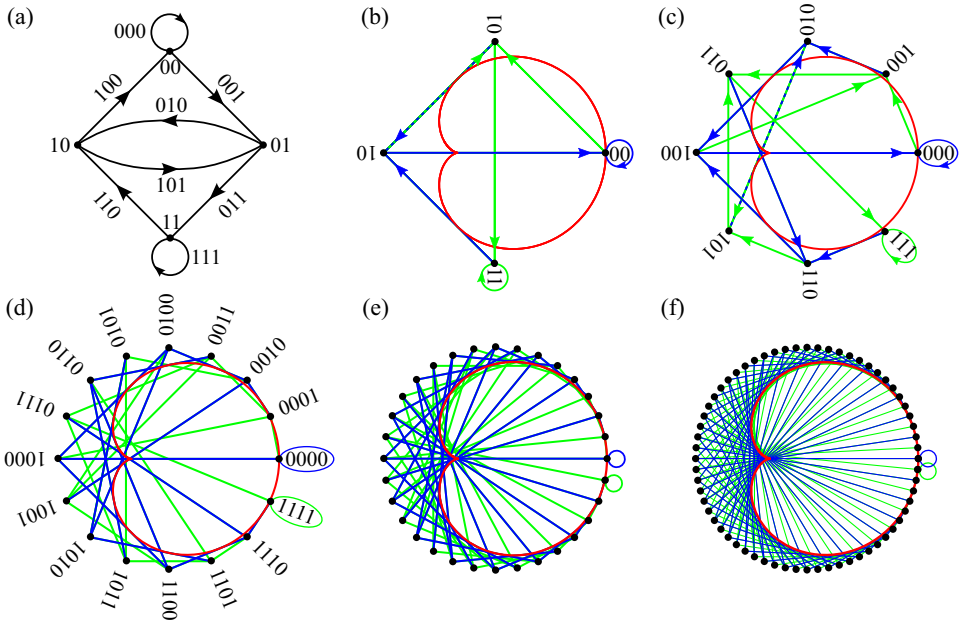


Figure 1. de Bruijn (k, n) graphs for $k = 2$ and (a,b) $n = 2$, (c) $n = 3$, (d) $n = 4$, (e) $n = 5$, and (f) $n = 6$. (a) and (b) are the same graph, (a) drawn in the manner of de Bruijn [**1**]. For simplicity, arrows are omitted in (d–f), and labels in (e,f). Also plotted (in red) in (b–f) is the cardioid given by (2) with $k = 2$.

$r \in \{0, 1, \ldots, k-1\}$, and $\gamma(j, r) = kj + r \bmod k^n$. This is recorded in a de Bruijn (k, n) graph by a directed edge from the vertex marked $w_j^{(k,n)}$ to each vertex $w_{\gamma(j,r)}^{(k,n)}$. In the de Bruijn $(2, n)$ graphs of Figure 1, edges $w_j^{(2,n)} \to w_{\gamma(j,0)}^{(2,n)}$ and edges $w_j^{(2,n)} \to w_{\gamma(j,1)}^{(2,n)}$ are respectively shown in blue and green. Some de Bruijn (k, n) graphs for $k > 2$ are shown in Figure 2.

A de Bruijn sequence that includes each word of length n on k symbols corresponds to a Hamiltonian path in a de Bruijn (k, n) graph or, alternatively, to an Eulerian path in a de Bruijn $(k, n-1)$ graph since the edges of the $(k, n-1)$ graph correspond exactly to vertices of the (k, n) graph. Every de Bruijn graph has an Eulerian path since it is connected and each vertex has equal indegree and outdegree (namely equal to k).

Figure 1(a) follows de Bruijn's drawing. In all other examples of Figures 1 and 2, we have positioned the k^n vertices of a (k, n)—de Bruijn graph at angles $\theta_j = j2\pi/k^n$ on a circle and labeled them by the words $w_j^{(k,n)}$. Strikingly, edges of de Bruijn graphs as drawn in Figures 1(b–f) and Figure 2 envelop planar curves. Examining the sequence of graphs in Figure 1, we notice that edges $w_j^{(2,n)} \to w_{\gamma(j,0)}^{(2,n)}$ appear to be tangent to a cardioid. For small n, edges $w_j^{(2,n)} \to w_{\gamma(j,1)}^{(2,n)}$ are, in general, not tangent to cardioids but apparently approach tangency as n increases. Simple facts about epicycloids allow us to make these observations precise.

2. ASYMPTOTIC TANGENCY OF DE BRUIJN EDGES TO EPICYCLOIDS.

Epicycloids are planar curves traced by the path of a specified point on a circle as it rolls (without slipping) around another circle that is stationary, as illustrated in Figure 3(a). Normalize the stationary circle to radius 1, centered at the origin of the complex plane and let the rolling circle (with radius $1/(k-1)$, for some $k > 1$) start centered at $1 + 1/(k-1)$ on the real axis. The path $c(\theta)$ of the specified point chosen to start at $c(0) = ((k+1)/(k-1))e^{0i}$ is given by

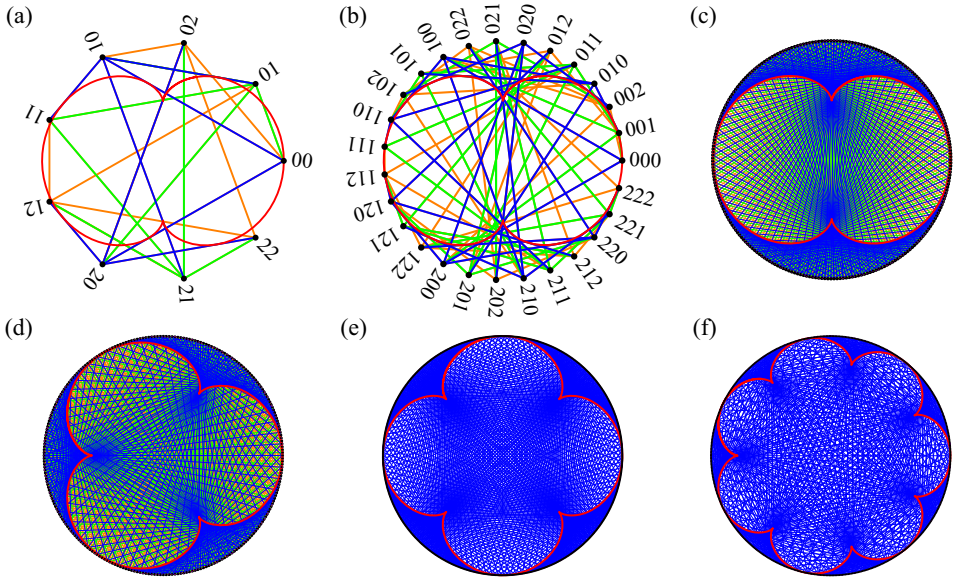


Figure 2. de Bruijn (k, n) graphs (a) $(k, n) = (3, 2)$, (b) $(k, n) = (3, 3)$, (c) $(k, n) = (3, 5)$, (d) $(k, n) = (4, 4)$, (e) $(k, n) = (5, 4)$, and (f) $(k, n) = (8, 3)$. In (a–c), edges $w_j^{(3,n)} \to w_{\gamma(j,r)}^{(3,n)}$ are shown in blue ($r = 0$), green ($r = 1$), and orange ($r = 2$). Arrows on edges are omitted, as well as labels in (c–f). In (e,f), only the edges $w_j^{(k,n)} \to w_{\gamma(j,0)}^{(k,n)}$ are plotted. The red curve in each panel is the epicycloid (2) for the corresponding k.

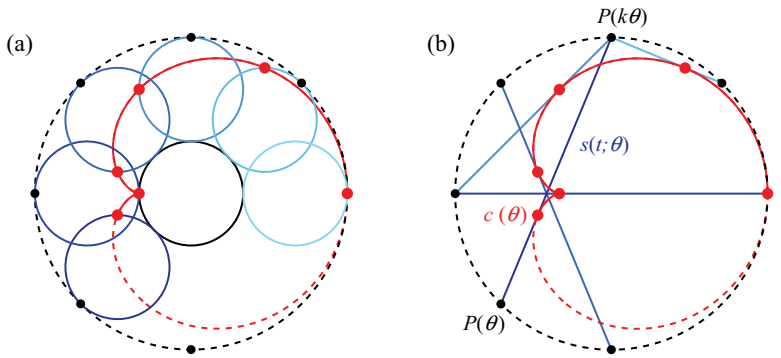


Figure 3. (a) Illustration of the generation of the epicycloid (2) with $k = 2$ as the trace of a point on a circle of radius $1/(k - 1)$ rolling about a circle of radius 1. (b) Illustration of the tangency of the chord (3) to the epicycloid (2) with $k = 2$.

$$c(\theta) = \frac{k\mathrm{e}^{\mathrm{i}\theta} + \mathrm{e}^{\mathrm{i}k\theta}}{k - 1}, \tag{2}$$

for $\theta \in [0, 2\pi)$, the angle from the real axis to the point of tangency of the two circles [**3**].

Let $P(\theta) = ((k + 1)/(k - 1))\mathrm{e}^{\mathrm{i}\theta}$ parameterize the circumscribing circle depicted as the dashed black circle in Figure 3. As shown by Beardon and Beardon [**3**], the chord

$$s(t; \theta) = (1 - t)\ P(\theta) + t\ P(k\theta) \tag{3}$$

between $P(\theta)$ and $P(k\theta)$ is tangent to the curve $c(\theta)$ at θ; this is illustrated in Figure 3(b).

This observation provides the link to de Bruijn graphs: Edges $w_j^{(k,n)} \to w_{\gamma(j,r)}^{(k,n)}$ are chords between $P(\theta_j)$ and $P(\theta_{kj+r})$. Edges in the family $r = 0$ are therefore tangent to the epicycloid (2) since $\theta_{kj} = k\theta_j$. For $r > 0$, the edges are not tangent to epicycloids. However, for each $r \in \{1, \ldots, k-1\}$, the angle θ_{kj+r} can be made arbitrarily close to $\theta_{kj} = k\theta_j$ by choosing n sufficiently large since $|\theta_{kj+r} - \theta_{kj}| = r2\pi/k^n$ for all $j \in \{0, 1, \ldots, k^n - 1\}$. It is in this sense that these edges approach tangency to epicycloids as $n \to \infty$.

REFERENCES

1. N. G. de Bruijn, A combinatorial problem, *Nederl. Akad. Wetensch. Proc.* **49** (1946) 758–764.
2. N. G. de Bruijn, Acknowledgement of priority to C. Flye Sainte-Marie on the counting of circular arrangements of 2^n zeros and ones that show each n-letter word exactly once, *Technological University Eindhoven Report 75-WSK-06* (1975) 1–14.
3. A. F. Beardon, L. A. Beardon, Circles, chords, and epicycloids, *The Mathematical Gazette* **73** no. 465 (1989) 192–197, http://dx.doi.org/10.2307/3618437.

Department of Mathematics, Colorado State University, Fort Collins, CO 80523, USA
motta@math.colostate.edu
shipman@math.colostate.edu
springer@math.colostate.edu

100 Years Ago in The American Mathematical Monthly
Edited by Vadim Ponomarenko

In Volume 10 of the eleventh edition of the *Encyclopedia Britannica*, under the word Fermat, there appears the following statement: "He died in the belief that he had found a relation which every prime number must satisfy, namely $2^{2^n} + 1 =$ a prime." It is very evident that most of the small prime numbers are not of this form. In the noted French encyclopedia, called *Nouveau Larousse*, there appears, under the word substitution, the following statement: "The notion of group due to Galois has opened a vast field." Every one ought to know that the notion of group is much older than the work of Galois. As these statements appear in works which are generally regarded as very reliable they are of especial interest [*sic*]. (G. A. Miller, in *School Science and Mathematics*.)

Excerpted from "Notes and News" **22** (1915) 36–38.

The Axiom of Choice, Well-Ordering, and Well-Classification

Hossein Hosseini Giv

Abstract. Let X be a nonempty set and $\mathcal{P}^*(X)$ denote the set of all nonempty subsets of X. In this note, we give a simple characterization of those choice functions on $\mathcal{P}^*(X)$ that are induced by a well-ordering on X. The arguments involved in this result lead us to an interesting theorem, which we refer to as the *well-classification theorem*. This is finally proved to be an equivalent form of the axiom of choice.

The axiom of choice, one of the most important axioms of mathematics, first appeared in *E. Zermelo*'s work [**5**] on the *well-ordering theorem*. Its usual form is as follows.

The axiom of choice. *If Γ is a nonempty set whose elements are nonempty sets, there exists a function f on Γ such that $f(X) \in X$ for every $X \in \Gamma$.*

The function f is called a *choice function*. Since this axiom does not say how to find such a function, it can be used to present *nonconstructive* proofs, that is, proofs in which the existence of a mathematical object is proved without describing the way it can be found or constructed. For this reason, the axiom raised controversy when it first appeared in the literature. Nevertheless, it soon became clear that without it, many important results of set theory and other disciplines like topology and functional analysis could not be proved.

Today, the axiom and some of its numerous equivalent forms, like *Zorn's lemma*, *Hausdorff maximal principle*, and the *well-ordering theorem*, play an essential role in modern mathematics [**2**]. For example, the axiom of choice is adjoined to the Zermelo–Fraenkel (ZF) axioms of set theory to obtain the most acceptable form of *axiomatic set theory*, known as ZFC (the letter C stands for choice) [**1, 4**]. Also, Zorn's lemma is utilized in the proof of such crucial results of linear algebra and functional analysis as the existence of Hamel bases and the Hahn–Banach theorem. A detailed list of the various equivalent forms of the axiom of choice can be found in [**3**].

Among the equivalent forms of the axiom of choice, the well-ordering theorem (sometimes the well-ordering principle) occupies an important place. Therefore, we state it.

Theorem 1. *(Well-ordering theorem) Every nonempty set can be equipped with a total ordering $\leq$ under which every nonempty subset has a minimal element.*

A total ordering with this property is called a *well-ordering*. The proof of Theorem 1, due to Zermelo, uses the axiom of choice and, accordingly, depends on the existence of a choice function. For this reason, the proof is nonconstructive and cannot be used for the construction of well-orderings in concrete situations.

Now, for a nonempty set X let $\mathcal{P}^*(X)$ denote the set of all nonempty subsets of X. Clearly, a well-ordering $\leq$ on X induces a choice function on $\mathcal{P}^*(X)$. This sends each $A \in \mathcal{P}^*(X)$ to its minimum with respect to $\leq$, denoted by $\min_{\leq} A$.

http://dx.doi.org/10.4169/amer.math.monthly.122.01.56
MSC: Primary 03E25

The above argument allows us to ask a natural question: Is every choice function on $\mathcal{P}^*(X)$ induced by a well-ordering $\leq$ on X as above?

A simple example shows that the answer is no. If X has three elements, $X = \{a, b, c\}$ for example, and we choose a from $\{a\}$ and $\{a, b\}$, b from $\{b\}$ and $\{b, c\}$, and c from $\{c\}$, $\{a, c\}$ and X, then the corresponding choice function is not induced by a well-ordering $\leq$ on X. If it was so, we would have $a \leq b \leq c \leq a$ and hence $a = b = c$, a contradiction.

Let us now take a closer look at this example. Can we explain it another way? First of all, we can use our choice function to define a relation $\mathcal{E}$ on $\mathcal{P}^*(X)$. If $A, B \in \mathcal{P}^*(X)$, write $A \mathcal{E} B$ if a common element is chosen from A and B. For instance $\{c\} \mathcal{E} \{a, c\}$. The relation is easily checked to be an equivalence relation on $\mathcal{P}^*(X)$.

At this point, we hope we can use $\mathcal{E}$ to explain why our choice function is not induced by a well-ordering. We first need to answer two questions. If X is any nonempty set, is it always true that every choice function on $\mathcal{P}^*(X)$ defines an equivalence relation on $\mathcal{P}^*(X)$? Can such relations be used to determine if the choice function is induced by a well-ordering?

The first question is answered in the following proposition, whose simple proof is omitted.

Proposition 1. *Let X be a nonempty set. If f is a choice function on $\mathcal{P}^*(X)$, then f defines an equivalence relation $\mathcal{E}$ on $\mathcal{P}^*(X)$ via*

$$A \mathcal{E} B \Leftrightarrow f(A) = f(B).$$

The relation $\mathcal{E}$ in the above proposition satisfies the following.

1. The set of all equivalence classes modulo $\mathcal{E}$ is
$$\mathcal{P}^*(X)/\mathcal{E} = \{\{x\}/\mathcal{E} : x \in X\},$$
where $\{x\}/\mathcal{E}$ is the equivalence class of $\{x\}$ modulo $\mathcal{E}$.
2. Every element of $\{x\}/\mathcal{E}$ contains x.

Call an equivalence relation $\mathcal{E}$ on $\mathcal{P}^*(X)$ with these two properties a *well-classification* of $\mathcal{P}^*(X)$. The above proposition therefore says that a choice function on $\mathcal{P}^*(X)$ defines a well-classification of $\mathcal{P}^*(X)$. Refer to this as the *well-classification of $\mathcal{P}^*(X)$ induced by the choice function*. With this terminology, we are now ready to answer our second question.

Theorem 2. *For some nonempty set X, let f be a choice function on $\mathcal{P}^*(X)$. Suppose that $\mathcal{E}$ is the well-classification of $\mathcal{P}^*(X)$ induced by f as in Proposition 1. Then f is induced by a well-ordering on X if and only if $\mathcal{E}$ has the following additional properties.*

1. *If $x \in X$ and $A \in \{x\}/\mathcal{E}$, then every set $B \subset A$ that contains x lies in $\{x\}/\mathcal{E}$.*
2. *If $x \in X$, $A \in \{x\}/\mathcal{E}$ and $y \in A$, then for every set C in $\{y\}/\mathcal{E}$, $C \cup \{x\}$ is an element of $\{x\}/\mathcal{E}$.*

Proof. First assume that f is induced by a well-ordering $\leq$ on X. If $y \in D$ for some $D \in \{x\}/\mathcal{E}$, then $x \leq y$. It follows that for $A \in \{x\}/\mathcal{E}$ and $B \subset A$ that contains x, $\min_{\leq} B = x$. This proves 1. Now, let x, A and y be as in 2. If $C \in \{y\}/\mathcal{E}$, then $\min_{\leq} C = y$. But, our choice of y shows that $x \leq y$. This gives $\min_{\leq} C \cup \{x\} = x$ and therefore that $C \cup \{x\} \in \{x\}/\mathcal{E}$, proving 2.

To prove the converse, suppose that $\mathcal{E}$ satisfies 1 and 2. For $x, y \in X$, write $x \leq y$ if some $A \in \{x\}/\mathcal{E}$ contains y. It is clear that for every $x \in X$, $x \leq x$. If $x, y \in X$

are such that $x \leq y$ and $y \leq x$, then it follows from our definition of $\leq$ and 1 that $\{x, y\} \in \{x\}/\mathcal{E} \cap \{y\}/\mathcal{E}$. Since $\mathcal{E}$ is a well-classification, this implies $x = y$. To verify transitivity, let $x, y, z \in X$ satisfy $x \leq y$ and $y \leq z$. Since $y \leq z$, $z \in B$ for some $B \in \{y\}/\mathcal{E}$. Since $x \leq y$, there is some $A \in \{x\}/\mathcal{E}$ such that $y \in A$, and 2 tells us that $B \cup \{x\} \in \{x\}/\mathcal{E}$. This gives $z \in B \cup \{x\}$, an element of $\{x\}/\mathcal{E}$, which means $x \leq z$. To complete the proof, let $Y \subset X$ be a nonempty set. Then there exists a unique $y \in X$ such that $Y \in \{y\}/\mathcal{E}$. Now our definition of $\leq$ shows that $\min_{\leq} Y = y$. ■

We now present another choice function that is not induced by any well-ordering.

Example 1. Consider $\mathbb{N}$, the set of positive integers, with its usual order $\leq$ as a well-ordering. We define a choice function f on $\mathcal{P}^*(\mathbb{N})$ as follows.

For every $n \in \mathbb{N}$, $f(\{n\}) = n$ and $f(\mathbb{N}) = 1$. If $\emptyset \neq B \subset \mathbb{N}$ is any other set, consider two cases: If $\min_{\leq} B^c = 1$, let $f(B) = \min_{\leq} B$, and if $\min_{\leq} B^c > 1$, set $f(B) = \min_{\leq} B^c - 1$.

Now, let $\mathcal{E}$ be the well-classification of $\mathcal{P}^*(\mathbb{N})$ induced by f as in Proposition 1. Then f is not induced by a well-ordering of $\mathbb{N}$ because of either of the following sample facts.

- The set $\{1, 2, 3\}$ is in $\{3\}/\mathcal{E}$, but its subset $\{2, 3\}$ that contains 3 lies in $\{2\}/\mathcal{E}$. This contradicts 1 of the above theorem.
- We have $2 \in \{1, 2, 3\}$ and this set is in $\{3\}/\mathcal{E}$. Now, $\{2\} \in \{2\}/\mathcal{E}$ and $\{2\} \cup \{3\}$ is not in $\{3\}/\mathcal{E}$, contradicting 2 of the above theorem.

It is worth mentioning that "well-classification" is an appropriate choice of terminology. In fact, if a well-classification $\mathcal{E}$ of $\mathcal{P}^*(X)$ can be found, then $\mathcal{P}^*(X)$ is classified in the most natural way. Every $A \in \mathcal{P}^*(X)$ lies in the equivalence class of a unique set $\{x\}$, and this shows that the mapping $x \mapsto \{x\}/\mathcal{E}$ is a one-to-one correspondence between X and $\mathcal{P}^*(X)/\mathcal{E}$.

So, if a well-classification $\mathcal{E}$ of $\mathcal{P}^*(X)$ exists, we say that $\mathcal{P}^*(X)$ *is well-classified*. In this case, a quotient of $\mathcal{P}^*(X)$, in the sense of equivalence relations, is in one-to-one correspondence with X. This, when compared with the fact that $\mathcal{P}^*(X)$ is usually much "larger" than X, can be of great interest.

So far, we have seen that the axiom of choice implies the following theorem.

Theorem 3. *(Well-classification theorem) If X is any nonempty set, then $\mathcal{P}^*(X)$ can be well-classified.*

Our concluding result shows that the well-classification theorem is an equivalent form of the axiom of choice.

Theorem 4. *The well-classification theorem implies the axiom of choice.*

Proof. Assume Theorem 3 and let Γ be a nonempty set whose elements are also sets and $X = \cup_{Y \in \Gamma} Y$. By Theorem 3, there is a well-classification $\mathcal{E}$ of $\mathcal{P}^*(X)$. Thus, for every $Y \in \Gamma$, there exists a unique $y \in X$ such that $Y \in \{y\}/\mathcal{E}$. This shows $y \in Y$. Now, the function $f : Y \in \Gamma \mapsto y \in X$ is a choice function. ■

REFERENCES

1. P. Bernays, *Axiomatic Set Theory.* North-Holland Publishing, Amsterdam, 1968.
2. T. J. Jech, *The Axiom of Choice.* North-Holland Publishing, Amsterdam, 1973.
3. H. Rubin, J. E. Rubin, *Equivalents of the Axiom of Choice.* Second printing. North-Holland Publishing, Amsterdam, 1970.

4. P. Suppes, *Axiomatic Set Theory.* D. Van Nostrand Company, Princeton, NJ, 1960.
5. E. Zermelo, Beweis, daß jede Menge wohlgeordnet werden kann, *Math. Annal.* **59** no. 4 (1904) 514–516.

Department of Mathematics, Faculty of Mathematics, University of Sistan and Baluchestan, Zahedan, Iran
giv@math.usb.ac.ir, hossein.giv@gmail.com
www.givmath.com

An Unbiased Marriage Theorem

We say that a bipartite graph H has *order* n if the vertex set of H has the bipartition $V = V_1 \cup V_2$, where $|V_1| = |V_2| = n$ and every edge is between the vertices of the two sets V_1 and V_2. Also for a subset S of vertices, let $N(S)$ denote the set of neighbors of the set S. We say that the bipartite graph H satisfies the (p, q)-*condition* if (i) for all subsets $I \subseteq V_1$ of cardinality at most p, the inequality $|I| \leq |N(I)|$ holds, and (ii) for all subsets $J \subseteq V_2$ of cardinality at most q, the inequality $|J| \leq |N(J)|$ holds. Note that the $(n, 0)$-condition is Hall's original condition in his marriage theorem [**1**].

Theorem 5. *Let H be a bipartite graph of order n satisfying the (p, q)-condition, and assume that $n \leq p + q$. Then the bipartite graph H contains a perfect matching.*

Proof. The proof is by induction on the order n. The induction basis is $n = 1$, which is straightforward. Now assume that it is true for all bipartite graphs of order less than n.

We consider two cases. First, assume that the inequalities in conditions (i) and (ii) are strict for all proper subsets I of V_1 and proper subsets J of V_2; that is, if for all $\emptyset \subsetneq I \subsetneq V_1$ with $|I| \leq p$ we have $|I| < |N(I)|$, and for all $\emptyset \subsetneq J \subsetneq V_2$ with $|J| \leq q$ we have $|J| < |N(J)|$. Match any vertex $x \in V_1$ with one of its neighbors y in V_2. Then the remaining graph $H' = H - \{x, y\}$ has order $n - 1$ and satisfies the (p, q)-condition. Hence H' has a matching M' by the induction hypothesis and we obtain the matching $M = M' \cup \{(x, y)\}$ for the graph H.

The second case is that there is a proper subset yielding an equality. Without loss of generality, assume that this is a subset of V_1; that is, we have $S \subsetneq V_1$ such that $1 \leq |S| = |N(S)| = k \leq p$. This divides the problem into finding a matching in two smaller graphs. First, we need to find a matching on graph H' induced by the vertex set $(V_1 - S) \cup (V_2 - N(S))$. Since there are no edges between $V_2 - N(S)$ and S, condition (ii) with the parameter q holds. Now pick I a subset of $V_1 - S$ of cardinality at most $p' = p - k$. Applying condition (i) to the disjoint union $I \cup S$ in H yields condition (i) holds with parameter p' for H'. Hence, H' of order $n - k$ satisfies the (p', q)-condition. Since $p' + q \geq n - k$, there is a matching M' in H'. Next, we need a matching on the graph H'' induced by the vertex set $S \cup N(S)$. But H'' satisfies the $(k, 0)$-condition, so it has a matching M'' by Hall's marriage theorem. Combining the matchings M' and M'' we obtain a matching for the bipartite graph H. ■

Note that the condition $n \leq p + q$ is necessary, since the graph $K_{p+1,p} \cup K_{q,q+1}$ of order $p + q + 1$ has no perfect matching but satisfies the (p, q)-condition.

ACKNOWLEDGMENT. The author thanks Margaret Readdy for her careful comments.

REFERENCE

1. P. Hall, On representation of subsets, *J. London Math. Soc.* **10** (1935) 26–30.

—Submitted by R. Ehrenhorg*, University of Kentucky

http://dx.doi.org/10.4169/amer.math.monthly.122.01.59
MSC: Primary 05C70
*The author was partially supported by National Security Agency grant H98230-13-1-0280.

Kronecker Square Roots and the Block Vec Matrix

Ignacio Ojeda

Dedicado a Carlos Benítez (in memoriam)

Abstract. Using the block vec matrix, I give a necessary and sufficient condition for factorization of a matrix into the Kronecker product of two other matrices. As a consequence, I obtain an elementary algorithmic procedure to decide whether a matrix has a square root for the Kronecker product.

1. INTRODUCTION. My statistician colleague, J.E. Chacón, asked me how to decide if a real given matrix A has a square root for the Kronecker product (i.e., if there exists a B such that $A = B \otimes B$) and, in the positive case, how to compute it. His questions were motivated by the fact that, provided that a certain real positive definite symmetric matrix has a Kronecker square root, explicit asymptotic expressions for certain estimator errors could be obtained. See [**1**] for a discussion of the importance of multivariate kernel density derivative estimation.

This note provides a suitable reference for the existence of square roots for the Kronecker product and it is organized as follows. First of all, I study the problem of the factorization of a matrix into a Kronecker product of two matrices by giving a necessary and sufficient condition under which this happens (Theorem 1). As a preparation for the main result, I introduce the *block vec matrix* (Definition 1). Now, the block vec matrix and Theorem 1 solve our problem in a constructive way.

2. KRONECKER PRODUCT FACTORIZATION. Throughout this note, $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{C}$ denote the sets of nonnegative integers, real numbers, and complex numbers, respectively. All matrices considered here have real or complex entries; $A^\top$ denotes the transpose of A and $\operatorname{tr}(A)$ denotes its trace.

The operator that transforms a matrix into a stacked vector is known as the *vec operator* (see, [**3**, Definition 4.2.9] or [**6**, §7.5]). If $A = \begin{pmatrix}\mathbf{a}_1|\dots|\mathbf{a}_n\end{pmatrix}$ is an $m \times n$ matrix whose columns are $\mathbf{a}_1, \dots, \mathbf{a}_n$, then $\operatorname{vec}(A)$ is the $mn \times 1$ matrix

$$\operatorname{vec}(A) = \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{pmatrix}.$$

The following definition generalizes the vec operation and is the key to all that follows.

Definition. Let $A = (A_{ij})$ be an $mp \times nq$ matrix partitioned into block matrices A_{ij}, each of order $p \times q$. The *block vec matrix* of A corresponding to the given partition is the $mn \times pq$ matrix

http://dx.doi.org/10.4169/amer.math.monthly.122.01.60
MSC: Primary 15A23, Secondary 15A69

$$\mathrm{vec}^{(p\times q)}(A) = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix}, \quad \text{where each } A_j = \begin{pmatrix} \mathrm{vec}(A_{1j})^\top \\ \vdots \\ \mathrm{vec}(A_{mj})^\top \end{pmatrix}.$$

If A is $m \times n$, it is instructive to verify the following identities corresponding to four natural ways to partition it:

- $p = q = 1 : \mathrm{vec}^{(1\times 1)}(A) = \mathrm{vec}(A)$,
- $p = m, q = n : \mathrm{vec}^{(m\times n)}(A) = \mathrm{vec}(A)^\top$,
- $p = 1, q = n$ (partition by rows): $\mathrm{vec}^{(1\times n)}(A) = A$,
- $p = m, q = 1$ (partition by columns): $\mathrm{vec}^{(m\times 1)}(A) = A^\top$.

If A is $mp \times nq$ and $\mathrm{vec}(A)$ is partitioned into nq blocks, each of size $mp \times 1$, then a computation reveals that $\mathrm{vec}^{(mp\times 1)}(\mathrm{vec}(A)) = A^\top$.

Let $B = (b_{ij})$ be an $m \times n$ matrix and let C be a $p \times q$ matrix. The *Kronecker product* of B and C, denoted by $B \otimes C$, is the $mp \times nq$ matrix

$$B \otimes C = \begin{pmatrix} b_{11}C & b_{12}C & \cdots & b_{1n}C \\ b_{21}C & b_{22}C & \cdots & b_{2n}C \\ \vdots & \vdots & & \vdots \\ b_{m1}C & b_{m2}C & \ldots & b_{mn}C \end{pmatrix}.$$

The following result is straightforward; see [**7**, Theorem 1].

Lemma 1. *Let B be an $m \times n$ matrix and let C be a $p \times q$ matrix. Then*

$$\mathrm{vec}^{(p\times q)}(B \otimes C) = \mathrm{vec}(B)\mathrm{vec}(C)^\top.$$

In particular, $\mathrm{vec}^{(p\times q)}(B \otimes C)^\top = \mathrm{vec}^{(m\times n)}(C \otimes B)$.

It may be (but need not be) possible to factor a given matrix, suitably partitioned, as a Kronecker product of two other matrices. For example, a zero matrix can always be factored as a Kronecker product of a zero matrix and any matrix of suitable size. The following theorem provides a necessary and sufficient condition for a Kronecker factorization.

Theorem 1. *Let $A = (A_{ij})$ be a nonzero $mp \times nq$ matrix, partitioned into blocks of order $p \times q$. There exist matrices B (of order $m \times n$) and C (of order $p \times q$) such that $A = B \otimes C$ if and only if* $\mathrm{rank}(\mathrm{vec}^{(p\times q)}(A)) = 1$.

Proof. If $A = B \otimes C$ is a factorization of the stated form, then B, C, $\mathrm{vec}(B)$, and $\mathrm{vec}(C)$ must all be nonzero. Lemma 1 ensures that

$$\mathrm{rank}(\mathrm{vec}^{(p\times q)}(A)) = \mathrm{rank}(\mathrm{vec}^{(p\times q)}(B \otimes C)) = \mathrm{rank}(\mathrm{vec}(B)\mathrm{vec}(C)^\top) = 1.$$

Conversely, since $A \neq 0$, there are indices r and s such that $A_{rs} \neq 0$ and hence $\mathrm{vec}(A_{rs}) \neq 0$. Since $\mathrm{rank}(\mathrm{vec}^{(p\times q)}(A)) = 1$, each row of $\mathrm{vec}^{(p\times q)}(A)$ is a scalar multiple of any nonzero row. Thus, there are scalars b_{ij} such that each $\mathrm{vec}(A_{ij}) = b_{ij}\,\mathrm{vec}(A_{rs})$. This means that $A = B \otimes C$, in which $B = (b_{ij})$ and $C = A_{rs}$. ∎

Notice that the preceding proof provides a simple construction for a pair of Kronecker factors for A if $\operatorname{rank}(\operatorname{vec}^{(p\times q)}(A)) = 1$.

The block vec matrix can be used to detect not only whether a given matrix has a Kronecker factorization of a given form but also, if it does not, how closely it can be approximated in the Frobenius norm by a Kronecker product. A best approximation is determined by the singular value decomposition of the block vec matrix. For details, see [**7**], where the block vec matrix is called the *rearrangement matrix*.

Example. Consider

$$A = \left(\begin{array}{cc} 2 & 1 \\ 2 & 0 \\ \hline 3 & 0 \\ 0 & 3 \end{array}\right).$$

Since

$$\operatorname{vec}^{(2\times 2)}(A) = \begin{pmatrix} 2 & 2 & 1 & 0 \\ 3 & 0 & 0 & 3 \end{pmatrix}$$

has rank 2, Theorem 1 ensures that $A \neq B \otimes C$, for any B and C of order 2×1 and 2×2, respectively.

The set of matrices that factor into the Kronecker product of two other matrices have the following interpretation in algebraic geometry.

Remark. Let A be a real $mp \times nq$ matrix and consider the big matrix

$$M = \begin{pmatrix} I_{mn} \otimes \mathbf{u}_{pq}^{\top} \\ \mathbf{u}_{mn}^{\top} \otimes I_{pq} \end{pmatrix}$$

where $\mathbf{u}_\bullet$ is the all-ones vector of dimension $\bullet$. Let K be equal to $\mathbb{R}$ or $\mathbb{C}$ and let

$$\varphi_M : K[X_{11}, \ldots, X_{m1}, X_{12}, \ldots, X_{mn\,pq}] \longrightarrow K[t_1, \ldots, t_{mn+pq}]$$

be the $K-$algebra map such that $\varphi(\mathbf{X}^{\mathbf{u}}) = \mathbf{t}^{M\mathbf{u}}$, with $\mathbf{X}^{\mathbf{u}} := X_{11}^{u_1} \cdots X_{mn\,pq}^{u_{mnpq}}$ and $\mathbf{t}^{\mathbf{v}} := t_1^{v_1} \cdots t_{mn+pq}^{v_{mn+pq}}$. Then $X = (x_{ij})$ has rank 1 if and only if $\operatorname{vec}(X)^{\top}$ is a zero of $\ker \varphi_M$ (see, e.g., [**2**, Section 2]). Therefore, by Theorem 1, the set of $mn \times pq$ matrices that factor into the Kronecker product of an $m \times n$ matrix and a $p \times q$ matrix is the following *algebraic set*

$$\mathcal{V}(\ker \varphi_M) = \left\{A \in K^{mn\times pq} \mid \operatorname{vec}(\operatorname{vec}^{(p\times q)}(A))^{\top} \in \ker \varphi_M\right\}.$$

In algebraic statistics, $\ker \varphi_M$ is the ideal associated to two independent random variables with values in $\{1, \ldots, mn\}$ and $\{1, \ldots, pq\}$. Thus, as it is well know in statistics, *factorization means independence and vice versa*.

We can now give a solution to the problem that motivates this paper.

Corollary 1. *If A is a nonzero $m^2 \times n^2$ matrix and B an $n \times n$ matrix, then*

(a) $A = B \otimes B$ *if and only if* $\mathrm{vec}^{(m\times n)}(A) = \mathrm{vec}(B)\mathrm{vec}(B)^\top$,

(b) *if* $A = B \otimes B$, *then* $\mathrm{vec}^{(m\times n)}(A)$ *is symmetric and has rank one.*

Is the necessary condition in the preceding corollary sufficient? It is not surprising that the answer depends on the field. For example, if $m = n = 1$ and $A = (-1)$, then $\mathrm{vec}^{(1\times 1)}(A) = (-1)$ is symmetric and has rank one, but A has no real Kronecker square root; it does have complex Kronecker square roots $B = (\pm \mathrm{i})$, but these are the only ones.

Theorem 2. *Let* A *be an* $m^2 \times n^2$ *real or complex matrix. Suppose that* $\mathrm{vec}^{(m\times n)}(A)$ *is symmetric and has rank one.*

(a) *There is an* $m \times n$ *matrix* B *such that* $A = B \otimes B$.

(b) *If* B *and* C *are* $m \times n$ *matrices such that* $A = B \otimes B = C \otimes C$, *then* $C = \pm B$.

(c) *If* A *is real, there is a real* $m \times n$ *matrix* B *such that* $A = B \otimes B$ *if and only if* $\mathrm{tr}(\mathrm{vec}^{(m\times n)}(A)) > 0$.

Proof. Any complex symmetric matrix has a special singular value decomposition, that is unique in a certain way; see [**4**, Corollary 4.4.4: Autonne's theorem]. In the case of a rank one symmetric matrix Z whose largest (indeed, only nonzero) singular value is σ, Autonne's theorem says that there is a unit vector $\mathbf{u}$ such that $Z = \sigma \mathbf{u}\mathbf{u}^\top$. Moreover, if $\mathbf{v}$ is a unit vector such that $Z = \sigma \mathbf{v}\mathbf{v}^\top$, then $\mathbf{v} = \pm \mathbf{u}$. If we use Autonne's theorem to represent the block vec matrix as $\mathrm{vec}^{(m\times n)}(A) = \sigma \mathbf{u}\mathbf{u}^\top = (\sigma^{1/2}\mathbf{u})(\sigma^{1/2}\mathbf{u})^\top$ and define B by $\mathrm{vec}(B) = (\sigma^{1/2}\mathbf{u})$, we have $\mathrm{vec}^{(m\times n)}(A) = \mathrm{vec}(B)\mathrm{vec}(B)^\top$. The preceding corollary now ensures that $A = B \otimes B$ and the assertion in (b) follows from the uniqueness part of Autonne's theorem.

Now suppose that A is real. If there is a real B such that $A = B \otimes B$, then $\mathrm{tr}(\mathrm{vec}^{(m\times n)}(A) = \mathrm{tr}(\mathrm{vec}(B)\mathrm{vec}(B)^\top) = \mathrm{vec}(B)^\top B$ is positive since it is the square of the Euclidean norm of the (necessarily nonzero) real vector $\mathrm{vec}(B)$. Conversely, the spectral theorem ensures that any real symmetric matrix can be represented as $Q\Lambda Q^\top$, in which Q is real orthogonal and Λ is real diagonal. Since the block vec matrix is real symmetric and has rank one, we can take $\Lambda = \mathrm{diag}(\lambda, 0, \ldots, 0)$ and represent $\mathrm{vec}^{(m\times n)}(A) = \lambda \mathbf{q}\mathbf{q}^\top$ in which $\mathbf{q}$ is the first column of Q. If $\lambda = \mathrm{tr}(\mathrm{vec}^{(m\times n)}(A)) > 0$, then $\mathrm{vec}^{(m\times n)}(A) = (\lambda^{1/2}\mathbf{q})(\lambda^{1/2}\mathbf{q})^\top = \mathrm{vec}(B)\mathrm{vec}(B)^\top$, in which $\mathrm{vec}(B) = \lambda^{1/2}\mathbf{q}$ (and hence also B) is real. ∎

The uniqueness part of the preceding theorem has some perhaps surprising consequences.

Corollary 2. *Let* A *be a nonzero* $m^2 \times n^2$ *real or complex matrix and suppose that* $A = B \otimes B$ *for some* $m \times n$ *matrix* B.

(a) *A is symmetric if and only if B is either symmetric or skew symmetric.*

(b) *A is not skew symmetric.*

(c) *A is Hermitian if and only if B is either Hermitian or skew Hermitian.*

(d) *A is Hermitian positive definite if and only if B is Hermitian and definite (positive or negative).*

(e) *A is skew Hermitian if and only if* $\mathrm{e}^{\mathrm{i}\pi/4}B$ *is Hermitian.*

(f) *A is unitary if and only if B is unitary.*

(g) *If B is real, then A is real orthogonal if and only if B is real orthogonal.*

(h) *A is complex orthogonal if and only if either B or $\mathrm{i}B$ is complex orthogonal.*

Proof.

(a) $A^\top = B^\top \otimes B^\top$, so $A = A^\top$ if and only if $A = B \otimes B = B^\top \otimes B^\top$, which holds if and only if $B^\top = \pm B$.

(b) If $A^\top = -A^\top$, then $-A = -B \otimes B = (\mathrm{i}B) \otimes (\mathrm{i}B) = B^\top \otimes B^\top$ and hence $B^\top = \pm \mathrm{i}B = \pm \mathrm{i}(B^\top)^\top = \pm \mathrm{i}(\pm \mathrm{i}B)^\top = -B^\top$, so $B = 0$.

(c) $A = A^*$ if and only if $A = B \otimes B = B^* \otimes B^*$; that is to say, $B^* = \pm B$.

(d) Using (c) and the fact that the eigenvalues of $B \otimes B$ are the pairwise products of the eigenvalues of B, we can exclude the possibility that B is skew Hermitian since in that case its nonzero eigenvalues (there must be at least one) would be purely imaginary. Hence, $B \otimes B$ would have at least one negative eigenvalue.

(e) Under our hypothesis, the following statements are equivalent (i) $A = -A^*$, (ii) $A = B \otimes B = -B^* \otimes B^* = (iB)^* \otimes (iB)^*$ (iii) $B^* = \pm \mathrm{i}B$ (iv) $(\mathrm{e}^{\mathrm{i}\pi/4}B)^* = \pm \mathrm{e}^{\mathrm{i}\pi/4}B$, and so the claim follows.

(f) $A^{-1} = B^{-1} \otimes B^{-1} = B^* \otimes B^*$ if and only if $B^* = \pm B^{-1}$, which is equivalent to $BB^* = \pm I$. However, $BB^* = -I$ is not possible since BB^* is positive definite.

(g) Follows from (f).

(h) $A^{-1} = B^{-1} \otimes B^{-1} = B^\top \otimes B^\top$ if and only if $B^\top = \pm B^{-1}$. That is to say, $BB^\top = \pm I$ or, equivalently, either $BB^\top = I$ or $(\mathrm{i}B)(\mathrm{i}B)^\top = I$. ∎

ACKNOWLEDGMENT. I would like to thank J.E Chacón for pointing me toward the question of the square roots of matrices for the Kronecker product. I also thank the anonymous referee for the comments and useful suggestions. The author is supported by the project MTM2012-36917-C03-01, National Plan I+D+I and by Junta de Extremadura (FEDER funds).

REFERENCES

1. J. E. Chacón, T. Duong, M. Wand, Asymptotics for general multivariate kernel density derivative estimators, *Statist. Sinica* **21** no. 2 (2011) 807–840.
2. D. Geiger, C. Meek, B. Sturmfels, On the toric algebra of graphical models, *Ann. Statist.* **34** no. 3 (2006) 1463–1492.
3. R. A. Horn, C. R. Johnson, *Topics in Matrix Analysis*. Cambridge Univ. Press, Cambridge, UK, 1991.
4. R. A. Horn, C. R. Johnson, *Matrix Analysis*. Second edition, Cambridge Univ. Press, Cambridge, UK, 2013.
5. J. R. Magnus, H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Revised reprint of the 1988 original. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, 1999.
6. J. R. Schott, *Matrix Analysis for Statistics*. Second edition. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2005.
7. C. Vanloan, N. Pitsianis, Approximation with Kronecker products, in *Linear Algebra for Large-Scale and Real-Time Applications*. Edited by M. S. Moonen, G. H. Golub, and B. L. de Moor, Kluwer Academic Publishers, New York 1993, 293–314.

Departamento de Matemáticas, Universidad de Extremadura, E-06071 Badajoz, España
ojedamc@unex.es

On a Formula of S. Ramanujan

Pablo A. Panzone

Abstract. Certain finite families of rational functions have the property that the coefficients of the expansions in power series of their elements are related by simple algebraic expressions. We prove an identity in the spirit of S. Ramanujan using a result of T. N. Sinha.

A remarkable formula of S. Ramanujan states that if

$$\frac{1+53x+9x^2}{1-82x-82x^2+x^3}=\sum_{i\geq 0}\alpha_i x^i,$$

$$\frac{2-26x-12x^2}{1-82x-82x^2+x^3}=\sum_{i\geq 0}\beta_i x^i,$$

and

$$\frac{2+8x-10x^2}{1-82x-82x^2+x^3}=\sum_{i\geq 0}\gamma_i x^i,$$

then

$$\alpha_i^3+\beta_i^3=\gamma_i^3+(-1)^i.$$

This was proved by M. Hirschhorn in [**2**]. Two more proofs were given in [**4**] and [**3**], the last one is a *proof by example* which means that one verifies an identity by showing that it holds for a finite number of values (see [**6**], p. 9). Recently J. McLaughlin proved a very nice and much more complicated result in [**5**]. This result is also discussed in [**1**].

In this note we prove the following theorem using a modified version of a result of T. N. Sinha.

Theorem. *If*

$$\sum_{i=0}^{\infty}a_i(1)x^i=\frac{15-61x+21x^2}{1-2x-2x^2+x^3},\quad \sum_{i=0}^{\infty}a_i(2)x^i=\frac{13-77x-7x^2}{1-2x-2x^2+x^3},$$

$$\sum_{i=0}^{\infty}a_i(3)x^i=\frac{27-135x+13x^2}{1-2x-2x^2+x^3},\quad \sum_{i=0}^{\infty}a_i(4)x^i=\frac{-11-49x+33x^2}{1-2x-2x^2+x^3},$$

http://dx.doi.org/10.4169/amer.math.monthly.122.01.65
MSC: Primary 11A99, Secondary 11B34

$$\sum_{i=0}^{\infty} a_i(5)x^i = \frac{17 - 75x - 71x^2}{1 - 2x - 2x^2 + x^3}, \quad \sum_{i=0}^{\infty} a_i(6)x^i = \frac{3 - 17x - 91x^2}{1 - 2x - 2x^2 + x^3},$$

$$\sum_{i=0}^{\infty} a_i(7)x^i = \frac{29 - 79x - 51x^2}{1 - 2x - 2x^2 + x^3}, \quad \sum_{i=0}^{\infty} a_i(8)x^i = \frac{-1 + 31x - 79x^2}{1 - 2x - 2x^2 + x^3},$$

$$\sum_{i=0}^{\infty} a_i(9)x^i = \frac{-13 + 45x - 59x^2}{1 - 2x - 2x^2 + x^3}, \quad \sum_{i=0}^{\infty} b_i(1)x^i = \frac{-1 - 9x + 13x^2}{1 - 2x - 2x^2 + x^3},$$

$$\sum_{i=0}^{\infty} b_i(2)x^i = \frac{13 - 67x + 33x^2}{1 - 2x - 2x^2 + x^3}, \quad \sum_{i=0}^{\infty} b_i(3)x^i = \frac{-13 - 5x - 7x^2}{1 - 2x - 2x^2 + x^3},$$

$$\sum_{i=0}^{\infty} b_i(4)x^i = \frac{17 - 115x + 21x^2}{1 - 2x - 2x^2 + x^3}, \quad \sum_{i=0}^{\infty} b_i(5)x^i = \frac{29 - 129x + x^2}{1 - 2x - 2x^2 + x^3},$$

$$\sum_{i=0}^{\infty} b_i(6)x^i = \frac{15 - 81x - 59x^2}{1 - 2x - 2x^2 + x^3}, \quad \sum_{i=0}^{\infty} b_i(7)x^i = \frac{1 - 23x - 79x^2}{1 - 2x - 2x^2 + x^3},$$

$$\sum_{i=0}^{\infty} b_i(8)x^i = \frac{3 - 7x - 51x^2}{1 - 2x - 2x^2 + x^3}, \quad \sum_{i=0}^{\infty} b_i(9)x^i = \frac{-11 + 51x - 71x^2}{1 - 2x - 2x^2 + x^3},$$

$$\sum_{i=0}^{\infty} b_i(10)x^i = \frac{27 - 35x - 91x^2}{1 - 2x - 2x^2 + x^3},$$

then

$$\sum_{j=1}^{10} b_i(j)^k - \sum_{j=1}^{9} a_i(j)^k = (-1)^{ik},$$

for $k = 1, \ldots, 8$ *and all* $i = 0, 1, 2, 3, \ldots$.

The proof will follow from two lemmas. In the next lemma we write for short $A_i = A_i(m, n)$.

Lemma 1. *If*

$$A_1 = m^2 - mn - n^2, \quad A_2 = 15m^2 - 25mn - 21n^2, \quad A_3 = 13m^2 - 71mn + 7n^2,$$

$$A_4 = 27m^2 - 95mn - 13n^2, \quad A_5 = -11m^2 - 27mn - 33n^2,$$

$$A_6 = 17m^2 - 129mn + 71n^2, \quad A_7 = 3m^2 - 105mn + 91n^2,$$

$$A_8 = 29m^2 - 101mn + 51n^2, \quad A_9 = -m^2 - 49mn + 79n^2,$$

$$A_{10} = -13m^2 - 27mn + 59n^2,$$

and

$$B_1 = -m^2 + 3mn - 13n^2, \quad B_2 = 13m^2 - 21mn - 33n^2,$$

$$B_3 = -13m^2 - 25mn + 7n^2, \quad B_4 = 17m^2 - 77mn - 21n^2,$$

$$B_5 = 29m^2 - 99mn - n^2, \quad B_6 = 15m^2 - 125mn + 59n^2,$$

$$B_7 = m^2 - 101mn + 79n^2, \quad B_8 = 3m^2 - 55mn + 51n^2,$$

$$B_9 = -11m^2 - 31mn + 71n^2, \quad B_{10} = 27m^2 - 99mn + 91n^2,$$

then the following identity

$$\sum_{i=1}^{10} A_i^k = \sum_{i=1}^{10} B_i^k, \tag{1}$$

is valid for $k = 1, \ldots, 8$ *where* m, n *are arbitrary.*

The proof is just a check. We remark that these equalities are the heart of the proof and they are similar to a result of T. N. Sinha, see [**7**].

Next, recall that the Fibonacci sequence is defined by

$$F_0 = 0, \quad F_1 = 1, \quad F_{i+2} = F_{i+1} + F_i,$$

and one has the formula

$$F_i = \frac{1}{\sqrt{5}}\{\alpha^i - \beta^i\},$$

with $\alpha = \frac{\sqrt{5}+1}{2}$; $\beta = \frac{-\sqrt{5}+1}{2}$. Hint: observe that α, β are the roots of $x^2 - x - 1 = 0$. Therefore, α^i, β^i both satisfy the Fibonacci recurrence for $\alpha^{i+2} - \alpha^{i+1} - \alpha^i = \alpha^i(\alpha^2 - \alpha - 1) = 0$.

Lemma 2. *If* $i = 0, 1, 2, \ldots$, *then*

$$F_{i+1}^2 - F_i(F_{i+1} + F_i) = F_{i+1}^2 - F_i F_{i+2} = (-1)^i.$$

Also

$$\sum_{i=0}^{\infty} F_i^2 x^i = \frac{x(1-x)}{1-2x-2x^2+x^3},$$

$$\sum_{i=0}^{\infty} F_{i+1}^2 x^i = \frac{(1-x)}{1-2x-2x^2+x^3} \quad \textit{and}$$

$$\sum_{i=0}^{\infty} F_{i+1} F_i x^i = \frac{x}{1-2x-2x^2+x^3}.$$

Proof. One has

$$F_{i+1}^2 - F_i F_{i+2} = \frac{(\alpha^{i+1}-\beta^{i+1})^2}{5} - \frac{(\alpha^i-\beta^i)(\alpha^{i+2}-\beta^{i+2})}{5}$$

$$= \frac{-2(\alpha\beta)^{i+1}}{5} + \frac{(\alpha\beta)^i(\alpha^2+\beta^2)}{5}.$$

The first identity of the lemma follows from this noticing that $\alpha\beta = -1$, $\alpha^2+\beta^2 = 3$.

The first series is

$$\sum_{i=0}^{\infty} F_i^2 x^i = \sum_{i=0}^{\infty} \frac{(\alpha^i-\beta^i)^2}{5} x^i = \frac{1}{5}\sum_{i=0}^{\infty} (\alpha^{2i}+\beta^{2i}-2(-1)^i)x^i$$

$$= \frac{1}{5}\left\{\frac{1}{1-\alpha^2 x} + \frac{1}{1-\beta^2 x} - \frac{2}{1+x}\right\},$$

and the result follows.

The other series are proved in a similar way. ∎

Proof of theorem. In the identity of Lemma 1 set

$$m = F_{i+1}, \quad n = F_i.$$

Then

$$A_1 = m^2 - mn - n^2 = F_{i+1}^2 - F_i(F_{i+1}+F_i) = F_{i+1}^2 - F_i F_{i+2} = (-1)^i,$$

where the last equality follows from Lemma 2.

The rational functions stated in the theorem are obtained letting $a_i(1), \ldots, a_i(9)$ correspond to $A_2, \ldots, A_{10}$ and $b_i(1), \ldots, b_i(10)$ correspond to $B_1, \ldots, B_{10}$ with the above definition of m and n. In what follows we calculate the first rational function

stated, i.e., that corresponding to $a_i(1)$, but we omit the lengthy calculations for the rest.

One has

$$a_i(1) = A_2 = 15F_{i+1}^2 - 25F_{i+1}F_i - 21F_i^2.$$

Therefore, using Lemma 2

$$\sum_{i=0}^{\infty} a_i(1)x^i = \frac{15 - 61x + 21x^2}{1 - 2x - 2x^2 + x^3},$$

as stated.

Now the theorem follows from identity (1). ■

ACKNOWLEDGMENT. We wish to thank a referee for pointing out reference [**5**] and for a careful reading of the original manuscript.

REFERENCES

1. G. Andrews, B. Berndt, *Ramanujan's Lost Notebook: Part IV*. Springer-Verlag, New York, 2013.
2. M. D. Hirschhorn, An amazing identity of Ramanujan, *Math. Mag.* **68** no. 3 (1995) 199–201, http://dx.doi.org/10.2307/2691416.
3. M. D. Hirschhorn, A proof in the spirit of Zeilberger of an amazing identity of Ramanujan, *Math. Mag.* **69** (1996) 267–269, http://dx.doi.org/10.2307/2690530.
4. M. D. Hirschhorn, J. H. Han, Another look at an amazing identity of Ramanujan, *Math. Mag.* **79** no. 4 (2006) 302–304, http://dx.doi.org/10.2307/27642956.
5. J. McLaughlin, An identity motivated by an amazing identity of Ramanujan, *Fibonacci Quart.* **48** (2010) 34–38.
6. M. Petkovšek, H. S. Wilf, D. Zeilberger, *A=B*. A K Peters/CRC Press, Wellesley, MA, 1996.
7. T. N. Sinha, On the Tarry-Escott problem, *Amer. Math. Monthly* **73** no. 3 (1966) 280–285, http://dx.doi.org/10.2307/2315345.

INMABB-Departamento de Matemática, Universidad Nacional del Sur, Bahía Blanca, Buenos Aires, Argentina, 8000
pablopanzone@hotmail.com

Continuity is an Adjoint Functor

Edward S. Letzter

Abstract. We explain how the definition of continuity for functions between topological spaces can be rephrased as an adjointness condition between naturally arising functors.

1. INTRODUCTION. The emergence of category theory, introduced by S. Eilenberg and S. Mac Lane in the 1940s (see [**2**]), was among the most important mathematical developments of the twentieth century. The profound impact of the theory continues to this day, and categorical methods are currently used, for example, in algebra, geometry, topology, mathematical physics, logic, and theoretical computer science. (Hints of this breadth can be found, e.g., in [**1**], [**6**], and [**7**].)

A *category* is comprised of *objects* and *morphisms* between objects. Standard elementary examples include $\mathcal{S}et$, where the objects are sets and the morphisms are set functions; $\mathcal{V}ec_{\mathbb{F}}$, where the objects are vector spaces over a field $\mathbb{F}$ and the morphisms are $\mathbb{F}$-linear transformations; $\mathcal{G}rp$, where the objects are groups and the morphisms are group homomorphisms; and $\mathcal{T}op$, where the objects are topological spaces and the morphisms are continuous functions.

Relationships among different categories are established via *functors* between them. A pair of *adjoint functors*, as formulated in 1958 by D. M. Kan [**3**], determines a particularly close tie between two categories. Adjoint functors are essential tools in category theory, and their introduction was a significant milestone in its development.

While it is commonly held that adjoint functors "occur almost everywhere" [**6**, p. 107]), at least in many areas of mathematics, the typical first examples presented to students may not immediately reveal the fundamental importance of the ideas involved. (One such typical example, a left adjoint to a "forgetful functor," is described at the end of the brief review provided in the next section.)

Our aim in this note, then, is to illustrate how a natural example of adjoint functors can be "found" in the definition of a continuous map between topological spaces. In particular, we show, for a set function $\varphi : X \to Y$ between topological spaces X and Y, that φ is continuous if and only if certain naturally arising functors are adjoint.

Remark. The main result presented in this note was recorded in a more abstract, and apparently more obscure, setting in [**5**]. Moreover, in noncommutative algebraic geometry, certain adjoint functor pairs serve as morphisms between (not explicitly defined) noncommutative spaces. (This approach follows [**8**] and [**10**]; see also [**9**].)

The reader is referred, for example, to [**4**] for an accessible general introduction to category theory and its history.

2. A BRIEF REVIEW. As mentioned above, a category $\mathcal{C}$ consists of *objects* and *morphisms* between objects. The morphisms in $\mathcal{C}$ from an object A to an object B comprise a set denoted $\mathrm{Hom}_{\mathcal{C}}(A, B)$. A morphism in $\mathcal{C}$ from A to B is also referred to as a *$\mathcal{C}$-morphism* and denoted $A \to B$.

http://dx.doi.org/10.4169/amer.math.monthly.122.01.70
MSC: Primary 18A40, Secondary 54C05

These morphisms must satisfy the following conditions.

(1) For each pair of $\mathcal{C}$-morphisms $j : D \to E$ and $k : E \to F$, there is a *composition* morphism $k \circ j : D \to F$, such that

$$\ell \circ (k \circ j) \quad \text{and} \quad (\ell \circ k) \circ j$$

produce the same morphism $D \to G$, for all $\mathcal{C}$-morphisms $j : D \to E, k : E \to F$, and $\ell : F \to G$.

(2) For each object of $\mathcal{C}$ there is an identity morphism $\mathrm{id}_{\mathcal{C}}$ from that object to itself, such that the compositions

$$A \xrightarrow{\mathrm{id}_{\mathcal{C}}} A \xrightarrow{f} B \quad \text{and} \quad A \xrightarrow{f} B \xrightarrow{\mathrm{id}_{\mathcal{C}}} B$$

are both equal to f, for all $\mathcal{C}$-morphisms $f : A \to B$.

Adjoint functors. Let $\mathcal{C}$ and $\mathcal{D}$ be categories. A *(covariant) functor* $\Phi : \mathcal{C} \to \mathcal{D}$ assigns to each object A of $\mathcal{C}$ an object $\Phi(A)$ of $\mathcal{D}$, and to each $\mathcal{C}$-morphism $f : A \to B$ a $\mathcal{D}$-morphism

$$\Phi(f) : \Phi(A) \to \Phi(B),$$

such that

$$\Phi(\mathrm{id}_{\mathcal{C}}) = \mathrm{id}_{\mathcal{D}} \quad \text{and} \quad \Phi(\ell \circ k) = \Phi(\ell) \circ \Phi(k),$$

for all $\mathcal{C}$-morphisms $k : E \to F$ and $\ell : F \to G$.

Now consider a pair of functors

$$\Phi : \mathcal{C} \to \mathcal{D} \quad \text{and} \quad \Psi : \mathcal{D} \to \mathcal{C}.$$

Also suppose, for all objects L of $\mathcal{C}$ and M of $\mathcal{D}$, that there exists a bijective function

$$\mathrm{Hom}_{\mathcal{D}}(\Phi(L), M) \xrightarrow{\beta} \mathrm{Hom}_{\mathcal{C}}(L, \Psi(M)),$$

assigning to each $\mathcal{D}$-morphism

$$r : \Phi(L) \to M$$

a $\mathcal{C}$-morphism

$$\beta(r) : L \to \Psi(M).$$

We then say that (Φ, Ψ) is an *adjoint pair* provided

$$(*) \qquad \beta(t \circ r) = \Psi(t) \circ \beta(r) \quad \text{and} \quad \beta(r \circ \Phi(s)) = \beta(r) \circ s,$$

for all $\mathcal{C}$-morphisms $s : L' \to L$ and all $\mathcal{D}$-morphisms $t : M \to M'$. We indicate that the bijection β satisfies the two conditions in $(*)$ by saying that β is *natural in L and M*.

Example. The following is a standard first example of an adjoint pair: Let $\mathbb{F}$ be a field, and as above let $\mathcal{V}ec_{\mathbb{F}}$ denote the category whose objects are $\mathbb{F}$-vector spaces and whose morphisms are $\mathbb{F}$-linear transformations. Let $\Psi : \mathcal{V}ec_{\mathbb{F}} \to \mathcal{S}et$ be the "forgetful functor," assigning to each vector space its underlying set of vectors and assigning to each linear transformation its underlying set function. Let $\Phi : \mathcal{S}et \to \mathcal{V}ec_{\mathbb{F}}$ be the functor assigning to each set S the $\mathbb{F}$-vector space $\mathbb{F}S$ with basis S and assigning to each set map $S \to S'$ the $\mathbb{F}$-linear extension $\mathbb{F}S \to \mathbb{F}S'$. Then (Φ, Ψ) is an adjoint pair. (Details and analogous examples can be found, e.g., in Chapter IV of [**6**].)

3. CONTINUITY VIA ADJOINT PAIRS. For the remainder, let X and Y be topological spaces, and let $\varphi : X \to Y$ be a set function. Recall that φ will be continuous if and only if $\varphi^{-1}(V)$ is closed in X for all closed subsets V of Y.

The category of closed subsets of a topological space. Define $\mathcal{C}losed(X)$ to be the category whose objects are closed subsets of X and whose morphisms are described as follows. Let U and U' be closed subsets of X. If U is a subset of U' then there is exactly one morphism, the inclusion map, from U to U'. If U is not a subset of U' then the set of morphisms from U to U' is empty. Similarly define $\mathcal{C}losed(Y)$.

The category of *open* subsets of a topological space (see, e.g., [**6**, p. 35]) is basic to sheaf theory.

An adjointness criterion. Adjoint pairs of functors between the categories $\mathcal{C}losed(X)$ and $\mathcal{C}losed(Y)$ can be described in a particularly simple way, as follows. Suppose that

$$\Phi : \mathcal{C}losed(X) \to \mathcal{C}losed(Y) \quad \text{and} \quad \Psi : \mathcal{C}losed(Y) \to \mathcal{C}losed(X)$$

are functors. Then for U in $\mathcal{C}losed(X)$ and V in $\mathcal{C}losed(Y)$, there exists a bijection

$$\mathrm{Hom}_{\mathcal{C}losed(Y)}(\Phi(U), V) \quad \xrightarrow{\beta} \quad \mathrm{Hom}_{\mathcal{C}losed(X)}(U, \Psi(V))$$

exactly when one (and only one) of the following two cases holds.

Case 1: Both $\Phi(U) \subseteq V$ and $V \subseteq \Psi(V)$.
Case 2: Both $\Phi(U) \nsubseteq V$ and $U \nsubseteq \Psi(V)$.

It is not hard to verify that if a bijection β as above does exist, then it must be natural in U and V in the sense of $(*)$. (Also note that if β exists, it must be unique.) We deduce that (Φ, Ψ) is an adjoint pair exactly when the statement

$$(\dagger) \qquad \Phi(U) \subseteq V \quad \text{if and only if} \quad U \subseteq \Psi(V)$$

holds true for all U in $\mathcal{C}losed(X)$ and V in $\mathcal{C}losed(Y)$.

From a function to a pair of functors. Consider the assignments

$$T_\varphi : \mathcal{C}losed(X) \longrightarrow \mathcal{C}losed(Y), \quad U \longmapsto \overline{\varphi(U)},$$

and

$$T^\varphi : \mathcal{C}losed(Y) \longrightarrow \mathcal{C}losed(X), \quad V \longmapsto \overline{\varphi^{-1}(V)},$$

where $\overline{S}$ denotes the closure of an arbitrary subset of X or Y. It is straightforward to check that T_φ and T^φ are functors.

Our aim now is to prove the following.

Theorem. *The function φ is continuous if and only if (T_φ, T^φ) is an adjoint pair.*

Proof. To start, we claim that the following four conditions are equivalent for all U in $\mathcal{C}losed(X)$ and V in $\mathcal{C}losed(Y)$.

1. (T_φ, T^φ) is an adjoint pair.
2. $T_\varphi(U) \subseteq V$ if and only if $U \subseteq T^\varphi(V)$.
3. $\overline{\varphi(U)} \subseteq V$ if and only if $U \subseteq \overline{\varphi^{-1}(V)}$.
4. $\varphi(U) \subseteq V$ if and only if $U \subseteq \overline{\varphi^{-1}(V)}$.

The equivalence of (1), (2), and (3) follows directly from (†). To see why (3) is equivalent to (4), recall that the closure of $\varphi(U)$ in Y is the smallest closed subset containing $\varphi(U)$, and so

$$\varphi(U) \subseteq V \quad \text{if and only if} \quad \overline{\varphi(U)} \subseteq V.$$

Next, it is also true, for all U in $\mathcal{C}losed(X)$ and V in $\mathcal{C}losed(Y)$, that if $\varphi(U) \subseteq V$ then

$$U \subseteq \varphi^{-1}(\varphi(U)) \subseteq \varphi^{-1}(V) \subseteq \overline{\varphi^{-1}(V)}.$$

Hence, it follows from the equivalence of (1) and (4) above that (T_φ, T^φ) is an adjoint pair exactly when

$$(\ddagger) \qquad U \subseteq \overline{\varphi^{-1}(V)} \implies \varphi(U) \subseteq V,$$

for all U in $\mathcal{C}losed(X)$ and V in $\mathcal{C}losed(Y)$.

Now suppose that φ is a continuous function. As noted above, $\varphi^{-1}(V)$ is closed in X for all closed subsets V of Y, and so

$$\varphi^{-1}(V) = \overline{\varphi^{-1}(V)}.$$

Therefore, for all U in $\mathcal{C}losed(X)$ and V in $\mathcal{C}losed(Y)$, if

$$U \subseteq \varphi^{-1}(V) = \overline{\varphi^{-1}(V)},$$

then

$$\varphi(U) \subseteq \varphi(\varphi^{-1}(V)) \subseteq V.$$

Consequently, (T_φ, T^φ) is an adjoint pair.

Conversely, suppose that (T_φ, T^φ) is an adjoint pair, and fix an arbitrary V in $\mathcal{C}losed(Y)$. Set

$$U := \overline{\varphi^{-1}(V)}.$$

By (‡),

$$\varphi(U) \subseteq V,$$

and so

$$\overline{\varphi^{-1}(V)} = U \subseteq \varphi^{-1}(\varphi(U)) \subseteq \varphi^{-1}(V).$$

Therefore,

$$\varphi^{-1}(V) = \overline{\varphi^{-1}(V)}.$$

In particular, $\varphi^{-1}(V)$ is closed, and so φ is continuous. The theorem follows. ■

ACKNOWLEDGMENT. I am grateful to the referees for their helpful comments.

REFERENCES

1. B. Coecke, È. O. Paquette, Categories for the practising physicist, in *New Structures for Physics*. Edited by B. Coecke. Lecture Notes in Phys., Vol. 813, Springer-Verlag, Heidelberg, 2011, 173–286, http://dx.doi.org/10.1007/978-3-642-12821-9_3.
2. S. Eilenberg, S. MacLane, Natural isomorphisms in group theory, *Proc. Nat. Acad. Sci. USA* **28** (1942) 537–543, http://dx.doi.org/10.1073/pnas.28.12.537.
3. D. M. Kan, Adjoint functors, *Trans. Amer. Math. Soc.* **87** (1958) 294–329.
4. F. W. Lawvere, S. H. Schanuel, *Conceptual Mathematics: A First Introduction to Categories*. Second edition. Cambridge Univ. Press, UK, 2009, http://dx.doi.org/10.1017/cbo9780511804199.
5. E. S. Letzter, On continuous and adjoint morphism between non-commutative spectra, *Proc. Edinb. Math. Soc.* **49** no. 2 (2006) 367–381, http://dx.doi.org/10.1017/s0013091504000628.
6. S. Mac Lane, *Categories for the Working Mathematician*. Second edition. Graduate Texts in Mathematics, Vol. 5, Springer-Verlag, New York, 1998.
7. B. C. Pierce, *Basic Category Theory for Computer Scientists*. Foundations of Computing Series. MIT Press, Cambridge, MA, 1991.
8. A. L. Rosenberg, *Noncommutative Algebraic Geometry and Representations of Quantized Algebras*. Mathematics and its Applications. Vol. 330, Kluwer Academic Publishers, Dordrecht, 1995, http://dx.doi.org/10.1007/978-94-015-8430-2.
9. S. P. Smith, Maps between non-commutative spaces, *Trans. Amer. Math. Soc.* **356** (2004) 2927–2944.
10. M. Van den Bergh, Blowing up of non-commutative smooth surfaces, *Mem. Amer. Math. Soc.* **154** (2001), http://dx.doi.org/10.1090/memo/0734.

Department of Mathematics, Temple University, Philadelphia PA 19122
letzter@temple.edu

PROBLEMS AND SOLUTIONS

Edited by **Gerald A. Edgar, Doug Hensley, Douglas B. West**
with the collaboration of Itshak Borosh, Paul Bracken, Ezra A. Brown, Randall Dougherty, Tamás Erdélyi, Zachary Franco, Christian Friesen, Ira M. Gessel, László Lipták, Frederick W. Luttmann, Vania Mascioni, Frank B. Miles, Richard Pfiefer, Dave Renfro, Cecil C. Rousseau, Leonard Smiley, Kenneth Stolarsky, Richard Stong, Walter Stromquist, Daniel Ullman, Charles Vanden Eynden, Sam Vandervelde, and Fuzhen Zhang.

Proposed problems and solutions should be sent in duplicate to the MONTHLY *problems address on the back of the title page. Proposed problems should never be under submission concurrently to more than one journal, nor posted by the proposer to the internet before the due date for solutions. Submitted solutions should arrive before May 31, 2015. Additional information, such as generalizations and references, is welcome. The problem number and the solver's name and address should appear on each solution. An asterisk (*) after the number of a problem or a part of a problem indicates that no solution is currently available.*

PROBLEMS

11810. *Proposed by Ovidiu Furdui, Technical University of Cluj-Napoca, Cluj-Napoca, Romania.* Let $H_n = \sum_{k=1}^{n} 1/k$, and let ζ be the Riemann zeta function. Find

$$\sum_{n=1}^{\infty} H_n \left(\zeta(3) - \sum_{k=1}^{n} \frac{1}{k^3} \right).$$

11811. *Proposed by Vazgen Mikayelyan, Yerevan State University, Yerevan, Armenia.* Let $\langle a \rangle$ and $\langle b \rangle$ be infinite sequences of positive numbers. Let $\langle x \rangle$ be the infinite sequence given for $n \geq 1$ by

$$x_n = \frac{a_1^{b_1} \cdots a_n^{b_n}}{\left(\dfrac{a_1 b_1 + \cdots + a_n b_n}{b_1 + \cdots + b_n} \right)^{b_1 + \cdots + b_n}}.$$

(**a**) Prove that $\lim_{n \to \infty} x_n$ exists.
(**b**) Find the set of all c that can occur as that limit, for suitably chosen $\langle a \rangle$ and $\langle b \rangle$.

11812. *Proposed by Cristian Chiser, Craiova, Romania.* Let f be a twice continuously differentiable function from $[0, 1]$ into $\mathbb{R}$. Let p be an integer greater than 1. Given that $\sum_{k=1}^{p-1} f(k/p) = -\frac{1}{2}(f(0) + f(1))$, prove that

$$\left(\int_0^1 f(x)\, dx \right)^2 \leq \frac{1}{5!\, p^4} \int_0^1 (f''(x))^2\, dx.$$

http://dx.doi.org/10.4169/amer.math.monthly.122.01.75

11813. *Proposed by Greg Oman, University of Colorado-Colorado Springs, Colorado Springs, CO.* Let X be a set, and let $*$ be a binary operation on X (that is, a function from $X \times X$ to X). Prove or disprove: there exists an uncountable set X and a binary operation $*$ on X such that for any subsets Y and Z of X that are closed under $*$, either $Y \subseteq Z$ or $Z \subseteq Y$.

11814. *Proposed by Cezar Lupu, University of Pittsburgh, Pittsburgh, PA.* Let ϕ be a continuously differentiable function from $[0, 1]$ into $\mathbb{R}$, with $\phi(0) = 0$ and $\phi(1) = 1$, and suppose that $\phi'(x) \neq 0$ for $0 \leq x \leq 1$. Let f be a continuous function from $[0, 1]$ into $\mathbb{R}$ such that $\int_0^1 f(x)\,dx = \int_0^1 \phi(x) f(x)\,dx$. Show that there exists t with $0 < t < 1$ such that $\int_0^t \phi(x) f(x)\,dx = 0$.

11815. *Proposed by George Apostolopoulos, Messolonghi, Greece.* Let x, y, and z be positive numbers such that $x + y + z = 3$. Prove that

$$\frac{x^4 + x^2 + 1}{x^2 + x + 1} + \frac{y^4 + y^2 + 1}{y^2 + y + 1} + \frac{z^4 + z^2 + 1}{z^2 + z + 1} \geq 3xyz.$$

11816. *Proposed by Sabin Tabirca, University College Cork, Cork, Ireland.* Let ABC be an acute triangle, and let B_1 and C_1 be the points where the altitudes from B and C intersect the circumcircle. Let X be a point on arc BC, and let B_2 and C_2 denote the intersections of XB_1 with AC and XC_1 with AB, respectively. Prove that the line B_2C_2 contains the orthocenter of ABC.

SOLUTIONS

If the Sum of the Squares is the Square of the Sum, ...

11671 [2012, 800]. *Proposed by Sam Northshield, SUNY-Plattsburgh, Plattsburgh, NY.* Show that if relatively prime integers a, b, c, d satisfy

$$a^2 + b^2 + c^2 + d^2 = (a + b + c + d)^2,$$

then $|a + b + c|$ can be written as $m^2 - mn + n^2$ for some integers m and n.

Solution by Richard Stong, Center for Communications Research, San Diego, CA. Let $\omega = e^{2\pi i/3}$ be a primitive cube root of unity. Note that $m^2 - mn + n^2$ is the norm of $m + n\omega$ in the number ring $\mathbb{Z}[\omega]$. This ring is a unique factorization domain. The primes that split in this number ring are 3 and all primes congruent to 1 modulo 3. Thus a positive integer can be written in the form $m^2 - mn + n^2$ if and only if every prime congruent to 2 modulo 3 divides it an even number of times.

Let $g = \gcd(a + b + c, a + b + d, a + c + d, b + c + d)$. Now $(a + b + d) + (a + c + d) + (b + c + d) - 2(a + b + c) = 3d$ and symmetrically, and since $\gcd(a, b, c, d) = 1$, g is a divisor of 3.

Thus for any prime p congruent to 2 modulo 3 that divides $a + b + c$, we can choose one of $a + b + d$, $a + c + d$, and $b + c + d$ that is not divisible by p. Rewriting the given equality as

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

we see that p divides the right side with even multiplicity and hence divides $a + b + c$ with even multiplicity. By the remarks above, $a + b + c$ can be written in the form $m^2 - mn + n^2$ for some integers m and n.

Also solved by G. Apostolopoulos (Greece), R. Chapman (U. K.), P. P. Dályay (Hungary), Y. J. Ionin, O. P. Lossers (Netherlands), C. R. Pranesachar (India), M. A. Prasad (India), J. P. Robertson, T. Viteam (Chile), GCHQ Problem Solving Group (U. K.), and the proposer.

Carlson's Inequality

11680 [2012, 880]. *Proposed by Benjamin Bogoşel, University of Savoie, Savoie, France, and Cezar Lupu, University of Pittsburgh, Pittsburgh, PA.* Let $x_1, \ldots, x_n$ be nonnegative real numbers. Show that

$$\left(\sum_{i=1}^{n} \frac{x_i}{i}\right)^4 \leq 2\pi^2 \sum_{i,j=1}^{n} \frac{x_i x_j}{i+j} \sum_{k,l=1}^{n} \frac{x_k x_l}{(k+l)^3}.$$

Solution by Boukharfane Radouan, Quebec, Canada. This inequality is a direct application of the integral version of Carlson's inequality. Recall that this equality states that if f is a nonnegative function defined on $[0, \infty)$ such that $f(t)$ and $tf(t)$ are square-integrable, then

$$\left(\int_0^\infty f(t)\,dt\right)^4 \leq \pi^2 \left(\int_0^\infty (f(t))^2\,dt\right)\left(\int_0^\infty t^2 (f(t))^2\,dt\right).$$

For the current problem we apply Carlson's inequality to the function $f(t) = \sum_{k=1}^{n} x_k e^{-kt}$. Then we compute

$$\int_0^\infty f(t)\,dt = \sum_{k=1}^{n} x_k \int_0^\infty e^{-kt}\,dt = \sum_{k=1}^{n} \frac{x_k}{k},$$

$$\int_0^\infty (f(t))^2\,dt = \sum_{k,j=1}^{n} x_k x_j \int_0^\infty e^{-(k+j)t}\,dt = \sum_{k,j=1}^{n} \frac{x_k x_j}{k+j},$$

and

$$\int_0^\infty t^2 (f(t))^2\,dt = \sum_{k,j=1}^{n} x_k x_j \int_0^\infty t^2 e^{-(k+j)t}\,dt = 2\sum_{k,j=1}^{n} \frac{x_k x_j}{(k+j)^3}.$$

Putting these pieces together gives the desired inequality.

Editorial comment. Reference: F. Carlson, Une inégalité, *Ark. Mat. Astron. Fys.* **25B** (1934) 1–5. Some solvers provided Hardy's proof for Carlson's inequality.

Also solved by G. Apostolopoulos (Greece), P. Bracken, R. Chapman (U. K.), P. P. Dályay (Hungary), M. Omarjee (France), R. Stong, R. Tauraso (Italy), and the proposer.

Automorphisms Cannot One-Up their Group

11681 [2012, 880–881]. *Proposed by Des MacHale, University College Cork, Cork Ireland.* For any group G, let $\mathrm{Aut}(G)$ denote the group of automorphisms of G.
(**a**) Show that there is no finite group G with $|\mathrm{Aut}(G)| = |G| + 1$.
(**b**) Show that there are infinitely many finite groups G with $|\mathrm{Aut}(G)| = |G|$.
(**c**) Find all finite groups G with $|\mathrm{Aut}(G)| = |G| - 1$.

Solution by the Missouri State University Problem Solving Group, Missouri State University, Springfield, MO. For **(b)**, it is well known that $\mathrm{Aut}(S_n) \cong S_n$ when $n \notin \{2, 6\}$, so $\{S_n : n \notin \{2, 6\}\}$ is such an infinite family.

Now consider **(a)** and **(c)**. Let $\mathrm{Inn}(G)$ denote the group of inner automorphisms of G, that is, the group of mappings τ_b defined by $\tau_b(x) = bxb^{-1}$. Let $Z(G)$ be the

center of G. An elementary group-theoretic argument shows that $\mathrm{Inn}(G) \cong G/Z(G)$, so $|\mathrm{Inn}(G)|$ divides $|G|$. Since $\mathrm{Inn}(G)$ is a (normal) subgroup of $\mathrm{Aut}(G)$, the size of $\mathrm{Inn}(G)$ divides $|\mathrm{Aut}(G)|$. In **(a)** and **(c)**, $|\mathrm{Inn}(G)|$ divides two relatively prime integers, so $|\mathrm{Inn}(G)| = 1$. Hence G is Abelian.

We claim that also G is cyclic. If not, then $G \cong \mathbb{Z}_{p^\alpha} \oplus \mathbb{Z}_{p^\beta} \oplus H$ with p a prime and $1 \le \alpha \le \beta$. Define $f\colon G \to G$ by $f(x, y, z) = (x + y, y, z)$. This is an automorphism of order p^α, so p^α divides both $|G|$ and $|\mathrm{Aut}(G)|$. From the contradiction $p^\alpha \nmid 1$, we conclude that G is cyclic.

Since G is cyclic, $|\mathrm{Aut}(G)| = \varphi(G) < |G|$, so **(a)** cannot occur.

We claim that case **(c)** can occur if and only if G is cyclic of prime order. If $G \cong \mathbb{Z}_p$ with p a prime, then $|\mathrm{Aut}(G)| = \varphi(p) = p - 1 = |G| - 1$, as claimed. Otherwise, $|G| = n = pk$ with p a prime and $k > 1$. Now both p and $2p$ do not exceed n and are not relatively prime to n; hence $|\mathrm{Aut}(G)| = \varphi(|G|) < |G| - 1$. Thus if G is not cyclic of prime order, then $|\mathrm{Aut}(G)| < |G| - 1$.

Editorial comment. Bruce Burdick used similar ideas to prove that $|\mathrm{Aut}(G)| = |G| + 2$ if and only if $G \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$, and $|\mathrm{Aut}(G)| = |G| - 2$ if and only if $G \cong \mathbb{Z}_4$.

Also solved by A. J. Bevelacqua, R. Black & A. Lizzi & N. Monson, P. Budney, B. Burdick, R. Chapman (U. K.), P. P. Dályay (Hungary), D. Fleischman, S. M. Gagola Jr., O. Geupel (Germany) (part (b) only), N. Grivaux (France), Y. J. Ionin, J Konieczny, C. Lanski, C. Leuridan (France), J. H. Lindsey II, O. P. Lossers (Netherlands), C. P. Rupert, J. H. Smith, R. Stong, D. Tyler, the GCHQ Problem Solving Group (U. K.), TCDmath Problem Group (Ireland), NSA Problems Group, and the proposer.

An Alternating Sum of Squares of Alternating Sums

11682 [2012, 881]. *Proposed by Ovidiu Furdui, Technical University of Cluj-Napoca, Cluj-Napoca, Romania.* Compute

$$\sum_{n=0}^{\infty} (-1)^n \left(\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{n+k} \right)^2.$$

Solution by Brian Bradie, Christopher Newport University, Newport News, VA. The sum equals $\pi^2/24$.

Since

$$\int_0^1 \frac{x^n}{1+x}\,dx = \int_0^1 x^n \left(\sum_{k=1}^{\infty} (-1)^{k-1} x^{k-1} \right) dx = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{n+k},$$

we have

$$\begin{aligned}
\sum_{n=0}^{\infty} (-1)^n \left(\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{n+k} \right)^2 &= \sum_{n=0}^{\infty} (-1)^n \left(\int_0^1 \frac{x^n}{1+x}\,dx \right)^2 \\
&= \sum_{n=0}^{\infty} (-1)^n \int_0^1 \frac{x^n}{1+x}\,dx \int_0^1 \frac{y^n}{1+y}\,dy = \sum_{n=0}^{\infty} (-1)^n \int_0^1 \int_0^1 \frac{x^n y^n}{(1+x)(1+y)}\,dy\,dx \\
&= \int_0^1 \int_0^1 \frac{1}{(1+x)(1+y)(1+xy)}\,dy\,dx.
\end{aligned}$$

Now

$$\int_0^1 \frac{1}{(1+y)(1+xy)}\,dy = \left. \frac{\ln(1+y) - \ln(1+xy)}{1-x} \right|_{y=0}^{y=1} = -\frac{\ln((1+x)/2)}{1-x},$$

so

$$\int_0^1 \int_0^1 \frac{1}{(1+x)(1+y)(1+xy)}\,dy\,dx = -\int_0^1 \frac{\ln((1+x)/2)}{(1-x)(1+x)}\,dx$$
$$= -\frac{1}{2}\int_0^1 \frac{\ln((1+x)/2)}{1+x}\,dx - \frac{1}{2}\int_0^1 \frac{\ln((1+x)/2)}{1-x}\,dx. \qquad (*)$$

For the first term in $(*)$, we compute

$$-\frac{1}{2}\int_0^1 \frac{\ln((1+x)/2)}{1+x}\,dx = -\frac{1}{4}\ln^2\left(\frac{1+x}{2}\right)\Bigg|_0^1 = \frac{1}{4}\ln^2 2.$$

For the second term in $(*)$, the substitution $u = (1-x)/2$ yields

$$-\frac{1}{2}\int_0^1 \frac{\ln((1+x)/2)}{1-x}\,dx = \frac{1}{2}\int_{1/2}^0 \frac{\ln(1-u)}{u}\,du = \frac{1}{2}\left(\frac{\pi^2}{12} - \frac{\ln^2 2}{2}\right).$$

(For the last step, if h is the integrand, think about the integral of h over $(0, 1)$ and $(1/2, 1)$, and use integration by parts.) Therefore

$$\sum_{n=0}^{\infty}(-1)^n\left(\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{n+k}\right)^2 = \frac{1}{4}\ln^2 2 + \frac{1}{2}\left(\frac{\pi^2}{12} - \frac{1}{2}\ln^2 2\right) = \frac{\pi^2}{24}.$$

Also solved by U. Abel (Germany), D. Beckwith, M. Benito & Ó. Ciaurri & E. Fernández & L. Roncal (Spain), R. Boukharfane (Canada), K. N. Boyadzhiev, B. Burdick, R. Chapman (U. K.), H. Chen, P. P. Dályay (Hungary), O. Geupel (Germany), M. L. Glasser, J. P. Grivaux (France), O. Kouba (Syria), O. P. Lossers (Netherlands), O. Oloa (France), M. Omarjee (France), P. Perfetti (Italy), A. Stenger, R. Stong, R. Tauraso (Italy), D. B. Tyler, J. Van Hamme (Belgium), M. Vowe (Switzerland), GWstat Problem Solving Group, and the proposer.

Special Gergonne Points

11683 [2012, 881]. *Proposed by Raimond Struble, Santa Monica, CA.* Given a triangle ABC, let F_C be the foot of the altitude from the incenter to AB. Define F_B and F_C similarly. Let C_A be the circle with center A that passes through F_B and F_C, and define C_B and C_C similarly. The *Gergonne point* of a triangle is the point at which segments AF_A, BF_B, and CF_C meet. Determine, up to similarity, all isosceles triangles such that the Gergonne point of the triangle lies on one of the circles C_A, C_B, or C_C.

Solution by Bruce S. Burdick, Roger Williams University, Bristol, RI. The triangle is isosceles. Assume that $\angle B$ and $\angle C$ are congruent. Then $\overline{AF_A}$ is a line of symmetry of the triangle, so it is perpendicular to BC, and it is a common tangent to C_B and C_C. Thus the Gergonne point, call it G, can lie on either C_B or C_C only if it coincides with F_A, but that implies that ABC is a degenerate triangle. Thus G is on C_A.

Let the lengths of the sides opposite A, B, C be a, b, c, respectively. Let $s = \frac{1}{2}(a + b + c)$. The distance from A to F_B is $s - a$, so the radius of C_A is $s - a$, and the distance from A to G is $s - a$. Also, $\overline{AF_A}$ is the bisector of $\angle A$. The observation that

$$\text{Area}(ABF_B) = \text{Area}(ABG) + \text{Area}(AGF_B)$$

allows us to write

$$\frac{1}{2}\,c(s-a)\sin A = \frac{1}{2}\,c(s-a)\sin\frac{A}{2} + \frac{1}{2}(s-a)^2\sin\frac{A}{2},$$
$$c\sin\frac{A}{2}\cos\frac{A}{2} = \frac{1}{2}(s-a+c)\sin\frac{A}{2},$$
$$c\cos\frac{A}{2} = \frac{1}{4}(b+c-a+2c).$$

Since $\triangle ABF_A$ is a right triangle and $b = c$, we have

$$\sqrt{c^2 - \frac{1}{4}a^2} = c - \frac{1}{4}a, \quad c^2 - \frac{1}{4}a^2 = c^2 - \frac{1}{2}ac + \frac{1}{16}a^2,$$
$$0 = \frac{5}{16}a^2 - \frac{1}{2}ac, \quad a = \frac{8}{5}c.$$

It follows that if the Gergonne point of an isosceles triangle lies on one of the circles C_A, C_B, C_C, then the three sides, in some order, are in the ratio $5 : 5 : 8$.

Also solved by R. Boukharfane (Canada), P. P. Dályay (Hungary), C. Delorme (France), A. Ercan (Turkey), O. Geupel (Germany), M. Goldenberg & M. Kaplan, J.-P. Grivaux (France), K. Hanes, A. Johnston, N. Komanda, J. H. Lindsey II, O. P. Lossers (Netherlands), J. Minkus, C. P. Pranesachar (India), R. Stong, M. Vowe (Switzerland), H. Widmer (Switzerland), J. Zacharias, GCHQ Problem Solving Group (U. K.), and the proposer.

Möbius Estimates

11684 [2013, 76]. *Proposed by Raymond Mortini, Université Paul Verlaine, Metz, France, and Rudolf Rupp, Georg-Simon-Ohm Hochschule Nürnberg, Nuremberg, Germany.* For complex a and z, let

$$\phi_a(z) = \frac{a-z}{1-\overline{a}z}, \quad \rho(a,z) = \frac{|a-z|}{|1-\overline{a}z|}.$$

(**a**) Show that whenever $-1 < a, b < 1$,

$$\max_{|z|\le 1} |\phi_a(z) - \phi_b(z)| = 2\rho(a,b)$$
$$\max_{|z|\le 1} |\phi_a(z) + \phi_b(z)| = 2.$$

(**b**) For complex α, β with $|\alpha| = |\beta| = 1$, let

$$m(z) = m_{a,b,\alpha,\beta}(z) = |\alpha\phi_a(z) - \beta\phi_b(z)|.$$

Determine the maximum and minimum, taken over z with $|z| = 1$, of $m(z)$.

Solution by the proposers.
(**b**) Observe that ϕ_a is its own inverse. Let $c = (b-a)/(1-a\overline{b})$ and let

$$\lambda = -\frac{1-a\overline{b}}{1-\overline{a}b}.$$

Since ϕ_b is a bijection of the unit circle onto itself,

$$\max_{|z|=1} \left|\alpha\phi_a(z) - \beta\phi_b(z)\right| = \max_{|z|=1} \left|\alpha\overline{\beta}\phi_a\big(\phi_b(z)\big) - z\right| = \max_{|z|=1} \left|\alpha\overline{\beta}\lambda\phi_c(z) - z\right|.$$

The same identities hold when the maximum is replaced by the minimum. Put $\gamma = \alpha\overline{\beta}\lambda$, and let $-\pi < \arg\gamma \le \pi$. For $|z| = 1$, let $H(z) = |\gamma\phi_c(z) - z|$. We have

$$H(z) = \left|\gamma\,\frac{z(c\overline{z}-1)}{1-\overline{c}z} - z\right| = \left|\gamma\,\frac{1-c\overline{z}}{1-\overline{c}z} - 1\right| = \left|\gamma\,\frac{w}{\overline{w}} + 1\right|,$$

where $w = 1 - c\overline{z} = 1 - c/z$. As z moves around the unit circle, w moves around the circle $|w-1| = |c|$. Write $w = |w|e^{i\theta}$. Note that θ varies on the interval $[-\theta_m, \theta_m]$, where $|\theta_m| < \pi/2$ and $\sin\theta_m = |c| = \rho(a,b)$. Now

$$H(z) = \left|\gamma e^{2i\theta} + 1\right| = 2\left|\cos\left(\frac{\arg\gamma}{2} + \theta\right)\right|.$$

Hence

$$\max_{|z|=1} H(z) = 2\max\left\{\left|\cos\left(\frac{\arg\gamma}{2} + \theta\right)\right| : |\theta| \le \arcsin\rho(a,b)\right\} \qquad (*)$$

and

$$\min_{|z|=1} H(z) = 2\min\left\{\left|\cos\left(\frac{\arg\gamma}{2} + \theta\right)\right| : |\theta| \le \arcsin\rho(a,b)\right\}.$$

(**a**) Specialize $(*)$ by taking $a, b \in (-1,1)$ and $\alpha = \beta = 1$, so that $\gamma = -1$. By the maximum principle, the maximum on the disk is achieved on the boundary, so

$$\max_{|z|\le 1} |\phi_a(z) - \phi_b(z)| = 2\max\left\{|\sin\theta| : |\theta| \le \arcsin\rho(a,b)\right\} = 2\rho(a,b).$$

For the other part of (**a**), instead specialize $(*)$ by taking $a, b \in (-1,1)$ and $\alpha = 1$, $\beta = -1$, so that $\gamma = 1$. This gives

$$\max_{|z|\le 1} |\phi_a(z) + \phi_b(z)| = 2\max\left\{|\cos\theta| : |\theta| \le \arcsin\rho(a,b)\right\} = 2.$$

Also solved by P. P. Dályay (Hungary) and R. Stong. Part (a) only by A. Alt, D. Beckwith, D. Fleischman, O. P. Lossers (Netherlands), and T. Smotzer.

The Reciprocal of the Thue-Morse Constant

11685 [2013, 76]. *Proposed by Donald Knuth, Stanford University, Stanford, CA.* Prove that

$$\prod_{n=0}^{\infty}\left(1 + \frac{1}{2^{2^k}-1}\right) = \frac{1}{2} + \sum_{k=0}^{\infty}\frac{1}{\prod_{j=0}^{k-1}\left(2^{2^j}-1\right)}.$$

In other words, prove that

$$(1+1)\left(1+\tfrac{1}{3}\right)\left(1+\tfrac{1}{15}\right)\left(1+\tfrac{1}{255}\right)\cdots = \tfrac{1}{2} + 1 + 1 + \tfrac{1}{3} + \tfrac{1}{3\cdot 15} + \tfrac{1}{3\cdot 15\cdot 255} + \cdots.$$

Solution by Traian Viteam, Punta Arenas, Chile. For $n \ge 0$,

$$\prod_{k=0}^{n}\left(1 + \frac{1}{2^{2^k}-1}\right) - \prod_{k=0}^{n-1}\left(1 + \frac{1}{2^{2^k}-1}\right) = \prod_{k=0}^{n-1} 2^{2^k} \Big/ \left(\prod_{k=0}^{n} 2^{2^k} - 1\right)$$

$$= 2^{2^n-1} \Big/ \left(\prod_{k=0}^{n} 2^{2^k} - 1\right) = \frac{1}{2}\left(\frac{1}{\prod_{j=0}^{n-1}(2^{2^j}-1)} + \frac{1}{\prod_{j=0}^{n}(2^{2^j}-1)}.\right)$$

Summing from $n = 0$ to $n = N$ yields

$$\prod_{k=0}^{N}\left(1+\frac{1}{2^{2^k}-1}\right)-1=\frac{1}{2}+\sum_{k=1}^{N}\frac{1}{\prod_{j=0}^{k-1}(2^{2^j}-1)}+\frac{1}{2}\frac{1}{\prod_{j=0}^{N}(2^{2^j}-1)}$$

for all N. Letting N tend to infinity yields the desired result.

Editorial comment. The proposer noted that this is the special case $x = 1/2$ of

$$\frac{1}{\prod_{k=0}^{\infty}\left(1-x^{2^k}\right)}=1-x+2\sum_{k=0}^{\infty}\frac{x^{2^k}}{\prod_{j=0}^{k-1}\left(1-x^{2^j}\right)}.$$

The left side is the reciprocal of the generating function $\mu(x)$ of the Thue-Morse sequence, and $\mu(1/2)$ is the Thue-Morse constant, which is the subject of Section 6.8 in *Mathematical Constants* by Steven R. Finch, Cambridge University Press (2003), pp. 436–441.

Also solved by R. Barnes, D. Beckwith, R. Boukharfane (Canada), B. Burdick, R. Chapman (U. K.), J. Fabrykowski & T. Smotzer, O. Geupel (Germany) C. Georghiou (Greece), Y. J. Ionin, O. Kouba (Syria), K. Kyun (Korea), J. H. Lindsey II, O. P. Lossers (Netherlands), R. Martin (Germany), J. Martinez (Spain), M. Omarjee (France), H. Roelants (Belgium), R. Sachdev (India), J. Schlosberg, R. Tauraso (Italy), M. Wildon (U. K.), BSI Problems Group (Germany), GCHQ Problem Solving Group (U. K.), TCDmath Problem Group (Ireland), and the proposer.

A Fast-Growing Function

11688 [2013, 77]. *Proposed by Samuel Alexander, The Ohio State University, Columbus, OH.* Consider $f\colon \mathbb{N} \to \mathbb{N}$ such that $\lim_{a\to\infty} \inf_{b,c,d\in\mathbb{N},b<a} f(a, c, d) - f(b, c, d) = \infty$. Show that for $B \in \mathbb{N}$, there exists $k \in \mathbb{N}$ such that

$$f(a, c, d) = k \quad \Rightarrow \quad \max\{c, d\} > B.$$

Solution by Iosif Pinelis, Michigan Technological University, Houghton, MI. We say that a subset S of $\mathbb{N}$ has *density zero* if

$$\lim_{n\to\infty}\frac{1}{n}|S\cap[n]|=0,$$

where $[n] = \{1, \ldots, n\}$.

First we show that if $h : \mathbb{N} \to \mathbb{N}$ is a function satisfying $\lim_{a\to\infty} h(a) - h(a - 1) = \infty$, then $h(\mathbb{N})$ has density zero. For a positive integer m, there exists $a_m \in \mathbb{N}$ such that $h(a) - h(a - 1) \geq m$ for $a > a_m$. Hence for all $n \in \mathbb{N}$,

$$|h(\mathbb{N})\cap[n]|\leq|h([a_m])|+\frac{n}{m}+1,$$

and thus $\limsup_{n\to\infty} \frac{1}{n}|h(\mathbb{N}) \cap [n]| \leq 1/m$. Since this holds for all m, it follows that $h(\mathbb{N})$ has density zero.

Now the given hypothesis implies for fixed $c, d \in \mathbb{N}$ that $\lim_{a\to\infty} f(a, c, d) - f(a - 1, c, d) = \infty$, and thus the set $S_{c,d} = \{f(a, c, d)\colon a \in \mathbb{N}\}$ has density 0. Since the union of finitely many sets of density zero has density zero, for any $B \in \mathbb{N}$ the set $\bigcup_{c,d\leq B} S_{c,d}$ has density zero. Therefore some $k \in \mathbb{N}$ is not in this set, so $f(a, c, d) = k$ implies $\max\{c, d\} > B$.

Also solved by R. Chapman (U. K.), O. Geupel (Germany), B. Karaivanov, O. P. Lossers (Netherlands), R. Martin (Germany), R. Stong, H. Takeda (Japan), GCHQ Problem Solving Group (U. K.), TCDmath Problem Group (Ireland), and the proposer.

REVIEWS

Edited by **Jeffrey Nunemacher**
Mathematics and Computer Science, Ohio Wesleyan University, Delaware, OH 43015

The Mathematics of Encryption: An Elementary Introduction. By Margaret Cozzens and Steven J. Miller. American Mathematical Society, Providence, RI, 2013, xviii + 332 pp., ISBN 978-0-8218-8321-1, $49.00.

Reviewed by **Edward F. Schaefer**

The development of modern cryptography is the story of dealing with the problem of eavesdroppers in digital communications. A lot of beautiful mathematics is being used to solve these problems, and the solutions themselves motivate mathematics that does not yet exist.

An introductory cryptography course can be taught as a story. During the course, the student learns the basic building blocks of cryptography: symmetric key cryptography, public key cryptography for key agreement, public key cryptography for digital signatures, and hash functions. In the culmination of the course, the student learns how these four building blocks make up the protocol called Transport Layer Security (or TLS). This protocol (or something like it) is used whenever anything secure happens on the web or when you use an ATM.

Most textbooks on cryptography, that are written from a mathematical point of view, do not tell this story or do not tell it well. They focus on the two topics involving public key cryptography. These topics are elegant and use very interesting mathematics. Typically short shrift (or no shrift) is given to hash functions and modern symmetric key cryptography—the latter being the workhorse of cryptography. One can easily memorize the algorithms that implement public key cryptography for key agreement and signatures. The algorithms used for symmetric key cryptography and hash functions are so complicated that I never have students memorize them for exams. This is probably one of the reasons the mathematicians who write these books do not describe these algorithms. However, one needs to see such algorithms in order to get sufficient intuition into modern symmetric key cryptography and hash functions. In addition, both actually involve some interesting mathematics (admittedly not as interesting as that used for public key cryptography). Most mathematician authors are choosing beauty over giving readers a thorough understanding of the important components of cryptography, as it is used. Let us look at the story, see what kinds of mathematics arise in it, and compare how popular textbooks written from a mathematical viewpoint tell this story.

Until the mid-1970s, there was simply symmetric key cryptography. Alice and Bob want to exchange encrypted messages. They choose an encrypting function $f_e(x, y)$, where x represents possible messages to be encrypted (or *plaintexts*) and y represents possible encrypting keys. They must also agree on an encrypting key k_e. If M is a plaintext message, then the output $f_e(M, k_e)$ is called the encryption or *ciphertext*; it is the ciphertexts that get transmitted. There should be a decrypting function $f_d(x, y)$

http://dx.doi.org/10.4169/amer.math.monthly.122.01.83

(easily determined from f_e) and a decrypting key k_d such that for all plaintexts x, we have $f_d(f_e(x, k_e), k_d) = x$. It is assumed that for an eavesdropper, knowledge of f_e and the values $f_e(x, k_e)$, for many x's and a given k_e, is not enough to determine k_e, k_d, or any of the x's. For symmetric key cryptography, the value of k_d should be the same as k_e (true of modern systems) or can be determined from k_e in polynomial time. Informally, we can define a *symmetric key cryptosystem* to be the pair f_e and f_d.

Modern symmetric key cryptosystems consist of block ciphers (more popular) and stream ciphers. Popular block ciphers include DES, triple-DES, and AES, the Advanced Encryption Standard, which is the current world standard for symmetric key cryptography. Currently, for the most common implementation of AES, the plaintexts, ciphertexts, and keys are blocks of 128 bits. One can use ASCII to encode a plaintext message as a string of bits—there is nothing secret about encoding and decoding. A message is padded so that its length is a multiple of 128 and each block is encrypted. A permutation function takes a string of bits and replaces it with a permutation of the bits in the string. For a substitution function, each outbit bit is a function of multiple input bits. In World War II it was found that alternating substitution functions with permutation functions creates a strong block cipher, and this is how block ciphers are designed today. Some of the component substitution functions within AES involve some nice mathematics. For example, the only nonlinear part of AES comes from the following function from bytes to bytes: Encode a byte as the coefficients of a polynomial in x of degree 7. Assuming it was not the byte consisting of eight 0's (which is sent to itself), find the multiplicative inverse of that polynomial in the finite field $\mathbf{F}_2[x]/(x^8 + x^4 + x^3 + x + 1)$ and decode as a byte (see [**4**, p. 36]). Another component substitution function (which also provides diffusion) comes from encoding four bytes as the coefficients of a polynomial in $\mathbf{F}_2[x, z]/(x^8 + x^4 + x^3 + x + 1, z^4 + 1)$, a finite ring that is not an integral domain. This polynomial is then multiplied by a fixed polynomial in that ring that has a multiplicative inverse. The product is then decoded to four bytes (see [**4**, p. 39]). Computer scientists use x for both variables in that ring, but they know what they mean.

A stream cipher, in its simplest form, requires that Alice and Bob agree on a pseudo-random bit generator and a seed, the seed acting as the key. For encryption, the output of the pseudo-random bit generator is XORed with the plaintext message to create the ciphertext message. For decryption, the output of the pseudo-random bit generator is XORed with the ciphertext to get the plaintext. If we XOR the plaintext with a truly random string, then we have provable perfect security; this is called a one time pad—it is extremely inefficient.

A currently popular stream cipher is RC4 (RC stands for *Ron's Code*, Ron being Ron Rivest—the R of RSA). In one implementation, we start with $S_0 = 0, \ldots, S_{255} = 255$. After some initial permutation of the values of the S_i's, involving the key, the following pseudocode creates ℓ pseudo-random bytes:

$i = 0$; $j = 0$
for $r = 0, \ldots, \ell - 1$ do
  $i = i + 1 \pmod{256}$
  $j = j + S_i \pmod{256}$
  swap S_i and S_j
  $t = S_i + S_j \pmod{256}$
  $R_r = S_t$
end for

Then the concatenation $R_0 R_1 R_2 \ldots$ is the output of the pseudo-random bit generator (see [**10**, pp. 397–398]).

The ordered values of the S_i's form a vector. Consider the intriguing way that an index within a vector and the value of one of its components are added. We do not yet have mathematics that can undo repetitions of such a process, which could help in the cryptanalysis. Bruce Schneier invented a pseudo-random number generator using a deck of cards. It is used for a stream cipher and is presented in Neal Stephenson's novel *Cryptonomicon* [**13**]. Similar to RC4, the position of a card in a deck and the value of a card interact.

As safe as modern symmetric key cryptosystems are, there is the problem that Alice and Bob need to agree on the key ahead of time. This requires copresence or Alice sending Bob the key using insecure methods. Neither of these is acceptable if you want to encrypt your credit card information for a web purchase from a distant vendor. This problem was solved in the mid-1970s using public key cryptography. In public key cryptography, each user has their own private key. From it they can create a public key in polynomial time. However, there is no known polynomial time algorithm for getting the private key from the public key. Getting the private key from the public key should require solving some difficult computational problem like factoring, the discrete logarithm problem in the multiplicative group of a finite field, or the discrete logarithm problem in the group of points on an elliptic curve with coordinates in a finite field. Alice and Bob can use public key cryptography to agree on a key for a symmetric key cryptosystem.

In some public key cryptosystems (like RSA), there is an encrypting function $p_e(x, y)$, where x represents possible plaintexts and y represents possible public keys. Bob makes public his public key k_B. To encrypt a message M (usually a key for a symmetric key cryptosystem that Alice will soon use with Bob), she computes $p_e(M, k_B)$, the ciphertext, and sends it to Bob. There should be a function $p_d(x, y)$, easily determined from p_e, such that if k' is any private key, then for all plaintexts x, we have $p_d(p_e(x, k_B), k'_B) = x$. Note that RSA is the most popular public key cryptosystem, not for security or efficiency reasons, but because of the marketing ability of Jim Bidzos, the former President of RSA Security.

You might wonder why one should bother with the symmetric key cryptosystem at this point, since Alice could encode her plaintext as one or more messages and encrypt each as above with the public key encrypting function. The reason is that public key cryptography is, in practice, much slower than symmetric key cryptography.

The mathematics used in the RSA cryptosystem consists of finding large primes, evaluating Euler's totient function (which gives the size of the unit group of the ring $\mathbf{Z}/n\mathbf{Z}$ in which one works) and computing modular inverses and exponentiation (see [**6**, Ch. IV, Sec. 2]). The Diffie–Hellman key agreement system for finite fields uses a large finite field of the form $\mathbf{F}_q$, where q is a power of 2 or a prime. There is also the need to find an element of the unit group of $\mathbf{F}_q$ of large prime order (this must exist for $\mathbf{F}_q$ to be useable) and exponentiation in the finite field (see [**10**, pp. 513–514]). The Diffie–Hellman key agreement system for elliptic curves uses the group of points of an elliptic curve with coordinates in a large finite field of one of the same two types as above (see [**6**, Ch. VI, Sec. 2]). Again, an element must be found with large prime order (this must exist for the elliptic curve to be useable) and there is exponentiation in the group. More gorgeous number theory and arithmetic geometry arise in finding useable finite fields and elliptic curves, improvements in implementation, and cryptanalytic attempts to crack such cryptography.

By solving the problem of how to agree remotely on a key for symmetric key cryptography we introduce a new problem. In the old days, when Alice and Bob met to agree on a shared key for symmetric key cryptography, they were assured of each other's identities. In the above scenario with a public encryption function and key

being used to agree on a key for symmetric key cryptography, how can Bob be sure that encrypted messages are actually coming from Alice? This problem is solved using digital signatures (as well as certificates), which also use public key cryptography.

One way of implementing digital signatures is to exploit public key cryptosystems using functions, as described earlier. Assume for any public and private key pair k, k' and any string x that $p_e(p_d(x, k'), k) = x$. If k_A and k'_A are Alice's public and private keys, then Alice can digitally sign the message M by computing and making public $p_d(M, k'_A)$. Of course Bob can verify the signature by confirming that $p_e(p_d(M, k'_A), k_A)$ and M are the same. This connects the message M to whomever has the k'_A associated to the public key k_A. Certificates connect public keys to actual entities (people, corporations, etc.) and also use digital signatures.

We need to consider what exactly should be signed. If Alice has encrypted a message using symmetric key cryptography, then she can sign the entire ciphertext. But that would be slow, and there is the additional issue that the ciphertext and signature could be tampered with during transmission (perhaps corresponding blocks of each are removed). The solution to this is a hash algorithm, which is derived from a hash function using the Merkle–Damgård construction (see [**9**, pp. 13–15]). A hash function $h(x, y)$ takes as inputs an m bit string and a t bit string and outputs a t bit string. To create a hash algorithm, a message M is encoded as a bit string and padded so that its length is a multiple of m. It is broken into m bit blocks $M_1, \ldots, M_r$. There is a t bit initialization vector; if that is a secret, shared by Alice and Bob, then the output hash is called a message authentication code or MAC. Let y_0 be the initialization vector. For $1 \leq i \leq r$ define $y_i = h(M_i, y_{i-1})$. Then y_r is the output hash or MAC. What typically gets signed is the hash (or MAC) of a plaintext message. We will elaborate on this when we describe TLS.

A good cryptographic hash algorithm has three properties, two of which we will not describe (the weakly and strongly collision-free properties). Most important for cryptography is the property of being one way; this means that when challenged with a t-bit string z, it is practically impossible to find any message M such that $h(M, y_0) = z$. Popular hash algorithms include SHA-i, for $i = 0, 1, 2, 3$ (and SHA stands for Secure Hash Algorithm), and MD5. Note that SHA-0, SHA-1, and MD5 have been recently shown not to have the strongly collision free property (see [**15**]); that does not make them unsafe for use in TLS.

The mathematics involved in popular hash functions includes alternating XORing m-bit strings with adding the integers they represent modulo 2^m (something else mathematics is not yet ready to undo), nonlinear functions on bit strings where, like a linear function, each output bit is equally likely to be a 0 or 1 (see [**10**, Ch. 18, Sec. 5]), and exponentiation of a 2×2 matrix modulo 5 (see [**1**, p. 8]). Hash functions tend to be even less elegant than block ciphers. This lack of beauty not only leads to their being left out of cryptography textbooks; it is one of the reasons there is a dearth of experts. One expert confided to me, during the competition leading to the new SHA-3, that there were so few people interested in hash functions that most of the candidates were disappointingly unremarkable.

Only after explaining the four basic building blocks of cryptography are we ready to give to a class a simplified explanation of how secure communication occurs online (see [**12**, Ch. 17, Sec. 2]). Alice and Bob (really their computers) arrange to communicate using TLS. We will assume that this implementation of TLS uses a public key cryptosystem that involves encrypting and decrypting functions, as described above, and specifies a block symmetric key cryptosystem and hash function to be used.

Alice and Bob first trade certificates and check each other's. Bob creates a key for the symmetric key cryptosystem and a MAC key and concatenates them. Bob finds

Alice's public key on her certificate and uses the public key encryption function to encrypt the concatenation for her and sends that encryption to Alice. Alice decrypts that and so now has the two keys. Alice takes the plaintext message she wants to send to Bob, breaks it into blocks, encrypts each block with the symmetric key cryptosystem and sends the corresponding ciphertexts to Bob. Alice then uses the MAC key and her plaintext message and sends that through the hash algorithm to get the output MAC. She takes the MAC and signs it using her private key. She takes the signed MAC and encrypts that using the symmetric key cryptosystem. We will call the output the encrypted signed MAC. (Don't you just love it?—a single object that incorporates almost an entire cryptography course.) She sends the encrypted signed MAC to Bob.

Bob now uses the symmetric key cryptosystem to decrypt the ciphertexts to get the plaintext message. He then uses the MAC key and sends the plaintext message he got through the hash algorithm. We will refer to the output as the putative MAC. Bob then uses the symmetric key cryptosystem and then Alice's public key on the encrypted signed MAC to get the MAC she computed. He then compares Alice's MAC with the putative MAC. If they are the same, then he knows that the ciphertexts were not tampered with and that it was Alice who sent all of this to him. If Bob now wants to respond to the message, he does the same thing Alice did, starting with the symmetric key encryption of his message. Think about this the next time you see https: in front of a URL in a browser or while you listen to the whirring of an ATM after you enter your card.

We now compare how textbooks on cryptography, written from a mathematical viewpoint, tell this story. Cozzens and Miller's book does cover all four of the building blocks of cryptography, but gives very short shrift to modern symmetric key cryptography and hash functions. The same is true of the text by Hoffstein, Pipher, and Silverman [**5**]. The texts by Coutinho [**3**], Koblitz [**6**], Kraft and Washington [**7**], and Lewand [**8**] all cover the two building blocks that use public key cryptography and do not include descriptions of modern symmetric key cryptography or hash functions.

The texts by Buchmann [**2**] and Smart [**11**] thoroughly cover all four of the building blocks of cryptography, but do not tie them together with TLS. Note that both authors have worked in both mathematics and computer science. The latter does not have homework. The only cryptography textbook, written from a mathematical point of view, that tells the complete story is by Trappe and Washington [**14**]. Note that the first author is an engineer and the second is a mathematician.

Let us finish by describing in more detail the book being reviewed. The authors assume no advanced mathematical knowledge—one homework exercise asks the reader to prove for positive integers a and b that $(x^a)^b = x^{ab}$. Indeed, the mathematics of cryptography can be taught proof-free and group-free; ideas from number theory are simply presented as tools. This is a good idea for undergraduate courses including computing students.

I am not entirely clear on the intended audience of this book. It is extremely readable with good motivation and explanations. Only a few parts, like the section on digital steganography (which includes hiding a message in a JPEG file without altering the appearance of the photo) are unclear. I enjoyed the numerous historical facts presented. The text part of the book would be appropriate for the armchair enthusiast. However, the book includes a lot of homework normally ignored by the armchair enthusiast. There are many outstanding homework problems requiring creative thinking. Some of the best perturb something presented in the text and ask the reader to determine the consequences. Strangely, there are many homework problems asking the reader to come up with a difficult number theoretic proof (like showing that Euler's totient function is multiplicative in the number theoretic sense) and these are completely out

of line with the level of the text. By ignoring the homework questions that require proofs, this book would be appropriate for an undergraduate course in cryptography aimed at non-STEM students.

REFERENCES

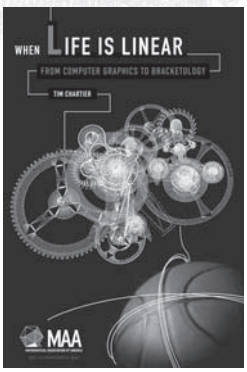
1. G. Bertoni, J. Daemen, M. Peeters, G. van Assche, The KECCAK reference (2011), http://keccak.noekeon.org/Keccak-reference-3.0.pdf.
2. J. Buchmann, *Introduction to Cryptography*. Second edition. Undergraduate Texts in Mathematics, Springer-Verlag, New York, 2004.
3. S. C. Coutinho, *The Mathematics of Ciphers: Number Theory and RSA Cryptography*. AK Peters, Natick, MA, 1999.
4. J. Daemen, V. Rijmen, *The Design of Rijndael: AES—The Advanced Encryption Standard*. Springer-Verlag, Berlin, 2002.
5. J. Hoffstein, J. Pipher, J. H. Silverman, *An Introduction to Mathematical Cryptography*. Undergraduate Texts in Mathematics, Springer-Verlag, New York, 2008.
6. N. Koblitz, *A Course in Number Theory and Cryptography*. Second edition. Graduate Texts in Mathematics, Vol. 97, Springer-Verlag, New York, 1994.
7. J. S. Kraft, L. C. Washington, *An Introduction to Number Theory with Cryptography*. CRC Press, Boca Raton, FL, 2013.
8. R. E. Lewand, *Cryptological Mathematics*. The Mathematical Association of America, Washington, DC, 2000.
9. R. C. Merkle, *Secrecy, Authentication, and Public Key Systems*, Ph.D. dissertation, Stanford University, Stanford, 1979.
10. B. Schneier, *Applied Cryptography: Protocols, Algorithms and Source Code in C*. Second edition. John Wiley & Sons, New York, 1996.
11. N. Smart, *Cryptography: An Introduction*. McGraw-Hill Education, Berkshire, 2003.
12. W. Stallings, *Cryptography and Network Security: Principles and Practices*. Fourth edition. Prentice-Hall of India, New Delhi, 2006.
13. N. Stephenson, *Cryptonomicon*. Avon, New York, 2002.
14. W. Trappe, L. Washington, *Introduction to Cryptography with Coding Theory*. Second edition. Pearson Prentice Hall, Upper Saddle River, NJ, 2006.
15. X. Wang, H. Yu, How to break MD5 and other hash functions, in *Advances in Cryptology*, Lecture Notes in Computer Science, Vol. 3494 (2005) 19–35, http://merlot.usc.edu/csac-f06/papers/Wang05a.pdf.

Santa Clara University, Santa Clara, CA 95053
eschaefer@scu.edu

New from the MAA

When Life is Linear: From Computer Graphics to Bracketology

By Tim Chartier



Catalog Code: NML-45
ISBN: 978-0-88385-649-9
140 pp., Paperbound, 2015
List Price: $50.00
Member Price: $40.00
Series: Anneli Lax NML

Tim Chartier has written the perfect supplement to a linear algebra course. Every major topic is driven by applications, such as computer graphics, cryptography, webpage ranking, sports ranking and data mining. Anyone reading this book will have a clear understanding of the power and scope of linear algebra.

—Arthur Benjamin, Harvey Mudd College

From simulating complex phenomenon on supercomputers to storing the coordinates needed in modern 3D printing, data is a huge and growing part of our world. A major tool to manipulate and study this data is linear algebra. This book introduces concepts of matrix algebra with an emphasis on application, particularly in the fields of computer graphics and data mining. Readers will learn to make an image transparent, compress an image and rotate a 3D wireframe model. In data mining, readers will use linear algebra to read zip codes on envelopes and encrypt sensitive information. The books details methods behind web search, utilized by such companies as Google, and algorithms for sports ranking which have been applied to creating brackets for March Madness and predict outcomes in FIFA World Cup soccer. The book can serve as its own resource or to supplement a course on linear algebra.

To order, visit maa-store.hostedbywebstore.com or call 800-331-1622.

MAA100
MATHEMATICAL ASSOCIATION OF AMERICA
CELEBRATING A CENTURY OF ADVANCING MATHEMATICS