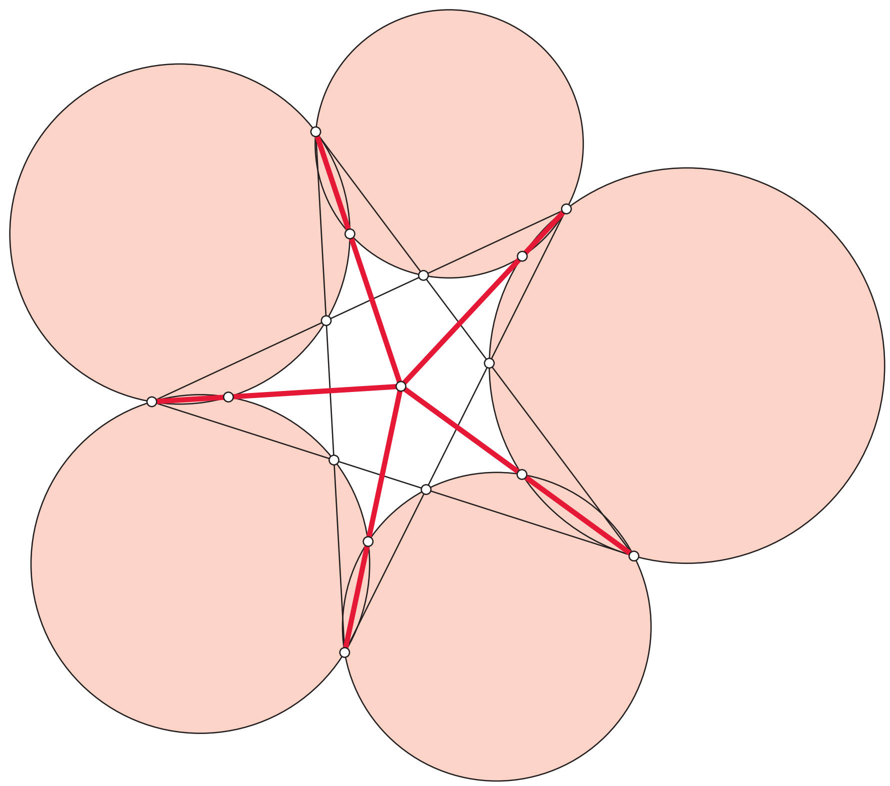
Vol. 87, No. 1, February 2014

MAA

MATHEMATICS MAGAZINE



Five lines meet (page 44)

- Maria Gaetana Agnesi on graphing functions
- Maclaurin's inequality and Bernoulli's inequality
- Singular, reducible, and degenerate quadratic forms

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

EDITORIAL POLICY

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

A more detailed statement of author guidelines appears in this *Magazine*, Vol. 83, at pages 73–74, and is available at the *Magazine*'s website www.maa.org/pubs/mathmag.html. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, or published by another journal or publisher.

Please submit new manuscripts by email directly to the editor at mathmag@maa.org. A brief message containing contact information and with an attached PDF file is preferred. Word-processor and DVI files can also be considered. Alternatively, manuscripts may be mailed to Mathematics Magazine, 132 Bodine Rd., Berwyn, PA 19312-1027. If possible, please include an email address for further correspondence.

Cover image: The image is from the article by Fisher, Hoehn, and Schröder, p. 44. Begin with the star; then each circle is determined by three points of intersection, and each red line is determined by two circles. The five red lines meet at a point.

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

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

Advertising correspondence should be addressed to

MAA Advertising
1529 Eighteenth St. NW
Washington DC 20036

Phone: (877) 622-2373
E-mail: tmarmor@maa.org

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

Change of address, missing issue inquiries, and other subscription correspondence:

MAA Service Center, *maahq@maa.org*

All at the address:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036

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

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

Printed in the United States of America

Vol. 87, No. 1, February 2014

MATHEMATICS MAGAZINE

LETTER FROM THE EDITOR

Antonella Cupillari's article, starting on the facing page, gives us a window into the teaching and learning of mathematics in the 18th century, when Maria Agnesi published her influential textbook. The article treats Agnesi's methods of graphing functions: "quite modern," but with traces of the old ways.

Maclaurin's inequality improves on the AM/GM inequality by inserting other "averages" between the arithmetic mean and the geometric mean. The article by Ben-Ari and Conrad show how Bernoulli's inequality, $1 + x/n \geq \sqrt[n]{(1+x)}$, can be refined in the same way, and how these inequalities are related to each other.

Quadratic forms are the subject of the article by Kronenthal and Lazebnik. The article is a definitive treatment of what it means for a form to be singular. The thoroughness and precision of the article make it useful for anyone who wants to learn more about quadratic forms in an analytic context.

As an expository journal, we still enjoy publishing new research results when they are likely to appeal to a large audience. Three excellent examples are in our Notes section. Reading a research paper motivates learning: The appeal of the result creates a strong incentive to master the arguments in the proof.

I found this to be true of the remarkable five-circle theorem of Fisher, Hoehn, and Schröder, illustrated on our cover. The theorem was discovered using a computer graphics program. The proof uses tools of geometry that arise often in Olympiad-style problems: Ceva, Menelaus, and the Power of a Point. I confess that I did not understand these tools before studying this proof. I understand them much better now; such is the motivating power of a good research result.

Lara Pudwell and Rachel Rockey tell us about de Bruijn sequences and de Bruijn arrays. Their main result is a proof of the existence of de Bruijn L-Arrays, which are a natural generalization of both. The authors point out that there is room for further generalizations, and they challenge us to find new constructions.

Connie Xu addresses a topic that concerned the classical Greeks. What properties distinguish a parabola form other curves? She provides a new characterization. Can you rewrite this theorem and proof using the style and methods of Archimedes? The author's brother, Conway Xu, was the author of a note in our February, 2010 issue. Both were high-school students at the time of publication. Has there been another brother-sister pair of authors in this MAGAZINE? In other MAA publications?

Also in the Notes Section, Harvey Diamond shows how to introduce the elementary transcendendental functions by starting with the differential equations that they solve. Tim Jones proves the irrationality of numbers like e^8, while Poo-Sung Park's note finds a use for the number 8^e.

Thanks to the Putnam Committee for providing the 2013 problems and solutions, which appear on page 71.

As noted on page 24, we are now using the Editorial Manager system for new submissions to the MAGAZINE. To submit an article (or any item), start at `http://www.editorialmanager.com/mathmag`.

Walter Stromquist, Editor

ARTICLES

Maria Gaetana Agnesi's Other Curves (More Than Just the Witch)

ANTONELLA CUPILLARI
Penn State Erie–The Behrend College
Erie, PA 16563
axc5@psu.edu

One of the first calculus books

The textbook *Instituzioni Analitiche ad uso della Gioventù Italiana* (Analytic Institutions for the use of the Italian Youth), by Maria Gaetana Agnesi, was published in Milan, Italy, in 1748. In more than a decade of preparation, Agnesi had studied mathematics intensely, reading authors from Euclid to L'Hôpital under the direction of some of the best Italian mathematicians of the time, hired and befriended by her wealthy father, and had corresponded with other European mathematicians. The amount of correspondence and number of contacts between Maria Agnesi and European mathematicians was quite impressive, especially because she never traveled, aside from visits to her family's country estate during the summer months. After her death, the bulk of her papers, consisting of about twenty-five volumes, were donated to the public library in Milan, where they are still preserved.

Figure 1 Maria Gaetana Agnesi (1718–1799)

Math. Mag. **87** (2014) 3–13. doi:10.4169/math.mag.87.1.3. © Mathematical Association of America

INSTITUZIONI
ANALITICHE
AD USO
DELLA GIOVENTU' ITALIANA
DI D.NA MARIA GAETANA
AGNESI
MILANESE
Dell' Accademia delle Scienze di Bologna.
TOMO I.

IN MILANO, MDCCXLVIII.
NELLA REGIA-DUCAL CORTE.
CON LICENZA DE' SUPERIORI.

Figure 2 Cover page of the *Instituzioni* with Agnesi's handwritten note

The *Instituzioni Analitiche* was quite well received in Europe, and the Pope offered her a professorial position at the University of Bologna. She never took the position. The book's publication marked her retirement from mathematics, to a life of charitable work that continued until her death and for which she used all her inheritance, dying in poverty [**2, 3**].

The fact that the *Instituzioni* [**1**] was written in Italian set it apart from the other works of the same time (such as Euler's *Introductio in analysin infinitorum*, also published in 1748). Agnesi intended it as a teaching instrument for a more general audience not fluent in Latin. It is a well-organized collection of the mathematics that was known at that time, covering topics from introductory algebra, geometry, and trigonometry, to what was the most advanced mathematics of Agnesi's time, calculus. The material is organized in four Books (*Libri*); each Book is divided into Chapters (*Capi*); important paragraphs and sections are numbered consecutively within each Book. Agnesi so carefully supervised the typesetting and the artwork that some of the printing presses of the firm Stampatore Richini (Richini Printing Company) were moved into her large family house. She helped the typesetters learn how to work with mathematical symbols and graphs.

Book One, "Dell'Analisi delle Quantità Finite" (Analysis of finite quantities), introduces algebraic concepts, from numbers to equations to the study of curves and tangents using only the "Algebra Cartesiana." Differential calculus is introduced in Book Two, "Del Calcolo Differenziale."

The witch

In this paper, we present material from Book One, Chapter V, "Of the construction of Loci of degree higher than second degree." This chapter occupies pages 351–415, together with figures collected at the end of Book One. One of the many curves described in Chapter V (in section 238, page 380 and Fig. 135) is the one known in the English-speaking world as "the Witch of Agnesi." This curve is usually represented by the modern equation $y = 1/(1 + x^2)$ and mentioned in most modern calculus books, as it is the derivative of the function $f(x) = \arctan x$. The *Instituzioni* includes the following geometric characterization.

PROBLEM III

238. Given the semicircle ADC (Fig. 135) with diameter AC; find outside of it a point M such that, after constructing MB perpendicular to the diameter AC, which cuts the circle in D, will produce $AB, BD :: AC, BM$. And because there are infinitely many points M that will satisfy the problem, find the curve they form.

Let M be one of such points, and after setting $AC = a$, $AB = x$, $BM = y$, by the properties of the circle one will have $BD = \sqrt{ax - xx}$, and because of the condition given in the problem, one will have $AB, BD :: AC, BM$, that is: $x, \sqrt{ax - xx} :: a, y$; and therefore $y = \dfrac{a\sqrt{ax - xx}}{x}$, or $y = \dfrac{a\sqrt{a - x}}{\sqrt{x}}$, which is the equation of the curve to be described, usually called the *Versiera*.

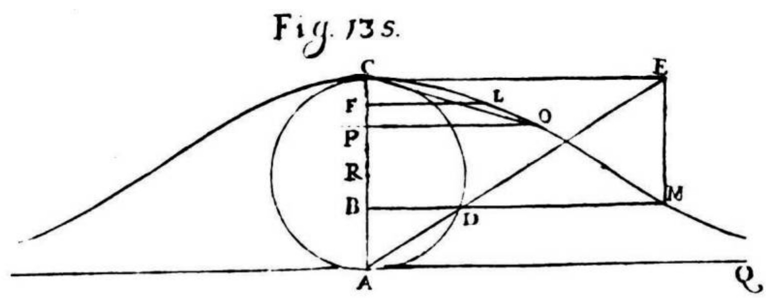


Figure 3 Agnesi's graph of the *Versiera*, the "Witch of Agnesi"

Agnesi used the notation $AB, BD :: AC, BM$ for the proportion that we will write (outside quotes) as $AB : BD = AC : BM$, and she follows common practice in writing the square of x as xx. Agnesi used the variable x for the vertical coordinate and y for the horizontal one. Rewriting her equation $y = a\sqrt{a - x}/\sqrt{x}$ as $x = a^3/(a^2 + y^2)$, setting $a = 1$, and exchanging the roles of the two variables, transforms her equation into the more familiar $y = 1/\sqrt{1 + x^2}$.

The catchy English name for this curve is due to the good reverend John Colson (1680–1760), who was so impressed with Agnesi's work that he learned Italian just to translate the *Instituzioni* into English. Instead of translating the words "*la Versiera*" (which refers to a change of "*verse*," or concavity) he translated the words "*l'aversiera,*" which meant "an evil female spirit, the wife of the devil"—hence, "the Witch."

INSTITUZIONI
ANALITICHE
LIBRO PRIMO
Dell' Analiſi delle Quantità finite.

L'Analiſi delle quantità finite, che comunemente chiamaſi Algebra Carteſiana, è un metodo, con cui trattando quantità finite ſi ſciolgono i Problemi; cioè da certe quantità, e condizioni date e cognite, ſi viene in cognizione d'altre incognite, e che ſi cercano, per mezzo di alcune operazioni, e metodi, che parte a parte mi propongo di ſpiegare ne' ſeguenti Capi.

A CAPO

Figure 4 First page of Book One. Of the Analysis of Finite Quantities

Colson's work, *Analytical Institutions, in four books; originally written by Donna Maria Gaetana Agnesi; Professor of the Mathematicks and Philosophy in the University of Bologna*, was edited by Rev. John Hellins and appeared in 1801, after Colson's death. Luckily, Rev. Hellins kept the curve's name as translated by Rev. Colson. The curve did not have a special place in the list of curves presented in the chapter, and it was neither the most interesting nor the most difficult to graph, but it is the one that made Agnesi famous and earned her a spot in many contemporary calculus books [**2**].

More curves

Agnesi's working is quite modern, in the sense that it is not hard to follow for someone who is used to modern textbooks (and knows Italian). But it still has traces of old-style ways of thinking. For example, Agnesi was still very conscious of the possible physical/geometrical meaning behind the variables she used. She referred to sums and differences of variables as "analytic quantities of one dimension" and used the word "rectangle" when considering the product of two variables. So equations would be dimensionally consistent as much as possible. For example, Agnesi would prefer to write $y^3 = aax$ or $y^3 = axx$ rather than $y^3 = x$ or $y^3 = xx$, even if the constant a had no other significance.

More than twenty curves are presented in Chapter V. Some curves are introduced algebraically, while others are described geometrically at first. No reason is given for the choice of some rather complicated equations. The list includes well-known curves, such as the cissoid of Diocles, the conchoids of Nicomedes, the equilateral hyperbolas, and the parabolas of Apollonius. The name "first parabolas" was used for the curves described by the equations $y^n = a^{n-1}x$, with $n \geq 2$. Later on, Agnesi used these curves to construct the graphs of curves represented by higher-degree equations, such as the "second cubic" $y^3 = axx$.

Several sections in Chapter V are devoted to general suggestions on how to graph curves and how to determine their geometric properties without using calculus. Agnesi remarked time and again that the use of coordinate axes perpendicular to each other was not crucial for the work presented, and that it was possible to use several different techniques to produce a graph. The first technique she used is a mix of point-plotting and discussion of asymptotes and concavities. Concavities are determined geometrically, using the coordinates of two points on the graph of the function and comparing the position of the line that joins them to the section of graph between them.

Consider Example III (page 361), in which Agnesi showed how to graph the curve represented by the equation $a^3 - zyy = 0$. The calculations assume that $a > 0$, but this assumption is never spelled out. Agnesi's graph, her Fig. 129, is our FIGURE 5.

The first step consists of rewriting the equation as $y = \pm\sqrt{a^3/z}$. The horizontal axis is the z-axis, with the point A corresponding to $z = 0$ on the axis. Agnesi wrote, "set $z = 0$, the equation is $y = \pm\sqrt{a^3/0}$, that is $y = \pm\infty$, and therefore BD, infinite on both sides of A will be an asymptote of the curve." Then she tried to find a point of intersection between the graph and the z-axis, by setting $y = 0$:

> ...and the equation will be $\pm\sqrt{\dfrac{a^3}{z}} = 0$, that is $0 = \dfrac{a^3}{z}$, or the same as saying $z = \dfrac{a^3}{0}$, and thus $z = \infty$. Therefore when $y = 0$ it will be $z = \infty$, and thus AC will be another asymptote.

She then found the coordinates of several points on the curve, such as $F(a, a)$, $G(a, -a)$, $I(2a, \sqrt{aa/2})$, and $K(2a, -\sqrt{aa/2})$.

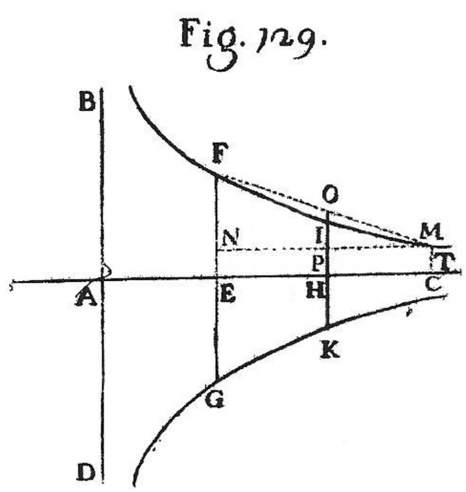


Figure 5 Agnesi's graph of the curve $a^3 - zyy = 0$

At this point, Agnesi observed that larger values of z yield smaller values of y, and the curve gets closer to the asymptote, without touching it, except possibly at an infinite distance from the point A. This discussion was followed by an explanation of why negative values of z cannot be used and a study of the concavity/convexity of the curve [**1**].

> To study whether the curve is concave, or convex with respect to its axis AC; I take $AC = 3a$, then it will be $CM = \sqrt{\frac{aa}{3}}$, and drawing FM, which meets in OH the line extended from HI, if needed, and drawing MN parallel to AC, it will be $NF = a - \sqrt{\frac{aa}{3}}$, and $PI = \sqrt{\frac{aa}{2}} - \sqrt{\frac{aa}{3}}$. Then use the similarity $MN, NF ::$ MP, PO, that is $2a, a = \sqrt{\frac{aa}{3}} :: a, PO$, and it will be $PO = \frac{a - \sqrt{\frac{aa}{3}}}{2}$, and therefore if PO is larger than PI, the curve is convex with respect to the axis AC, which will be proved in the following way.

The construction of the proof of the statement "PO is larger than PI" is presented in a slightly peculiar, yet correct, way. Agnesi started with the inequality to be proved, namely

$$\frac{a - \sqrt{\frac{aa}{3}}}{2} > \sqrt{\frac{aa}{2} - \sqrt{\frac{aa}{3}}},$$

and then proceeded to simplify it using a chain of equivalent inequalities. After several steps, she obtained the mathematical statement $9aa/3 > aa$ and concluded

> ...since it is true that $\frac{9aa}{3} > aa$, it is also true that $\frac{a - \sqrt{\frac{aa}{3}}}{2} > \sqrt{\frac{aa}{2}} - \sqrt{\frac{aa}{3}}$, that is, that PO is larger than PI, and as a consequence, that the curve is convex with respect to the axis AT.

Another detail to notice is that Agnesi had started the work on this curve using the interval $[a, 3a]$ on the z-axis and did not address how her conclusions could be extended to the whole positive z-axis. This is consistent with the techniques used to construct proofs at this time, when a special case was discussed, using methods that could be generalized.

Mathematical details get considerably more complicated in Example IV (page 363), where Agnesi introduced the curve with equation

$$y = \pm\sqrt{\frac{4ax + aa - 2xx \pm a\sqrt{aa + 8ax}}{2}}$$

whose graph is in FIGURE 6. At first, she determined the coordinates of the y-intercepts, namely $A(0, 0)$, $E(0, a)$, and $E(0, -a)$, sometimes using the same letter to identify two different, symmetrically-placed points.

Then she found the x-intercepts, namely $A(0, 0)$, $F(a, 0)$, and $B(3a, 0)$. For reasons that are not disclosed, she used $x = a/2$ to find more points on the graph. This calculation yields the points

$$G\left(\frac{a}{2}, \sqrt{\frac{5aa + 2a\sqrt{5aa}}{4}}\right) \quad \text{and} \quad I\left(\frac{a}{2}, \sqrt{\frac{5aa - 2a\sqrt{5aa}}{4}}\right)$$

and their mirror images with respect to the x-axis, called g and i. Agnesi did not simplify the equation of the curve and the radical expressions just obtained, in spite of having introduced the algebra of radical expressions earlier in Book One. She then determined the collection of values of x that would yield real values for the coordinate y. Using lengthy calculations, she was able to conclude that the graph of the curve exists only when $-a/8 \le x \le 3a$. The study of this curve concludes as follows.

> Letting $x = \frac{a}{8}$ [the correct value is $x = -\frac{a}{8}$] will give $y = \pm\frac{\sqrt{15aa}}{8}$, and thus, making KM positive and KN negative and equal to $\frac{\sqrt{15aa}}{8}$, the points M and N will be on the curve. I take $x = \frac{a}{16}$ [the correct value is $x = -\frac{a}{16}$], and thus $Y = \pm\frac{\sqrt{95aa \pm 128a\sqrt{\frac{aa}{2}}}}{16}$, that is four real values: two positive, relatively equal to the two negative values. And because the fourth proportional between $\frac{a}{8}$, $\frac{\sqrt{15aa}}{8}$, and $\frac{a}{16}$, that is $\frac{\sqrt{15aa}}{16}$ is smaller than $\sqrt{\frac{\sqrt{95aa + 128a\sqrt{\frac{aa}{2}}}}{16}}$, but it is larger than $\sqrt{\frac{\sqrt{95aa - 128a\sqrt{\frac{aa}{2}}}}{16}}$; the curve will have two branches above AK, one concave and the other convex, and it will also have two similar ones below, equal to the one above, as in Fig. 130.

The branches mentioned by Agnesi are the arc AM and what should be the arc MD. Indeed, the curve shown in Fig. 130 (our FIGURE 6) has some peculiar features, including the vertical lines AD (basically the y-axis) and MN, and an extra pair of loops, which are not easy to investigate using only algebra and Agnesi's complicated equation. But the reason behind her choice for the coefficients of the radicand defining the

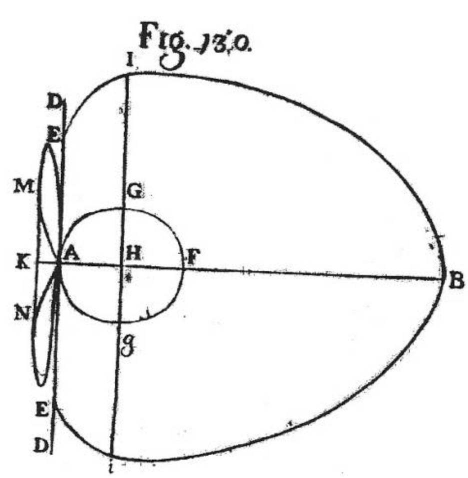


Figure 6 Should it be a limaçon?

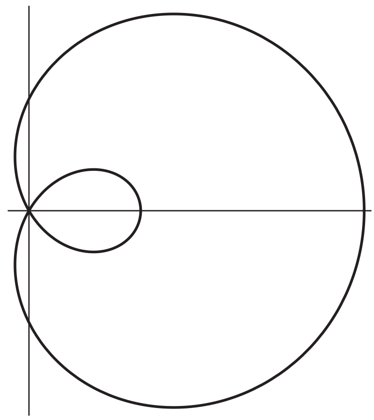


Figure 7 It is a limaçon.

curve becomes evident if we introduce polar coordinates and rewrite the equation in modern notation as

$$4r^2\left[(r-2a\cos\theta)^2-a^2\right]=0$$

or

$$4r^2\left[(r-2a\cos\theta)-a\right]\left[(r-2a\cos\theta)+a\right]=0.$$

The equation $r=\pm a+2a\cos\theta$ is a limaçon, shown in FIGURE 7.

This suggests that the graph in Fig. 130 is incorrect, but Reverend Colson, who translated the work into English and made some changes to the presentation of the material (e.g., he moved the graphs of the curves from the end of the chapter to their respective sections) did not comment on this issue.

The next example (Example V, page 371) involves the curve represented by the equation

$$y=\pm\sqrt{\frac{bbx-x^3+2axx-aax}{x-2a}},$$

with $0<b<a$, whose graph is given in FIGURE 8.

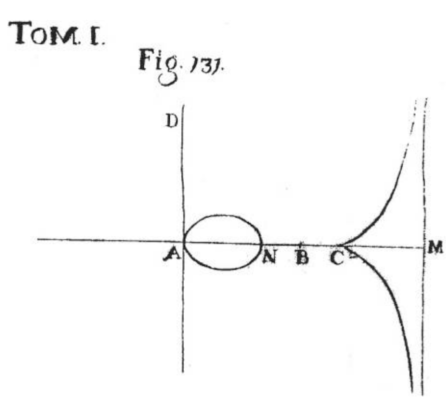


Figure 8 Agnesi's graph of a two-component curve

Again, Agnesi offered no clues as to the reason for including this curve and the choice of the rational expression under the square root. One interesting feature is offered by the fact that the graph of the equation consists of two separate parts, one finite and the other with an asymptotic behavior. In this case, there is no discussion of concavities, and the work included just determines the coordinates of a few points and the values of x that generate real values of y, which is the set $[0, a-b) \cup [a+b, 2a)$ (in modern notation). The equation of this curve can also be rewritten in polar coordinates as

$$r^2 \cos\theta - 2ar + (a^2 - b^2)\cos\theta = 0.$$

(We omit the details.) In this case, we obtain two equations, one for each of the two components of the curve.

Another way to graph

Agnesi used another method for graphing, which involved the use of easier-to-graph curves in order to construct other graphs. If the equation of a curve was quite complex, she would rewrite it using easier algebraic terms, and then used geometry to put together the pieces and construct the graph of the original equation. This process is better explained using one of the original examples, the curve from paragraph 245 (her Fig. 141, our FIGURE 9), in which the "first parabola" $aax = y^3$ is graphed. Agnesi described some properties of this curve before performing other calculations.

> It is clear that it [the first parabola] will have two branches, one positive and one negative, because when using a positive x value, y will also be positive, that is $y = \sqrt[3]{aax}$, and this will be the positive branch. But using the negative x, the y value will also be negative, that is $y = \sqrt[3]{-aax}$, (which is not an imaginary quantity) and this will be the negative branch. It is clear that the two branches go to infinity and are concave with respect to the axis AB.

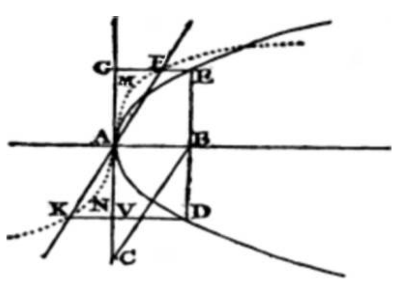


Figure 9 Graph of the curve $aax = y^3$ (dotted line); Fig. 141

Then Agnesi introduced a third variable z, defined by the lower-degree equation $yy = az$. The equation of the cubic can then be rewritten as $aax = azy$, or $ax = zy$. This equation is equivalent to the proportion $a : z = y : x$. Her technique consisted of constructing a geometric setting to give a meaning to this proportion, and then finding the points whose coordinates satisfy the proportion, which will be the points on the original curve. All these steps are illustrated in one graph in the *Instituzioni*. We have included two intermediate graphs to show more details of the first two steps, in FIGURE 10. Agnesi graphed the parabola $yy = az$ with horizontal axis AB and vertex A (used as the origin of the axes), and called it DAE. Then she carefully chose $AB = z$, $BE = y$, $BD = -y$, and AC to be a segment of length a.

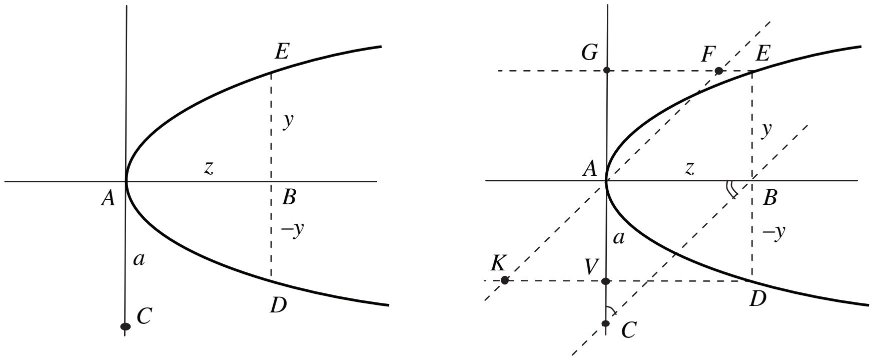


Figure 10 Steps toward Agnesi's Fig. 141

The projections of the segments BE and BD onto the y-axis determined the segments AG and AV, respectively. At this point, Agnesi constructed the line parallel to the segment CB through A, and named K and F its points of intersection with the horizontal lines through E and D, as shown in the following graph.

Using the similar triangles $\triangle ABC$ and $\triangle AGF$, it is possible to establish the relation $CA : AB = AG : GF$, which becomes $a : z = y : x$ after setting $x = GF$. So, the point F is on the graph of the curve with equation

$$a : z = y : x$$

$$a : \frac{yy}{a} = y : x$$

$$y^3 = aax.$$

Since a negative x value yields a negative y value, there is another point on the graph of the curve, placed symmetrically with respect to the origin from F, the point K. These are the only two points explicitly found.

Agnesi used the graph just obtained as an auxiliary one in the next paragraph of her book to produce the graph of the curve $y^4 = a^3x$ through the same geometric method. Using the lower-degree equation $y^3 = aax$ (and its graph, which is the curve QAD in FIGURE 11) allowed her to rewrite the equation $y^3 = a^3x$ as $zy = ax$ or $a : z = y : z$. So the problem became that of finding a geometric construction to give a meaning to this proportion.

She set $AC = GD = z$, $AK = QR = -z$, $CD = AG = y$, $KQ = AR = -y$, and let AB be a segment of length a. She then constructed the segments BC and BK.

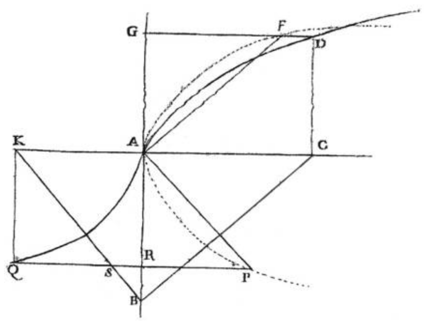


Figure 11 Agnesi's graph of the curve $y^4 = a^3x$ (dotted line) Fig. 142

Using the origin A, she drew the lines AF parallel to BC, and AP parallel to KB, and observed that the triangles $\triangle ABC$ and $\triangle GAF$ are similar. Therefore, $AB : AC = AG : GF$ or $a : z = y : x$, where $x = GF$. This implies that $a : y^3/(aa) = y : x$ and $y^4 = a^3x$. Thus the point F is on the graph of the curve, and so is the point P, its reflection with respect to the axis. Agnesi's graph does not emphasize this symmetry, and it also does not clearly show that the segments AK and AC have the same length; nonetheless, it shows the main characteristics of the curve, a sufficiently good result given the technical difficulty that she faced with the printing tools of the time.

Conclusion

We have to wonder how well Agnesi's name would be remembered if she had decided that the Versiera was not interesting enough to be included in Chapter V, or if Reverend Colson had had more time to learn Italian. Such is serendipity, and the irony of the situation in which a very knowledgeable, complex, and religious woman was made famous by a simple name: "the Witch."

REFERENCES

1. Maria Gaetana Agnesi, *Instituzioni Analitiche ad Uso della Gioventù Italiana,* Milano, 1748.
2. Antonella Cupillari, *A Biography of Maria Gaetana Agnesi, an Eighteenth-Century Woman Mathematician: With Translations of Some of Her Work from Italian into English,* Edwin Mellen Press, Lewiston, NY, 2007.
3. Massimo Mazzotti, *The World of Maria Gaetana Agnesi, Mathematician of God,* Johns Hopkins University Press, Baltimore, 2007.

Summary Maria Gaetana Agnesi's name is commonly associated with the curve known as "the Witch of Agnesi." However, the versiera (or versoria), as Agnesi called it, is only one of many curves she introduced in her mathematical compendium, the *Instituzioni* Analitiche (1748). Some of the other curves are much more interesting and complex than the versiera, and they are grouped in the lengthy section (pp. 351–415) titled "On the construction of Loci of degree higher than second degree." This article showcases Agnesi's presentation of some of them, made without using calculus. Instead, her tools of choice were geometry, algebra, and the method of using easier and already-known curves to build more challenging ones.

ANTONELLA CUPILLARI is the author of the books *The Nuts and Bolts of Proofs: An Introduction to Proofs* and *A Biography of Maria Gaetana Agnesi, an Eighteenth-Century Woman Mathematician: With Translations.* Her research focuses on history of mathematics and undergraduate education and she has published articles in both fields.

Maclaurin's Inequality and a Generalized Bernoulli Inequality

IDDO BEN-ARI
University of Connecticut
Storrs, CT 06269-3009
iddo.ben-ari@uconn.edu

KEITH CONRAD
University of Connecticut
Storrs, CT 06269-3009
kconrad@math.uconn.edu

One of the most famous inequalities in mathematics is the *arithmetic-geometric mean inequality*: For every positive integer n and $x_1, \dots, x_n > 0$,

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 x_2 \cdots x_n}, \tag{1}$$

and the inequality is strict unless the x_i's are all equal. Even students who are not active users of inequalities know (or should know!) this inequality. Are you aware that there is an extension of (1) that includes terms between the average on the left and the nth root on the right? It was first stated by Maclaurin in 1729 [**5**, pp. 80–81], but remains relatively unknown, except to aficionados of inequalities.

To place intermediate terms in (1), we will use the *elementary symmetric polynomials* in $x_1, \dots, x_n$, which are

$$e_k(x_1, \dots, x_n) = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} x_{i_1} x_{i_2} \cdots x_{i_k} = \sum_{\substack{I \subset \{1,\dots,n\} \\ \#I = k}} \prod_{i \in I} x_i$$

for $1 \leq k \leq n$. For instance, when $n = 3$,

$$e_1(x, y, z) = x + y + z, \quad e_2(x, y, z) = xy + xz + yz, \quad e_3(x, y, z) = xyz.$$

Elementary symmetric polynomials naturally arise as coefficients of the polynomial

$$(T - x_1)(T - x_2) \cdots (T - x_n) = T^n - e_1 T^{n-1} + e_2 T^{n-2} - \cdots + (-1)^n e_n.$$

In particular, $e_1 = x_1 + \cdots + x_n$ and $e_n = x_1 \cdots x_n$.

Each $e_k(x_1, \dots, x_n)$ is a sum of $\binom{n}{k}$ terms, and its average

$$E_k(x_1, \dots, x_n) = \frac{e_k(x_1, \dots, x_n)}{e_k(1, \dots, 1)} = \frac{e_k(x_1, \dots, x_n)}{\binom{n}{k}}$$

is called the kth *elementary symmetric mean* of $x_1, \dots, x_n$. When $n = 3$,

$$E_1(x, y, z) = \frac{x + y + z}{3}, \quad E_2(x, y, z) = \frac{xy + xz + yz}{3}, \quad E_3(x, y, z) = xyz.$$

Math. Mag. **87** (2014) 14–24. doi:10.4169/math.mag.87.1.14. © Mathematical Association of America

Now we can state Maclaurin's inequality: For $n \geq 2$ and positive $x_1, \ldots, x_n$,

$$\boxed{\frac{x_1 + \cdots + x_n}{n} \geq \sqrt{\frac{\sum\limits_{1 \leq i < j \leq n} x_i x_j}{\binom{n}{2}}} \geq \sqrt[3]{\frac{\sum\limits_{1 \leq i < j < k \leq n} x_i x_j x_k}{\binom{n}{3}}} \geq \cdots \geq \sqrt[n]{x_1 x_2 \cdots x_n},}$$

or equivalently,

$$E_1(x_1, \ldots, x_n) \geq \sqrt{E_2(x_1, \ldots, x_n)} \geq \sqrt[3]{E_3(x_1, \ldots, x_n)} \geq \cdots \geq \sqrt[n]{E_n(x_1, \ldots, x_n)}. \tag{2}$$

Moreover, the inequalities are all strict unless the x_i's are all equal. For example, when $n = 3$, Maclaurin's inequality says for positive x, y, and z that

$$\frac{x + y + z}{3} \geq \sqrt{\frac{xy + xz + yz}{3}} \geq \sqrt[3]{xyz},$$

and both inequalities are strict unless $x = y = z$.

The arithmetic-geometric mean inequality follows from Maclaurin's inequality (look at the first and last terms), and these inequalities are linked historically: The paper in which Maclaurin stated his inequality is also where the arithmetic-geometric mean inequality for n terms, not just two terms, first appeared [**5**, pp. 78–79].

The standard proof of Maclaurin's inequality [**1**, pp. 10–11; **3**, p. 52; **9**, p. 97, Theorem 4; **12**, Ch. 12] is based on Newton's inequality, which says

$$E_k(x_1, \ldots, x_n)^2 \geq E_{k-1}(x_1, \ldots, x_n) E_{k+1}(x_1, \ldots, x_n)$$

for $x_1, \ldots, x_n > 0$ and $1 \leq k \leq n - 1$, where $E_0(x_1, \ldots, x_n) = 1$. Recently, Maligranda [**6**, Theorem 3], [**7**] showed that the arithmetic-geometric mean inequality is equivalent to *Bernoulli's inequality*: For positive integers n and real $t > -1$,

$$(1 + t)^n \geq 1 + nt, \tag{3}$$

with the inequality being strict for $n > 1$ unless $t = 0$. Maligranda's work raises a question: Can Bernoulli's inequality generalize to complete the diagram below?

Arithmetic-Geometric Mean Inequality $\iff$ Bernoulli's Inequality

Maclaurin's Inequality $\iff$???

We will present such an extension of Bernoulli's inequality, use it in a new proof of Maclaurin's inequality, and show that the two inequalities are equivalent. After this, we describe two uses of Maclaurin's inequality, and an extension of Maclaurin's and Bernoulli's inequalities to graphs.

A generalized Bernoulli inequality

To generalize Bernoulli's inequality, we rewrite it in a more convenient form. When $-1 < t \leq -1/n$, the inequality (3) is obvious, since the left side is positive and the right side is negative or zero. When $t > -1/n$, set $x = nt$ and rewrite (3) as

$$1 + \frac{1}{n}x \geq \sqrt[n]{1+x} \tag{4}$$

for $x > -1$, with equality (for $n > 1$) if and only if $x = 0$. It is (4) that we will henceforth consider to be Bernoulli's inequality. Our extension of (4), which we call the *generalized Bernoulli inequality*, is the following: For integers $n \geq 2$ and real $x > -1$,

$$\boxed{1 + \frac{1}{n}x \geq \sqrt{1 + \frac{2}{n}x} \geq \sqrt[3]{1 + \frac{3}{n}x} \geq \cdots \geq \sqrt[n]{1 + \frac{n}{n}x},} \tag{5}$$

with the inequalities all strict unless $x = 0$. This is related to (4) in a similar way to the relation of Maclaurin's inequality to the arithmetic-geometric mean inequality.

To prove (5), all terms are equal when $x = 0$. For $x \neq 0$, we want to show that

$$\sqrt[k]{1 + \frac{k}{n}x} \;>\; \sqrt[k+1]{1 + \frac{k+1}{n}x}$$

for $n \geq 2$ and $1 \leq k \leq n-1$. Equivalently, we want to show that

$$\frac{1}{k}\log\left(1 + \frac{k}{n}x\right) > \frac{1}{k+1}\log\left(1 + \frac{k+1}{n}x\right),$$

where log is the natural logarithm. This will be derived from $\log t$ being *strictly concave*:

$$\log(\lambda u + (1-\lambda)v) > \lambda \log u + (1-\lambda)\log v \tag{6}$$

for distinct positive numbers u and v, and $0 < \lambda < 1$ (FIGURE 1). Since $1 + \frac{k}{n}x$ lies strictly between $u = 1$ and $v = 1 + \frac{k+1}{n}x$, write $1 + \frac{k}{n}x$ as a convex combination of u and v:

$$1 + \frac{k}{n}x = \lambda u + (1-\lambda)v,$$

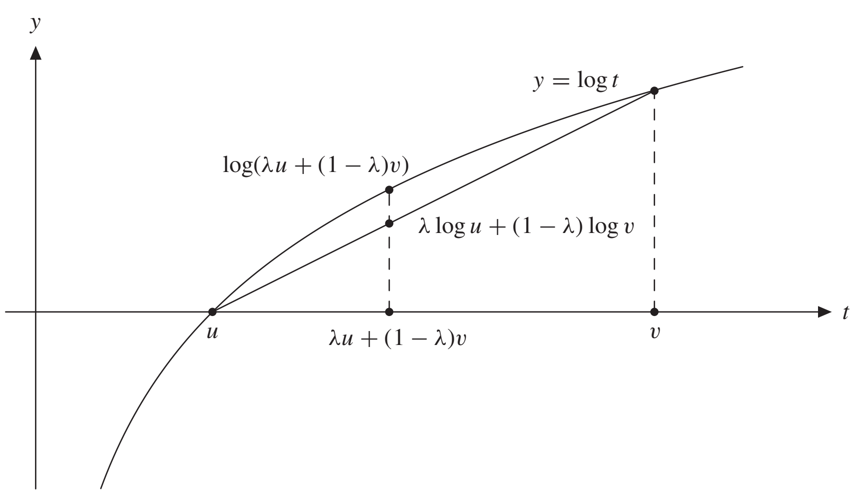


Figure 1 Strict concavity of log *t*

for $\lambda = \frac{1}{k+1}$. Then

$$\begin{aligned}\frac{1}{k}\log\left(1+\frac{k}{n}x\right) &= \frac{1}{k}\log\left(\lambda u + (1-\lambda)v\right)\\ &> \frac{1}{k}\left(\lambda\log u + (1-\lambda)\log v\right) \qquad \text{by (6)}\\ &= \frac{1}{k+1}\log\left(1+\frac{k+1}{n}x\right).\end{aligned}$$

That completes the proof of the generalized Bernoulli inequality (5). Now we want to derive Maclaurin's inequality from (5) by induction on n.

In the base case $n = 2$, Maclaurin's inequality is the arithmetic-geometric mean inequality for two terms, which has many proofs. Here is a proof from the generalized Bernoulli inequality when $n = 2$, which says that $1 + \frac{1}{2}x \geq \sqrt{1+x}$ for $x > -1$, with strict inequality unless $x = 0$. For positive x_1 and x_2,

$$\begin{aligned}\frac{x_1+x_2}{2} &= x_2\left(\frac{x_1/x_2}{2}+\frac{1}{2}\right)\\ &= x_2\left(1+\frac{1}{2}\left(\frac{x_1}{x_2}-1\right)\right)\\ &\overset{\text{gen. Bern.}}{\geq} x_2\sqrt{1+\left(\frac{x_1}{x_2}-1\right)}\\ &= \sqrt{x_1x_2},\end{aligned}$$

and this inequality is strict unless $x_1/x_2 - 1 = 0$, that is, $x_1 = x_2$.

Assume that Maclaurin's inequality holds for $n-1$ variables, where $n \geq 3$. We want to show that it holds for n variables, in the form (2). Since each $E_k(x_1, \ldots, x_n)$ is symmetric in the x_i's, without loss of generality $x_n = \max_i x_i$. To simplify notation, write

$$\begin{aligned}E_k &= E_k(x_1,\ldots,x_n) \qquad \text{for } 1 \leq k \leq n,\\ \varepsilon_k &= E_k(x_1,\ldots,x_{n-1}) \qquad \text{for } 1 \leq k \leq n-1,\end{aligned}$$

and set $\varepsilon_0 = 1$ and $\varepsilon_n = 0$.

We will use a recursive formula for the E_k's:

$$E_k(x_1,\ldots,x_n) = \left(1-\frac{k}{n}\right)E_k(x_1,\ldots,x_{n-1}) + \frac{k}{n}E_{k-1}(x_1,\ldots,x_{n-1})x_n \tag{7}$$

for $1 \leq k \leq n$, or in more abbreviated form,

$$E_k = \left(1-\frac{k}{n}\right)\varepsilon_k + \frac{k}{n}\varepsilon_{k-1}x_n. \tag{8}$$

This recursion follows from a recursive formula for elementary symmetric polynomials:

$$e_k(x_1,\ldots,x_n) = e_k(x_1,\ldots,x_{n-1}) + e_{k-1}(x_1,\ldots,x_{n-1})x_n \tag{9}$$

for $1 \le k \le n$, where we set $e_0(x_1, \ldots, x_{n-1}) = 1$ and $e_n(x_1, \ldots, x_{n-1}) = 0$:

$$E_k(x_1, \ldots, x_n) = \frac{e_k(x_1, \ldots, x_{n-1}) + e_{k-1}(x_1, \ldots, x_{n-1})x_n}{\binom{n}{k}} \qquad \text{by (9)}$$
$$= \frac{\binom{n-1}{k} E_k(x_1, \ldots, x_{n-1}) + \binom{n-1}{k-1} E_{k-1}(x_1, \ldots, x_{n-1})x_n}{\binom{n}{k}},$$

which is (7) after simplifying the ratios of binomial coefficients.

By Maclaurin's inequality for $n-1$ variables, $\varepsilon_{k-1}^{1/(k-1)} \ge \varepsilon_k^{1/k}$ for $2 \le k \le n-1$. Rewrite this in two ways:

$$\varepsilon_{k-1} \ge \varepsilon_k^{(k-1)/k} \qquad \text{and} \qquad \varepsilon_{k+1} \le \varepsilon_k^{(k+1)/k} \tag{10}$$

for $1 \le k \le n-1$. (The first inequality holds at $k = 1$ by the definition of ε_0 and the second inequality holds at $k = n-1$ by the definition of ε_n.) Using (8) and (10), when $1 \le k \le n-1$,

$$\begin{aligned} E_k &= \left(1 - \frac{k}{n}\right)\varepsilon_k + \frac{k}{n}\varepsilon_{k-1}x_n \\ &\ge \left(1 - \frac{k}{n}\right)\varepsilon_k + \frac{k}{n}\varepsilon_k^{(k-1)/k}x_n \\ &= \varepsilon_k\left(1 + \frac{k}{n}\left(\varepsilon_k^{-1/k}x_n - 1\right)\right) \end{aligned} \tag{11}$$

and

$$\begin{aligned} E_{k+1} &= \left(1 - \frac{k+1}{n}\right)\varepsilon_{k+1} + \frac{k+1}{n}\varepsilon_k x_n \\ &\le \left(1 - \frac{k+1}{n}\right)\varepsilon_k^{(k+1)/k} + \frac{k+1}{n}\varepsilon_k x_n \\ &= \varepsilon_k^{(k+1)/k}\left(1 + \frac{k+1}{n}\left(\varepsilon_k^{-1/k}x_n - 1\right)\right). \end{aligned} \tag{12}$$

Letting c_k denote the (positive) term $\varepsilon_k^{-1/k}x_n$ in (11) and (12), we combine these results to complete the proof:

$$E_k^{1/k} \overset{(11)}{\ge} \varepsilon_k^{1/k}\sqrt[k]{1 + \frac{k}{n}(c_k - 1)} \overset{\text{gen. Bern.}}{\ge} \varepsilon_k^{1/k}\sqrt[k+1]{1 + \frac{k+1}{n}(c_k - 1)} \overset{(12)}{\ge} E_{k+1}^{1/(k+1)}.$$

When does equality occur in (2)? From the way we used the generalized Bernoulli inequality (5) just above, $E_k^{1/k} > E_{k+1}^{1/(k+1)}$ if $c_k - 1 \ne 0$. How can $c_k - 1 = 0$, or equivalently, how can $\varepsilon_k^{1/k} = x_n$, when $1 \le k \le n-1$? Since

$$\varepsilon_k^{1/k} \le \varepsilon_1 = \frac{1}{n-1}(x_1 + \cdots + x_{n-1}),$$

and $x_n = \max_i x_i$, if any x_i is less than x_n then $\varepsilon_1 < x_n$, so $\varepsilon_k^{1/k} < x_n$. Thus all inequalities in (2) are strict unless each x_i is x_n, in which case all terms in (2) equal x_n.

The generalized Bernoulli inequality not only implies Maclaurin's inequality, but also follows from it. For $n \geq 2$ and $x > -1$, set $x_1 = \cdots = x_{n-1} = 1$ and $x_n = 1 + x$ in (7), to get for $1 \leq k \leq n-1$ that

$$E_k(\underbrace{1, \ldots, 1}_{n-1 \text{ terms}}, 1+x) = \left(1 - \frac{k}{n}\right) E_k(\underbrace{1, \ldots, 1}_{n-1 \text{ terms}}) + \frac{k}{n} E_{k-1}(\underbrace{1, \ldots, 1}_{n-1 \text{ terms}})(1+x).$$

Since $k < n$, the right side is $(1 - k/n) + (k/n)(1 + x) = 1 + (k/n)x$. This formula for $E_k(1, \ldots, 1, 1 + x)$ works when $k = n$ too:

$$E_n(1, \ldots, 1, 1+x) = 1 + x = 1 + \frac{n}{n} x.$$

Thus Maclaurin's inequality (2), when $x_1 = \cdots = x_{n-1} = 1$ and $x_n = 1 + x$, is the generalized Bernoulli inequality (5). And if the inequalities in Maclaurin's inequality are all strict unless the x_i's are all equal, then the inequalities in the generalized Bernoulli inequality are all strict unless $x = 0$.

The inequality (6) was used with $\lambda = \frac{1}{k+1}$ to prove the generalized Bernoulli inequality (5), which implies Maclaurin's inequality (2), which has the arithmetic-geometric mean inequality (1) as a special case. Let's complete the cycle by proving (6) with rational $\lambda \in (0, 1)$ from the arithmetic-geometric mean inequality, so Maclaurin's inequality and the generalized Bernoulli inequality are equivalent to (6) with rational λ.

Write any rational $\lambda \in (0, 1)$ as $\frac{k}{n}$ for an integer $n \geq 2$ and $k \in \{1, \ldots, n-1\}$. For $0 < u < v$, let $x_1 = x_2 = \cdots = x_k = u$ and $x_{k+1} = \cdots = x_n = v$. By the arithmetic-geometric mean inequality,

$$\begin{aligned} \lambda u + (1-\lambda)v &= \frac{(x_1 + \cdots + x_k) + (x_{k+1} + \cdots + x_n)}{n} \\ &> \sqrt[n]{(x_1 \cdots x_k)(x_{k+1} \cdots x_n)} \\ &= u^{\lambda} v^{1-\lambda}, \end{aligned}$$

where the inequality is strict since $x_1 \neq x_n$. Taking logarithms yields (6).

First application: convergence of a recursive sequence

The most interesting (to us) application of Maclaurin's inequality is to a recursion in n variables that generalizes Gauss's arithmetic-geometric mean recursion in two variables.

For any positive x and y, define sequences $\{x_j\}$ and $\{y_j\}$ for $j \geq 0$ by $x_0 = x$, $y_0 = y$, and

$$x_j = \frac{x_{j-1} + y_{j-1}}{2}, \quad y_j = \sqrt{x_{j-1} y_{j-1}}$$

for $j \geq 1$. The sequences $\{x_j\}$ and $\{y_j\}$ converge to a common limit, which Gauss denoted $M(x, y)$ and called the *arithmetic-geometric mean* of x and y. A proof of convergence is based on the arithmetic-geometric mean inequality for two numbers [**2**, Section 1].

The construction of $M(x, y)$ generalizes from two numbers to n numbers using elementary symmetric means: For $x_1, \ldots, x_n > 0$, define n sequences $\{x_{1,j}\}, \ldots, \{x_{n,j}\}$ for $j \geq 0$ by

$$x_{k,0} = x_k \text{ and } x_{k,j} = \sqrt[k]{E_k(x_{1,j-1}, \ldots, x_{n,j-1})}$$

for $j \geq 1$.

EXAMPLE 1. Let $x_1 = 1$, $x_2 = 2$, and $x_3 = 3$. TABLE 1 lists the first few iterations to 16 digits after the decimal point. Although $x_{1,0} < x_{2,0} < x_{3,0}$, we have $x_{1,j} > x_{2,j} > x_{3,j}$ for $j \geq 1$ by Maclaurin's inequality.

TABLE 1: An iteration of E_1, $\sqrt{E_2}$, and $\sqrt[3]{E_3}$ on three numbers

j	$x_{1,j}$	$x_{2,j}$	$x_{3,j}$
0	1	2	3
1	2	1.9148542155126762	1.8171205928321396
2	1.9106582694482719	1.9099276289927102	1.9091929427097283
3	1.9099262803835701	1.9099262335408387	1.9099261866980376
4	1.9099262335408155	1.9099262335408153	1.9099262335408151

This example illustrates the fact that for positive $x_1, \ldots, x_n$, the n sequences $\{x_{k,0}, x_{k,1}, x_{k,2}, \ldots\}$ for $1 \leq k \leq n$ converge to a common limit, which is called the *symmetric mean* of $x_1, \ldots, x_n$ and denoted $M(x_1, \ldots, x_n)$. The proof of convergence for $n \geq 2$ uses Maclaurin's inequality [**8**].

Gauss discovered a formula for $M(x, y)$, or rather its reciprocal, as an integral:

$$\begin{aligned}\frac{1}{M(x, y)} &= \frac{2}{\pi} \int_0^{\pi/2} \frac{dt}{\sqrt{x^2 \cos^2 t + y^2 \sin^2 t}} \\ &= \frac{2}{\pi} \int_0^{\infty} \frac{du}{\sqrt{(u^2 + x^2)(u^2 + y^2)}},\end{aligned}$$

where the second integral comes from the change of variables $u = y \tan t$ and makes evident the symmetry of $M(x, y)$ in x and y. It is still an open problem to give an explicit formula for $M(x_1, \ldots, x_n)$ in general for any $n \geq 3$.

Second application: products of random variables

We will express Maclaurin's inequality, and the case when it becomes an equality, in probabilistic terms. We will assume familiarity with some basic notions in probability theory, such as random variables and their expectation, which are covered in any undergraduate textbook on probability such as [**11**].

Fix positive numbers $x_1, \ldots, x_n$ and place n balls labeled by the x_i's in an urn. Select balls from the urn, one after another, without replacement, until all n balls are picked. Let X_j be the label of the jth ball that is picked, so X_j is a random variable with values in $\{x_1, \ldots, x_n\}$. The outcome of such sampling is a sequence of numbers $(X_1, \ldots, X_n)$. Since we sample without replacement, the value of X_j is affected by the values of $X_1, \ldots, X_{j-1}$, so $X_1, \ldots, X_n$ are *not* independent (unless the x_i's are the same). However, the X_j's are identically distributed. We explain this by an example.

EXAMPLE 2. If we have three balls, numbered as 1, 2, 3, they can be selected in 6 possible ways: 123, 132, 213, 231, 312, 321. If balls 1 and 2 have label x and ball 3 has label y, where $x \neq y$, then the labels we see when selecting the balls in all possible ways are xxy, xyx, xxy, xyx, yxx, yxx. In the first sampling, $X_1 = X_2 = x$ and $X_3 = y$. In the second sampling, $X_1 = X_3 = x$ and $X_2 = y$. Looking at how often x and y occur as a label for the first sampled ball, the second sampled ball, and the third sampled ball, we get x four times and y two times in each position, so X_1, X_2, and X_3 all have the same distribution: $\text{Prob}(X_j = x) = 2/3$ and $\text{Prob}(X_j = y) = 1/3$.

In our urn model, the expectation $\mathbb{E}(X_1 \cdots X_k)$ of the product $X_1 \cdots X_k$ equals $E_k(x_1, \ldots, x_n)$ for $1 \le k \le n$, so Maclaurin's inequality for n terms is equivalent to

$$\mathbb{E}(X_1) \ge \sqrt{\mathbb{E}(X_1 X_2)} \ge \cdots \ge \sqrt[n]{\mathbb{E}(X_1 \cdots X_n)}. \tag{13}$$

This should be contrasted with n independent and identically distributed random variables $X_1, \ldots, X_n$ with positive values, for which (13) has equality in place of $\ge$ everywhere. If $X_1, \ldots, X_n$ are all equal to a common random variable with positive values, then (13) has $\le$ in place of $\ge$ everywhere, by Jensen's inequality.

Maclaurin's inequality gives us information about the covariance of products of the X_j's in the urn model. Fix positive integers ℓ_1 and ℓ_2 such that $\ell_1 + \ell_2 \le n$, and set $Y_1 = X_1 \cdots X_{\ell_1}$ and $Y_2 = X_{\ell_1+1} \cdots X_{\ell_1+\ell_2}$. Since Y_2 has the same distribution as $X_1 \cdots X_{\ell_2}$, $\mathbb{E}(Y_2) = \mathbb{E}(X_1 \cdots X_{\ell_2})$. If $\ell_1 \le \ell_2$, then Maclaurin's inequality implies that $\mathbb{E}(Y_2)^{\ell_1/\ell_2} \le \mathbb{E}(Y_1)$, so

$$\begin{aligned} \mathbb{E}(Y_1 Y_2) &= \mathbb{E}(X_1 \cdots X_{\ell_1+\ell_2}) \le \mathbb{E}(Y_2)^{(\ell_1+\ell_2)/\ell_2} \\ &= \mathbb{E}(Y_2)^{\ell_1/\ell_2}\mathbb{E}(Y_2) \le \mathbb{E}(Y_1)\mathbb{E}(Y_2). \end{aligned} \tag{14}$$

If $\ell_2 \le \ell_1$, then Maclaurin's inequality implies that $\mathbb{E}(Y_1)^{\ell_2/\ell_1} \le \mathbb{E}(Y_2)$, so

$$\mathbb{E}(Y_1 Y_2) \le \mathbb{E}(Y_1)^{(\ell_1+\ell_2)/\ell_1} = \mathbb{E}(Y_1)\mathbb{E}(Y_1)^{\ell_2/\ell_1} \le \mathbb{E}(Y_1)\mathbb{E}(Y_2). \tag{15}$$

Thus $\mathbb{E}(Y_1 Y_2) \le \mathbb{E}(Y_1)\mathbb{E}(Y_2)$ either way, so

$$\operatorname{cov}(Y_1, Y_2) = \mathbb{E}(Y_1 Y_2) - \mathbb{E}(Y_1)\mathbb{E}(Y_2) \le 0.$$

When does $\operatorname{cov}(Y_1, Y_2) = 0$? This would make all inequalities in (14) or (15)—depending on if $\ell_1 \le \ell_2$ or $\ell_2 \le \ell_1$—into equalities. The first inequality in (14) or (15), as an equality, says that $E_{\ell_1+\ell_2}(x_1, \ldots, x_n)^{1/(\ell_1+\ell_2)}$ equals $E_{\ell_2}(x_1, \ldots, x_n)^{1/\ell_2}$ or $E_{\ell_1}(x_1, \ldots, x_n)^{1/\ell_1}$, and either of these implies $x_1 = \cdots = x_n$ by the rule for equality in Maclaurin's inequality. Conversely, if the x_i's are all equal, then Y_1 and Y_2 each take just one value, so $\operatorname{cov}(Y_1, Y_2) = 0$. Therefore, the condition for equality in Maclaurin's inequality is equivalent to the implication $\operatorname{cov}(Y_1, Y_2) = 0 \Rightarrow Y_1$ and Y_2 are constant for all positive integers ℓ_1 and ℓ_2 such that $\ell_1 + \ell_2 \le n$.

Is there a proof of Maclaurin's inequality in the form (13) that is based on probability?

Inequalities on graphs

The scope of Maclaurin's inequality extends to graphs. We will describe how this works and what the corresponding (equivalent) Bernoulli inequality looks like.

Let G be a graph with $n \ge 2$ vertices. We always assume that it has no loops (that is, no edge starts and ends at the same point) and no multiple edges (so that there is at most one edge between any two vertices). A *clique* in G is a complete subgraph of G: Any two vertices of the subgraph are connected by an edge. A k-clique is a clique with k vertices. For instance, a 1-clique is a vertex in G, a 2-clique is a pair of vertices in G and an edge connecting them, and a 3-clique is a set of 3 vertices in G and an edge connecting each pair of these vertices. FIGURE 2 has 1-cliques, 2-cliques, and 3-cliques, but no k-cliques for $k > 3$.

Let $m = m_G$ be the largest integer k such that G has a k-clique, so $1 \le m \le n$. Assign to each vertex v of G a variable X_v. Let $\mathbf{X}$ be the vector of these variables and

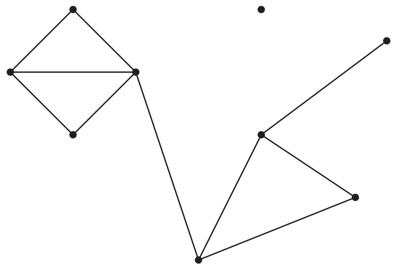


Figure 2 A graph

for $1 \le k \le m$, set

$$e_{k,G}(\mathbf{X}) = \sum_{k\text{-cliques } G_k} \prod_{v \in G_k} X_v.$$

This is a polynomial in the X_v's. Set

$$E_{k,G}(\mathbf{X}) = \frac{e_{k,G}(\mathbf{X})}{\binom{m}{k}}.$$

EXAMPLE 3. In the graph below, $m = 3$. Using the vertex labels from the picture,

$$E_{1,G} = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{3},$$
$$E_{2,G} = \frac{X_1X_2 + X_1X_3 + X_2X_3 + X_4X_5}{3}, \quad \text{and}$$
$$E_{3,G} = X_1X_2X_3.$$

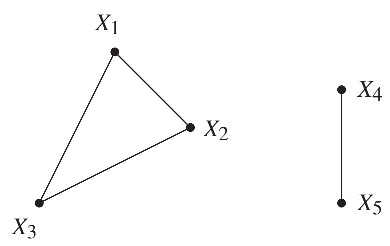


Figure 3 Graph for Example 3

EXAMPLE 4. The complete graph on n vertices, denoted K_n, is the graph with n vertices and one edge connecting every pair of vertices. For instance, K_4 is the edges and vertices of a tetrahedron. When G has n vertices, $m_G = n$ if and only if $G = K_n$.

If $G = K_n$, then $E_{k,G}$ is the kth elementary symmetric mean in the n variables attached to the vertices.

Pick numbers $x_v \ge 0$ for each vertex v in G and let $\mathbf{x}$ be the vector of these numbers. Khadzhiivanov [**4**] proved that

$$E_{1,G}(\mathbf{x}) \ge \sqrt{E_{2,G}(\mathbf{x})} \ge \sqrt[3]{E_{3,G}(\mathbf{x})} \ge \cdots \ge \sqrt[m]{E_{m,G}(\mathbf{x})}. \tag{16}$$

This is Maclaurin's inequality when $G = K_n$. Nikiforov [**10**] has given a recent account of (16), including the case of equality, which is more subtle when $G \ne K_n$.

Earlier we derived the generalized Bernoulli inequality from Maclaurin's inequality by setting each variable equal to 1, except for the last variable, which was set equal to $1 + x$ with $x > -1$. This suggests a natural extension of Bernoulli's inequality to graphs: In (16), set each x_v equal to 1 except for a single x_v, which is $1 + x$ with $x > -1$. When $G = K_n$, this is the generalized Bernoulli inequality. Due to the asymmetry in most graphs, there are usually several Bernoulli inequalities for a graph.

EXAMPLE 5. In Example 3, if we set X_1 (or X_2 or X_3) equal to $1 + x$ with $x > -1$ and the remaining four variables equal to 1, then (16) becomes

$$1 + \frac{2+x}{3} \geq \sqrt{1 + \frac{1+2x}{3}} \geq \sqrt[3]{1+x},$$

while if we set X_4 (or X_5) equal to $1 + x$ with $x > -1$ and the other four variables equal to 1, then (16) is

$$1 + \frac{2+x}{3} \geq \sqrt{1 + \frac{1+x}{3}} \geq \sqrt[3]{1+x}.$$

The equivalence of Maclaurin's inequality and the generalized Bernoulli inequality extends to graphs: Khadzhiivanov's inequality (16) for all G is equivalent to Bernoulli's inequality for all G, and in fact to Bernoulli's inequality for all G with $x = 0$ (that is, (16) for all G with all $x_v = 1$) because we are quantifying over all G. This is explained in [**10**]. For example, let's get the arithmetic-geometric inequality for two terms from (16) for all G with all $x_v = 1$. For positive integers a and b, build a graph with $a + b$ vertices and an edge connecting each of the first a vertices to each of the last b vertices. The inequality (16) on this graph says

$$\frac{\sum_{i=1}^{a+b} x_i}{2} \geq \sqrt{\left(\sum_{i=1}^{a} x_i\right)\left(\sum_{j=a+1}^{a+b} x_j\right)},$$

and when all x_i are 1 it is $(a + b)/2 \geq \sqrt{ab}$. Here, a and b are positive integers. For positive rational numbers a and b, write $a = A/C$ and $b = B/C$ for positive integers A, B, and C. Then $(A + B)/2 \geq \sqrt{AB}$, and dividing both sides by C gives us $(a + b)/2 \geq \sqrt{ab}$. We get $(a + b)/2 \geq \sqrt{ab}$ for all positive real numbers a and b by continuity from the rational case.

REFERENCES

1. E. F. Beckenbach and R. Bellman, *Inequalities,* Springer-Verlag, New York, 1983.
2. D. Cox, The arithmetic-geometric mean of Gauss, *L'Enseignement Mathématique* **30** (1984) 275–330.
3. G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities,* Cambridge Univ. Press, Cambridge, 1934.
4. N. Khadzhiivanov, Inequalities for graphs (Russian), *C. R. Acad. Bulgare Sci.* **30** (1977) 793–796.
5. C. Maclaurin, A Second Letter from Mr. Colin McLaurin to Martin Folkes, Esq.; Concerning the Roots of Equations, with the Demonstration of Other Rules in Algebra, *Phil. Trans.* **36** (1729) 59–96.
6. L. Maligranda, Why Hölder's inequality should be called Rogers' inequality, *Math. Inequal. Appl.* **1** (1998) 69–83.
7. L. Maligranda, The AM-GM inequality is equivalent to the Bernoulli inequality, *The Mathematical Intelligencer* **34** (2012) 1–2. http://dx.doi.org/10.1007/s00283-011-9266-8
8. Math Overflow, http://mathoverflow.net/questions/37576.
9. D. S. Mitrinovic, *Analytic Inequalities,* Springer-Verlag, New York, 1970.
10. V. Nikiforov, An extension of Maclaurin's inequality, http://arxiv.org/abs/math/0608199.
11. S. Ross, *A First Course in Probability* (8th ed.), Pearson Prentice-Hall, Upper Saddle River, NJ, 2010.
12. J. M. Steele, *The Cauchy-Schwarz Master Class: An Introduction to the Art of Inequalities,* Cambridge Univ. Press, Cambridge, 2004.

Summary Maclaurin's inequality is a natural, but nontrivial, generalization of the arithmetic-geometric mean inequality. We present a new proof that is based on an analogous generalization of Bernoulli's inequality. Applications of Maclaurin's inequality to iterative sequences and probability are discussed, along with graph-theoretic versions of the Maclaurin and Bernoulli inequalities.

IDDO BEN-ARI was born, raised, and educated in Haifa, Israel. He received his Ph.D. from the Technion, Israel Institute of Technology, in 2005, did a postdoc at UC Irvine, and joined the department of mathematics at the University of Connecticut in 2008, where he is now an assistant professor. He became interested in the topic of this paper after learning of Maclaurin's inequality from the co-author.

KEITH CONRAD received his undergraduate degree from Princeton in 1992 and his Ph.D. from Harvard in 1997. Following postdocs at Ohio State and UC San Diego, he has been at the University of Connecticut since 2003, where he is currently an associate professor. He first learned about Maclaurin's inequality, the topic of the paper he co-wrote for this MAGAZINE, while preparing a talk for the UConn math club on the arithmetic-geometric mean inequality.

Submissions to the Magazine

Beginning January 1, 2014, items proposed for inclusion in the MAGAZINE should be submitted using the MAGAZINE'S Editorial Manager system. Prepare your article in the form of a pdf, dvi, or doc file, and then follow the instructions at this website:

http://www.editorialmanager.com/mathmag/

Ordinarily manuscripts should not include the authors' names or affiliations. Questions concerning new submissions may be addressed to the Editor Elect at mathmag-editor-elect@maa.org.

Contributions to the Problems section should be sent to Bernardo M. Ábrego, Problems Editor, Department of Mathematics, California State University, Northridge, 18111 Nordhoff St, Northridge, CA 91330-8313 or by e-mail to mathmagproblems@csun.edu.

Other questions and correspondence may be sent to the Editor at mathmag@maa.org.

When Can You Factor a Quadratic Form?

BRIAN G. KRONENTHAL
Kutztown University of Pennsylvania
Kutztown, PA 19530
kronenthal@kutztown.edu

FELIX LAZEBNIK
University of Delaware
Newark, DE 19716
lazebnik@math.udel.edu

Can you factor either of the polynomials

$$Q_1 = 2X_3^2 + 2X_1X_2 - X_1X_3 - 4X_2X_3 - 6X_3X_4 + 3X_1X_4$$

or

$$Q_2 = 2X_3^2 + 2X_1X_2 - X_1X_3 - 4X_2X_3 - 6X_3X_4$$

into a product of polynomials of smaller degree? Feel free to use complex coefficients, but no computers or calculators are allowed.

It turns out that one of Q_1 and Q_2 factors, while the other does not. If you want to check your answer, or do not feel like trying, see the related examples ahead. Of course, we ultimately wish to consider the following general problem:

Given any quadratic form, determine whether it is a product of two linear forms.

If you are uncertain about the meaning of the terms "quadratic form" or "linear form," definitions will appear in the next section. Interestingly enough, this problem has a simple solution, and was answered several centuries ago. However, many people we talked to found it intriguing and were surprised that they had not thought about or seen this problem before. A solution will be a part of what we present in this article. However, our story begins elsewhere.

Problems with terminology?

Undoubtedly, every reader has tried to clarify a notion in one source by consulting another, only to be frustrated that the presentations are inconsistent in vocabulary or notation. Recently, this happened to us in a study of conics. While reading Peter Cameron's *Combinatorics: Topics, Techniques, and Algorithms* [**8**], we encountered a definition of a nonsingular quadratic form $Q = Q(x, y, z)$ as one that

. . . cannot be transformed into a form in less than three variables by any non-singular linear substitution of the variables x, y, z.

For example, this definition implies that the quadratic form $Q = X_1^2 + 2X_1X_2 + X_2^2$ is singular because $Q = (X_1 + X_2)^2$, and applying the nonsingular linear transformation of variables that maps X_1 to $X_1 - X_2$ and X_2 to itself, Q can be rewritten as X_1^2 (a form in only one variable). In contrast, the quadratic form $X_1^2 + 4X_1X_2 + X_2^2$ is

Math. Mag. **87** (2014) 25–36. doi:10.4169/math.mag.87.1.25. © Mathematical Association of America

nonsingular because it cannot be rewritten in only one variable via a nonsingular linear transformation of variables. All this will be made more precise and explained later in this paper.

Though the definition Cameron presents was clear, it did not appear to translate into simple criteria for determining whether a given quadratic form was nonsingular. In searching for such a test, we found that various sources used the word "singular" to describe quadratic forms in (what seemed to be) completely different ways. Complicating matters further was that terms such as "degenerate" and "reducible" started to appear, and these three words were often used interchangeably. For more details regarding this usage, look ahead to the section titled "Related terminology in the literature," and in particular TABLE 1.

In all, we found five criteria related to degeneracy (or nondegeneracy) of quadratic forms $Q = Q(x, y, z)$ in the literature. While we believed them to be equivalent, few sources proved the equivalence of even two of them, and we found only two sources that proved the equivalence of three. Three of the five criteria have immediate generalizations to n dimensions.

Our main motivation for writing this paper was to show once and for all, for ourselves and for the record, that the several conditions that are widely used as definitions are actually equivalent. We found the writing process instructive, and we hope the reader will find what we present to be useful. In particular, many of the proofs we used draw ideas from the basic principles of analysis, algebra, linear algebra, and geometry. We think that some of these equivalences can serve as useful exercises in related undergraduate courses, as they help to stress the unity of mathematics.

Notation and an example

The main object of our study will be a quadratic form and its associated quadric. We will now define these terms. Additional definitions and related results can be found in Hoffman and Kunze [**17**] or in Shilov [**31**], for example. Let $\mathbb{F}$ be a field whose characteristic, denoted char($\mathbb{F}$), is not 2. Examples include the fields of rational numbers, real numbers, and complex numbers, as well as finite fields containing an odd (prime power) number of elements. We view $\mathbb{F}^n$ as the n-dimensional vector space over $\mathbb{F}$. Any $(n-1)$-dimensional subspace of $\mathbb{F}^n$ is called a *hyperplane*. By $\mathbb{F}[X_1, \ldots, X_n]$, we denote the ring of polynomials with (commuting) indeterminants $X_1, \ldots, X_n$ and coefficients in $\mathbb{F}$. It will often be convenient to view a polynomial of k indeterminants as a polynomial of one of them, with coefficients being polynomials of the other $k-1$ indeterminants. For instance, a polynomial in $\mathbb{F}[X_1, X_2, X_3]$ may be viewed as an element of $\mathbb{F}[X_2, X_3][X_1]$, i.e., a polynomial of X_1 whose coefficients are polynomials of X_2 and X_3.

For $f \in \mathbb{F}[X_1, \ldots, X_n]$, let $\mathcal{Z}(f) = \{(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}^n : f(\alpha_1, \ldots, \alpha_n) = 0\}$. The "$\mathcal{Z}$" in $\mathcal{Z}(f)$ stands for the zeros of f. To illustrate, if we consider the polynomial $X_1^2 - X_2$ (with $n = 2$) over the field $\mathbb{R}$ of real numbers, then the graph of $\mathcal{Z}(X_1^2 - X_2)$ in the Cartesian coordinate system with axes X_1 and X_2 is the parabola $X_2 = X_1^2$. For $\boldsymbol{f} = (f_1, \ldots, f_m)$, where all $f_i \in \mathbb{F}[X_1, \ldots, X_n]$, we define $\mathcal{Z}(\boldsymbol{f})$ to be the intersection of all $\mathcal{Z}(f_i)$.

A polynomial $Q \in \mathbb{F}[X_1, \ldots, X_n]$ of the form

$$Q = Q(X_1, \ldots, X_n) = \sum_{1 \le i, j \le n} a_{ij} X_i X_j,$$

where $a_{ij} \in \mathbb{F}$ for all i, j, is called a *quadratic form*. For example, $X_1^2 + X_1 X_2$ is a quadratic form, and so are each of Q_1 and Q_2 from the introduction. We may view a quadratic form either as an abstract algebraic object, or as a function $Q : \mathbb{F}^n \to \mathbb{F}$ defined by $(\alpha_1, \ldots, \alpha_n) \mapsto Q(\alpha_1, \ldots, \alpha_n)$. We use the same notation for the algebraic object and the corresponding function. The set $\mathcal{Z}(Q)$ is often referred to as the *quadric* corresponding to Q, and it is the zero-set of Q when Q is viewed as a function.

Since $X_i X_j = X_j X_i$, we have some flexibility in choosing the coefficients; the coefficients a_{ij} and a_{ji} are interchangeable. We will always choose the coefficients to satisfy $a_{ij} = a_{ji}$. For every Q, we define the $n \times n$ matrix of coefficients $M_Q = (a_{ij})$. So for the quadratic form $Q = X_1^2 + X_1 X_2$ in $n = 2$ variables mentioned just above, the associated matrix is

$$M_Q = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 0 \end{pmatrix}.$$

Let V denote the set of all degree-one polynomials in $\mathbb{F}[X_1, \ldots, X_n]$ having zero constant term, called *linear forms*, together with the zero polynomial. A few examples of linear forms are X_1, X_2, and $X_1 + X_2$ (if $n \geq 2$), and $7X_1 + 2X_5 + X_6$ (if $n \geq 6$). The set V can be viewed as a vector space over $\mathbb{F}$ with basis $\{X_1, \ldots, X_n\}$.

For any linear transformation $\varphi : V \to V$ and any quadratic form Q, we can substitute $\varphi(X_i)$ for X_i in Q for all $i = 1, \ldots, n$. After simplifying the result by combining like terms, we again obtain a quadratic form, which we denote by

$$\widetilde{Q}(X_1, \ldots, X_n) = Q(\varphi(X_1), \ldots, \varphi(X_n)).$$

For $T = (t_1, \ldots, t_n) \in \mathbb{F}^n$, $\widetilde{Q}(T) = Q(\varphi(X_1)(T), \ldots, \varphi(X_n)(T))$.

Next, a comment on derivatives. We treat partial derivatives formally, as is done in algebra. For example, to differentiate Q with respect to X_1, we view Q as an element of $\mathbb{F}[X_2, \ldots, X_n][X_1]$:

$$Q = a_{11}X_1^2 + (2a_{12}X_2 + \cdots + 2a_{1n}X_n)X_1 + \sum_{2 \leq i, j \leq n} a_{ij} X_i X_j$$

and thus

$$\frac{\partial Q}{\partial X_1} = 2a_{11}X_1 + (2a_{12}X_2 + \cdots + 2a_{1n}X_n).$$

In other words, we treat Q as a polynomial in X_1, and differentiate it with respect to X_1. Partial derivatives with respect to the other indeterminants are defined similarly. The gradient $\boldsymbol{\nabla Q}$ of Q is then defined as $\boldsymbol{\nabla Q} = \left(\frac{\partial Q}{\partial X_1}, \ldots, \frac{\partial Q}{\partial X_n}\right)$.

Before stating our main result in the next section, we illustrate it with some examples.

EXAMPLES. Let $n = 4$ and let $\mathbb{F}$ be $\mathbb{R}$, the field of real numbers. The following examples are also valid over all fields of characteristic different from 2 (that is, fields in which $2 \neq 0$ as field elements). Consider the quadratic forms mentioned at the beginning of this article, namely

$$Q_1 = 2X_3^2 + 2X_1X_2 - X_1X_3 - 4X_2X_3 - 6X_3X_4 + 3X_1X_4$$

and

$$Q_2 = 2X_3^2 + 2X_1X_2 - X_1X_3 - 4X_2X_3 - 6X_3X_4.$$

We now examine several properties of Q_1 and Q_2.

1. The matrix associated with Q_1 is
$$M_{Q_1} = \begin{pmatrix} 0 & 1 & -1/2 & 3/2 \\ 1 & 0 & -2 & 0 \\ -1/2 & -2 & 2 & -3 \\ 3/2 & 0 & -3 & 0 \end{pmatrix}.$$
It has row reduced echelon form
$$\begin{pmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & -1/2 & -3/2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$
and thus M_{Q_1} has rank 2. In contrast,
$$M_{Q_2} = \begin{pmatrix} 0 & 1 & -1/2 & 0 \\ 1 & 0 & -2 & 0 \\ -1/2 & -2 & 2 & -3 \\ 0 & 0 & -3 & 0 \end{pmatrix}$$
has rank 4. That is, M_{Q_1} is a singular matrix and M_{Q_2} is nonsingular.
2. Define a linear transformation $\varphi_1 : V \to V$ by
$$\begin{aligned} \varphi_1(X_1) &= X_1 + 2X_3 \\ \varphi_1(X_2) &= \frac{1}{2}X_2 + \frac{1}{2}X_3 - \frac{3}{2}X_4 \\ \varphi_1(X_3) &= X_3 \\ \varphi_1(X_4) &= X_4. \end{aligned}$$
It is a straightforward verification that φ is nonsingular, and that
$$\widetilde{Q}_1 = Q_1(\varphi_1(X_1), \varphi_1(X_2), \varphi_1(X_3), \varphi_1(X_4)) = X_1X_2,$$
which contains only $r = 2$ indeterminants. Furthermore, we comment (without proof) that for any nonsingular linear transformation $\varphi : V \to V$, the simplified polynomial $Q_1(\varphi(X_1), \varphi(X_2), \varphi(X_3), \varphi(X_4))$ contains *at least* two indeterminants: The argument is the same as in the proof of Theorem 1.

   In contrast, for any nonsingular linear transformation $\varphi : V \to V$,
$$\widetilde{Q}_2 = Q_2(\varphi(X_1), \varphi(X_2), \varphi(X_3), \varphi(X_4))$$
contains all four indeterminants. This also follows from the proof of Theorem 1.
3. We now consider the zeros of the gradient fields of Q_1 and Q_2. We have
$$\nabla \boldsymbol{Q_1} = \begin{pmatrix} \partial Q_1/\partial X_1 \\ \partial Q_1/\partial X_2 \\ \partial Q_1/\partial X_3 \\ \partial Q_1/\partial X_4 \end{pmatrix} = \begin{pmatrix} 2X_2 - X_3 + 3X_4 \\ 2X_1 - 4X_3 \\ -X_1 - 4X_2 + 4X_3 - 6X_4 \\ 3X_1 - 6X_3 \end{pmatrix} = 2M_{Q_1}\boldsymbol{X},$$
where $\boldsymbol{X} = (X_1, X_2, X_3, X_4)^t$ is the transpose of (X_1, X_2, X_3, X_4). Therefore, $\mathcal{Z}(\nabla \boldsymbol{Q_1})$ is the null space of the matrix $2M_{Q_1}$. In much the same way,
$$\nabla \boldsymbol{Q_2} = \begin{pmatrix} 2X_2 - X_3 \\ 2X_1 - 4X_3 \\ -X_1 - 4X_2 + 4X_3 - 6X_4 \\ -6X_3 \end{pmatrix} = 2M_{Q_2}\boldsymbol{X}$$

implies that $\mathcal{Z}(\nabla \boldsymbol{Q_2})$ is the null space of the matrix $2M_{Q_2}$. However, while $\mathcal{Z}(\nabla \boldsymbol{Q_1})$ has dimension $2 = 4 - 2 = n - r$, $\mathcal{Z}(\nabla \boldsymbol{Q_2})$ has dimension 0 (i.e., $\mathcal{Z}(\nabla \boldsymbol{Q_2})$ contains only the zero vector).

4. $Q_1 = (X_1 - 2X_3)(2X_2 - X_3 + 3X_4)$, a product of two polynomials that are not scalar multiples of each other. In contrast, Q_2 does not factor into a product of linear polynomials (even over the complex numbers). These results offer a solution to the problem posed at the beginning of this article.
5. It is clear from the above factorization of Q_1 that $\mathcal{Z}(Q_1)$ is the union of two hyperplanes whose equations are $X_1 - 2X_3 = 0$ and $2X_2 - X_3 + 3X_4 = 0$. As $X_1 - 2X_3$ and $2X_2 - X_3 + 3X_4$ are not scalar multiples of one another, these hyperplanes are distinct. This contrasts with $\mathcal{Z}(Q_2)$, which contains only the zero vector.

The main result

The main result of this paper is the following pair of theorems. They establish the equivalence of several definitions of degeneracy (or nondegeneracy) of quadratic forms.

THEOREM 1. *Let $n \geq 2$, $\mathbb{F}$ be a field, and* $\text{char}(\mathbb{F}) \neq 2$. *Let $Q = \sum_{1 \leq i,j \leq n} a_{ij} X_i X_j$ be a nonzero quadratic form in $\mathbb{F}[X_1, \ldots, X_n]$. Then the following statements are equivalent.*

1. *The matrix $M_Q = (a_{ij})$ has rank r.*
2. *There exists a nonsingular linear transformation $\varphi : V \to V$ such that the transformed polynomial $\widetilde{Q} = Q(\varphi(X_1), \ldots, \varphi(X_n))$ contains precisely r of the indeterminants $X_1, \ldots, X_n$; furthermore, for any other nonsingular linear transformation, this number is at least r.*
3. *$\mathcal{Z}(\nabla \boldsymbol{Q})$ is a vector space of dimension $n - r$.*

Define r, the *rank* of a quadratic form Q, as in any of the three equivalent statements listed above. Then r is an integer such that $1 \leq r \leq n$. If $r = 1$ or 2, then we can supplement the above three statements with two more, which appear frequently in the context of conics or quadratic surfaces. In the following, let $\mathbb{K}$ denote a field that is either equal to $\mathbb{F}$, or is a quadratic extension $\mathbb{F}(m)$ of $\mathbb{F}$ for some $m \in \mathbb{K} \setminus \mathbb{F}$ such that $m^2 \in \mathbb{F}$.

THEOREM 2. *Let n, $\mathbb{F}$, and Q be as in Theorem 1, and let $r = 1$ or 2. Then the following statements are equivalent.*

1. *The matrix $M_Q = (a_{ij})$ has rank r.*
2. *There exists a nonsingular linear transformation $\varphi : V \to V$ such that the transformed polynomial $\widetilde{Q} = Q(\varphi(X_1), \ldots, \varphi(X_n))$ contains precisely r of the indeterminants $X_1, \ldots, X_n$; furthermore, for any other nonsingular linear transformation, this number is at least r.*
3. *$\mathcal{Z}(\nabla \boldsymbol{Q})$ is a vector space of dimension $n - r$.*
4. *Q is a product of two linear forms with coefficients in $\mathbb{K}$. These forms are scalar multiples of one another if $r = 1$, and are not if $r = 2$.*
5. *$\mathcal{Z}(Q)$ is a hyperplane in $\mathbb{K}^n$ for $r = 1$. $\mathcal{Z}(Q)$ is the union of two distinct hyperplanes in $\mathbb{K}^n$ for $r = 2$.*

Statements 1, 2, 3, and 5 primarily use terms of linear algebra, with statement 3 having an analysis flavor and statement 5 a geometric one. Statement 4 is algebraic. The statements in Theorems 1 and 2 correspond to the properties of Q_1 and Q_2 discussed in the above examples.

We are now ready for our main definition.

DEFINITION. Let Q be as in Theorem 1, and suppose that Q has rank r. If $1 \leq r < n$, we call Q *degenerate* and *singular*. If instead $r = n$, we call Q *nondegenerate* and *nonsingular*.

If $r = 1$ or 2, we call Q *reducible*. If instead $r \geq 3$, we call Q *irreducible*.

We make several comments about this definition.

- As a nonzero quadratic form can be factored only into the product of two linear forms, our definition of irreducibility corresponds to the one in algebra for polynomials. Indeed, a quadratic form factors into linear forms (over some extension $\mathbb{K}$ of $\mathbb{F}$) for $n \geq 2$ if and only if $1 \leq r \leq 2$.
- Similar definitions can be applied to the case $n = 1$, where every nonzero form $Q = aX_1^2 = (aX_1)X_1$, $a \neq 0$, is called *reducible*, *nondegenerate*, and *nonsingular*.
- We wish to emphasize that $n = 3$ is the only case in which the notions of degeneracy, singularity, and reducibility are equivalent.
- Consider the example from the previous section. Note that since $n = 4$ and $r = 2$, our definition implies that Q_1 is degenerate, singular, and reducible. By contrast, Q_2 is nondegenerate, nonsingular, and irreducible.

Proof of Theorems 1 and 2

First we mention a few well-known results that we will need.

Let R^t denote the transpose of a matrix R, and let $\boldsymbol{X} = (X_1, \ldots, X_n)^t$ be the column vector of indeterminants. Then, in terms of matrix multiplication, we have $(Q) = \boldsymbol{X}^t M_Q \boldsymbol{X}$, a 1×1 matrix. From now on, we will view (Q) as the polynomial Q, and simply write $Q = \boldsymbol{X}^t M_Q \boldsymbol{X}$.

Let $\varphi : V \to V$ be a linear transformation, $Y_i = \varphi(X_i)$ for all i, and let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^t$. Then we define

$$\varphi(\boldsymbol{X}) = (\varphi(X_1), \ldots, \varphi(X_n))^t,$$

and therefore

$$\varphi(\boldsymbol{X}) = (Y_1, \ldots, Y_n)^t = \boldsymbol{Y}.$$

Letting B_φ be the matrix of φ with respect to a basis $\{X_1, \ldots, X_n\}$ of V, we have

$$\boldsymbol{Y} = B_\varphi \boldsymbol{X}.$$

This allows us to use matrix multiplication in order to determine $M_{\widetilde{Q}}$. Indeed, we have

$$\widetilde{Q} = \widetilde{Q}(X_1, \ldots, X_n) = \boldsymbol{X}^t M_{\widetilde{Q}} \boldsymbol{X}$$

and

$$\begin{aligned} \widetilde{Q} = Q(\varphi(X_1), \ldots, \varphi(X_n)) &= Q(Y_1, \ldots, Y_n) \\ &= \boldsymbol{Y}^t M_Q \boldsymbol{Y} \\ &= \left(B_\varphi \boldsymbol{X}\right)^t M_Q \left(B_\varphi \boldsymbol{X}\right) \\ &= \boldsymbol{X}^t \left(B_\varphi^t M_Q B_\varphi\right) \boldsymbol{X}. \end{aligned}$$

The equality $\boldsymbol{X}^t M_{\widetilde{Q}} \boldsymbol{X} = \boldsymbol{X}^t \left(B_\varphi^t M_Q B_\varphi\right) \boldsymbol{X}$, viewed as an equality of 1×1 matrices with polynomial entries, implies that

$$M_{\widetilde{Q}} = B_\varphi^t M_Q B_\varphi.$$

Recall that for any square matrix M and any nonsingular matrix N of the same dimensions, $\operatorname{rank}(MN) = \operatorname{rank}(M) = \operatorname{rank}(NM)$. Therefore, φ nonsingular implies that so is B_φ. Thus,

$$\operatorname{rank}(M_Q) = \operatorname{rank}(M_{\widetilde{Q}}).$$

Finally, we will need the fundamental fact (see, for example, [**31**, Section 7.33(b)]) that given any quadratic form Q with $\operatorname{rank}(M_Q) = r$, there exists a nonsingular linear transformation ψ of V such that if $Q' = Q(\psi(X_1), \dots, \psi(X_n))$, then $M_{Q'}$ is diagonal. In other words,

$$Q' = d_1 X_1^2 + \cdots + d_r X_r^2,$$

where all d_i are nonzero elements of $\mathbb{F}$. This implies that for any nonsingular linear transformation φ of V, $\widetilde{Q} = Q(\varphi(X_1), \dots, \varphi(X_n))$ contains at least r variables X_i. Indeed, if $\widetilde{Q}$ contained $r' < r$ variables, then $r = \operatorname{rank}(M_Q) = \operatorname{rank}(M_{\widetilde{Q}}) \leq r'$, a contradiction.

We are now ready to prove Theorems 1 and 2. Let (k) stand for statement k of Theorem 1 or Theorem 2, $k = 1, \dots, 5$. We need to prove that (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3), and, when $r = 1$ or 2, that (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3) $\Leftrightarrow$ (4) $\Leftrightarrow$ (5).

(1) $\Leftrightarrow$ (2). The statement follows immediately from our remarks above on the diagonal form of Q, as the rank of a diagonal matrix is the number of its nonzero entries.

(1) $\Leftrightarrow$ (3). It is straightforward to verify that $(2M_Q)\boldsymbol{X} = \nabla \boldsymbol{Q}$. Since $\operatorname{char}(\mathbb{F}) \neq 2$, $\mathcal{Z}(\nabla \boldsymbol{Q})$ coincides with the null space of M_Q. Hence,

$$\dim(\mathcal{Z}(\nabla \boldsymbol{Q})) = n - \operatorname{rank}(M_Q) = n - r.$$

This proves the equivalence of statements (1), (2), and (3), and thus concludes the proof of Theorem 1. This equivalence leads to the notion of the rank of a quadratic form Q, which was defined to be r as it occurs in any of these equivalent statements.

We now suppose that the rank of Q is $r = 1$ or 2.

(2) $\Leftrightarrow$ (4). We have $\operatorname{rank}(M_Q) = r = 1$ or 2. As explained above, there exists a nonsingular transformation ψ of V such that $Q' = Q(\psi(X_1), \dots, \psi(X_n)) = d_1 X_1^2 + \cdots + d_r X_r^2$, with all d_i being nonzero elements of $\mathbb{F}$, and $\operatorname{rank}(M_{Q'}) = \operatorname{rank}(M_Q) = r$. If $r = 1$, then $Q' = d_1 X_1^2$, where $d_1 \neq 0$. Then $Q = d_1(\psi^{-1}(X_1))^2$ factors as claimed. If instead $r = 2$, then $Q' = d_1 X_1^2 + d_2 X_2^2$, where both $d_i \neq 0$. Let $m \in \mathbb{K}$ such that $m^2 = -d_2/d_1$. Then

$$\begin{aligned} Q' &= d_1 X_1^2 + d_2 X_2^2 \\ &= d_1 \left(X_1^2 - \frac{-d_2}{d_1} X_2^2 \right) \\ &= d_1 \left(X_1^2 - m^2 X_2^2 \right) \\ &= (d_1 X_1 - d_1 m X_2)(X_1 + m X_2). \end{aligned}$$

The two factors of Q' are independent in the vector space of linear forms; otherwise, $m = -m$, which is equivalent to $m = 0$ because $\operatorname{char}(\mathbb{F}) \neq 2$. This implies that $d_2 = 0$, a contradiction. Hence,

$$Q = \left(d_1\psi^{-1}(X_1) - d_1 m\psi^{-1}(X_2)\right) \cdot \left(\psi^{-1}(X_1) + m\psi^{-1}(X_2)\right).$$

Since ψ^{-1} is nonsingular, $\psi^{-1}(X_1)$ and $\psi^{-1}(X_2)$ are linearly independent over $\mathbb{F}$. Hence, the factors of Q are linearly independent over $\mathbb{K}$. This proves that (2) $\Rightarrow$ (4).

Suppose Q factors over $\mathbb{K}$ such that $Q = a(a_1X_1 + \cdots + a_nX_n)^2$ with $a, a_i \in \mathbb{K}$ and $a \neq 0$. Permuting indices as necessary, suppose $a_1 \neq 0$; then apply the linear substitution φ defined by

$$\varphi(X_1) = \frac{1}{a_1}X_1 - \sum_{i=2}^{n} \frac{a_i}{a_1}X_i \quad \text{and} \quad \varphi(X_i) = X_i \text{ for all } 2 \le i \le n.$$

This transformation is nonsingular, and the resulting quadratic form $\widetilde{Q} = aX_1^2$ has $r = 1$ indeterminant, implying (2).

Suppose instead that Q factors over $\mathbb{K}$ with $Q = (a_1X_1 + \cdots + a_nX_n)(b_1X_1 + \cdots + b_nX_n)$, where $a_i, b_i \in \mathbb{K}$ such that the factors do not differ by a scalar multiple. This is equivalent to the vectors $(a_1, \dots, a_n)$ and $(b_1, \dots, b_n)$ being linearly independent. Therefore, permuting indices as necessary, the vectors $(a_1, a_2), (b_1, b_2) \in \mathbb{F}^2$ are linearly independent with $a_1 \neq 0$ and $b_2 \neq 0$. Apply the linear substitution φ defined by

$$\begin{aligned}
\varphi(X_1) &= \alpha_{11}X_1 + \alpha_{12}X_2 + \cdots + \alpha_{1n}X_n \\
\varphi(X_2) &= \alpha_{21}X_1 + \alpha_{22}X_2 + \cdots + \alpha_{2n}X_n \\
\varphi(X_i) &= X_i \text{ for all } i = 3, \dots, n,
\end{aligned}$$

where (α_{ij}) is the inverse of the matrix

$$\begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ b_1 & b_2 & b_3 & \cdots & b_n \\ 0 & 0 & & & \\ \vdots & \vdots & & I_{n-2} & \\ 0 & 0 & & & \end{pmatrix},$$

and I_{n-2} is the $(n-2) \times (n-2)$ identity matrix. This nonsingular transformation produces the quadratic form $\widetilde{Q} = X_1X_2$, which contains $r = 2$ indeterminants. Note that if there existed a linear transformation $\phi : V \to V$ such that $\hat{Q} = Q(\phi(X_1), \dots, \phi(X_n))$ had only one indeterminant, then

$$1 = \operatorname{rank}(M_{\hat{Q}}) = \operatorname{rank}(M_Q) = \operatorname{rank}(M_{\widetilde{Q}}) = 2,$$

a contradiction. This proves that (4) $\Rightarrow$ (2).

(4) $\Leftrightarrow$ (5). The implication (4) $\Rightarrow$ (5) is obvious, and we concentrate on the converse.

Let $a_1X_1 + \cdots + a_nX_n = 0$ be an equation of a hyperplane W of $\mathbb{K}^n$ such that $W \subseteq \mathcal{Z}(Q)$. Then for every solution $(\alpha_1, \dots, \alpha_n)$ of $a_1X_1 + \cdots + a_nX_n = 0$, $Q(\alpha_1, \dots, \alpha_n) = 0$. As not all a_i are zero, we may assume by permuting the indices as necessary that $a_1 \neq 0$. Dividing by a_1, we rewrite the equation $a_1X_1 + \cdots + a_nX_n = 0$ as $X_1 + a_2'X_2 + \cdots + a_n'X_n = 0$. Viewing Q as an element of $\mathbb{K}[X_2, \dots, X_n][X_1]$ and dividing it by $X_1 + (a_2'X_2 + \cdots + a_n'X_n)$ with remainder, we obtain

$$Q = q \cdot \left(X_1 + \left(a_2'X_2 + \cdots + a_n'X_n\right)\right) + t,$$

with quotient $q \in \mathbb{K}[X_2, \dots, X_n][X_1]$ and remainder $t \in \mathbb{K}[X_2, \dots, X_n]$. Now, we have $Q(w_1, \dots, w_n) = 0 = w_1 + (a_2'w_2 + \cdots + a_n'w_n)$ for every $(w_1, \dots, w_n) \in$

W. Furthermore, for every $(w_2, \ldots, w_n) \in \mathbb{K}^{n-1}$, there exists $w_1 \in \mathbb{K}$ such that $(w_1, \ldots, w_n) \in W$. This implies that $t(w_2, \ldots, w_n) = 0$ for every $(w_2, \ldots, w_n) \in \mathbb{K}^{n-1}$. The following lemma will allow us to conclude that $t = 0$.

LEMMA. *Let $n \geq 1$, $\mathbb{K}$ be a field such that* $\text{char}(\mathbb{K}) \neq 2$, *and $f \in \mathbb{K}[X_1, ..., X_n]$ be a quadratic form that vanishes on $\mathbb{K}^n$. Then $f = 0$.*

Proof. We proceed by induction. For $n = 1$, $f = f(X_1) = aX_1^2$ for some $a \in \mathbb{K}$. Then $0 = f(1) = a$, and so $f = 0$.

Suppose the statement holds for all quadratic forms containing fewer than n indeterminants. We write f as

$$f(X_1, \ldots, X_n) = aX_1^2 + f_2(X_2, \ldots, X_n)X_1 + f_3(X_2, \ldots, X_n),$$

where $a \in \mathbb{K}$, $f_2 \in \mathbb{K}[X_2, \ldots, X_n]$ is a linear form, and $f_3 \in \mathbb{K}[X_2, \ldots, X_n]$ is a quadratic form. As $\text{char}(\mathbb{K}) \neq 2$, we know that 0, 1, and -1 are distinct elements of $\mathbb{K}$. As f vanishes on $\mathbb{K}^n$, f vanishes at every point of the form $(\alpha, \alpha_2, \ldots, \alpha_n) \in \mathbb{K}^n$ with $\alpha \in \{0, 1, -1\}$. Therefore, we obtain the three equations

$$0 = f(0, \alpha_2, \ldots, \alpha_n) = f_3(\alpha_2, \ldots, \alpha_n)$$
$$0 = f(1, \alpha_2, \ldots, \alpha_n) = a + f_2(\alpha_2, \ldots, \alpha_n) + f_3(\alpha_2, \ldots, \alpha_n)$$
$$0 = f(-1, \alpha_2, \ldots, \alpha_n) = a - f_2(\alpha_2, \ldots, \alpha_n) + f_3(\alpha_2, \ldots, \alpha_n)$$

for all $(\alpha_2, \ldots, \alpha_n) \in \mathbb{K}^{n-1}$. By induction hypothesis, the first equation implies that $f_3 = 0$. Then the second and third equations together imply that $a = 0$ and that $f_2(\alpha_2, \ldots, \alpha_n) = 0$ for all $(\alpha_2, \ldots, \alpha_n) \in \mathbb{F}^{n-1}$. Let $f_2 = a_2X_2 + \cdots + a_nX_n$. If $f_2 \neq 0$, then there exists $a_i \neq 0$. Setting $\alpha_i = 1$ and $\alpha_j = 0$ for all $j \neq i$, we obtain

$$0 = f_2(0, \ldots, 0, 1, 0, \ldots, 0) = a_i \neq 0,$$

a contradiction. Thus, $f_2 = 0$, and so $f = 0$. The lemma is proved. ■

Indeed, this lemma implies that $t = 0$. Therefore, Q factors into a product of two linear forms, each defining a hyperplane in $\mathbb{K}^n$. We now proceed based on whether $r = 1$ or 2. If $r = 1$, then $\mathcal{Z}(Q)$ is a hyperplane of $\mathbb{K}^n$. Thus, both factors of Q must define $\mathcal{Z}(Q)$, and so they are nonzero scalar multiples of one another. If instead $r = 2$, then $\mathcal{Z}(Q)$ is the union of two hyperplanes of $\mathbb{K}^n$, corresponding to the two factors of Q. As the hyperplanes are distinct, the factors are not scalar multiples of one another. Therefore, (5) $\Rightarrow$ (4).

This concludes the proof of the theorem. ■

We end this section with the following comment. Let $Q = Q(X_1, \ldots, X_n)$ be a quadratic form of rank $r = 2$ over a field $\mathbb{F}$, $\text{char}(\mathbb{F}) \neq 2$. It easily follows from our arguments that Q is a product of two linear forms over $\mathbb{F}$ if and only if there exists a nonsingular linear transformation $\varphi : V \to V$ such that

$$Q(\varphi(X_1), \ldots, \varphi(X_n)) = X_1^2 - X_2^2.$$

Related terminology in the literature

We now present a table of sources that utilize the statements from Theorems 1 and 2. The letters used in the table are as follows: D for *degenerate*, S for *singular*, and R for *reducible*. The columns labeled (1) to (5) refer to the corresponding statements from the theorems. Entries with a $*$ indicate that the source mentions the statement, but does not use a particular term to describe it. In addition to providing the reader with additional materials, we use this table to illustrate the variety of ways in which the words *degenerate*, *reducible*, and *singular* are used to refer to these statements. In particular, we see that each word is used to describe multiple statements, and that statements (1), (2), (4), and (5) are each referred to by multiple words!

A few additional notes will prove useful before studying the table. While some sources consider forms (often in the $n = 3$ case), others focus on conics in the classical projective plane $PG(2, q)$. As conics are simply quadrics in projective space, it is not surprising that the characterizations of degenerate quadrics and degenerate conics are nearly identical. Thus, we will not make any further attempt to distinguish them. In addition, we found many sources that discuss equivalent notions, but not in the context of quadratic forms or conics; they have therefore been excluded from this table.

The table illustrates the vocabulary in some of the sources that are available to us. The decision to exclude (or include) a source should not be interpreted as a criticism (or endorsement).

TABLE 1: This table shows how the words degenerate (D), reducible (R), and singular (S) are used in the literature to refer to the statements in Theorems 1 and 2. A $*$ indicates that a source refers to the statement, but does not use a particular term to describe it.

	(1)	(2)	(3)	(4)	(5)
[**1**]		D or S			
[**2**]				R	$*$
[**3**]				D	D
[**4**]	S				
[**6**]					R
[**8**]		S			
[**9**]	$*$			D	
[**10**]	$*$		S	R	R
[**11**]	S	$*$			
[**15**]	$*$	$*$	S		
[**16**]	S				
[**19**]	$*$		$*$		$*$
[**20**]	D				
[**23**]	$*$			D	D
[**24**]	$*$		S		
[**25**]	$*$			D	D
[**26**]				$*$	R
[**28**]	D				
[**29**]	$*$		$*$	D or R	$*$
[**30**]	D	$*$			
[**31**]	S				
[**33**]	$*$				D
[**35**]	$*$		S	R	

Concluding remarks

The discussion in previous sections leads to many interesting questions. We will briefly describe some of them.

What about similar studies of higher-order forms? Let $1 \leq d \leq n$, and let f be a polynomial from $\mathbb{F}[X_1, \ldots, X_n]$ such that $f = \sum_{(i_1,\ldots,i_d)} a_{i_1\ldots i_d} X_{i_1} \cdots X_{i_d}$, where all $a_{i_1\ldots i_d} \in \mathbb{F}$, and summation is taken over all integer sequences $(i_1, \ldots, i_d)$, $1 \leq i_1 \leq i_2 \leq \cdots \leq i_d$. Then f is called a d-form of n indeterminants. 1-forms and 2-forms were discussed in the previous sections as linear and quadratic forms, respectively. When $d \geq 3$, the coefficients $a_{i_1\ldots i_d}$ can be considered as entries of a d-dimensional matrix, which has n^d entries total. Such matrices and their determinants have been studied for more than three hundred years, and by many mathematicians, including Cayley, Silvester, Weierstrass, Garbieri, Gegenbauer, Dedekind, Lecat, Oldenburger, and Sokolov. The time and space necessary to define related notions and results is much more that this article allows, and we refer the reader to a monograph by Sokolov [**32**]. It contains a complete classification of forms over the fields of real and complex numbers for $d = 3$ and $n = 2$ or 3, but discussion of even these cases is far from short. (The classification is with respect to the nonsingular linear transformation of variables). While the monograph is in Russian, it contains 231 references in a variety of languages.

The question of factorization of a general d-form with n indeterminants, over a field for $d \geq 3$, seems to be highly nontrivial, and we could not find any useful criteria for this.

The question of classification of quadratic forms over the integers with respect to the unimodular linear transformation of variables is classical, and significant progress was made in this direction in the 18th and 19th centuries, culminating with the work of Gauss. Its source is the problem of representing integers by a given quadratic form with integer coefficients. For example:

> *Let* $Q(X_1, X_2) = X_1^2 - X_1X_2 + 5X_2^2$. *Describe all ordered triples of integers* (n, a, b) *such that* $n = Q(a, b)$.

For related results and their extensions for forms over other rings, see O'Meara [**27**], Buell [**7**], and Conway [**12**]. For new directions and results related to multidimensional determinants, d-forms, see Gelfand, Kapranov, and Zelevinsky [**13**]. For cubic forms in algebra, geometry and number theory, see Manin [**22**].

Acknowledgment We thank attendees of the University of Delaware Discrete Mathematics/Algebra Seminar for their ruthless criticism.

REFERENCES

1. S. Barwick and G. Ebert, *Unitals in Projective Planes,* Springer, New York, 2008. http://dx.doi.org/10.1007/978-0-387-76366-8
2. L. M. Batten, *Combinatorics of Finite Geometries,* Cambridge University Press, Cambridge, 1997.
3. M. K. Bennett, *Affine and Projective Geometry,* John Wiley, New York, 1995.
4. G. Birkhoff and S. MacLane, *A Survey of Modern Algebra,* Macmillan, New York, 1941.
5. R. Bix, *Conics and Cubics, A Concrete Introduction to Algebraic Curves,* Springer, New York, 2006. http://dx.doi.org/10.1007/0-387-39273-4
6. E. Brieskorn and H. Knörrer, *Plane Algebraic Curves,* Birkhäuser, Basel, 1986. http://dx.doi.org/10.1007/978-3-0348-5097-1
7. D. A. Buell, *Binary Quadratic Forms, Classical Theory and Modern Computations,* Springer, New York, 1989. http://dx.doi.org/10.1007/978-1-4612-4542-1
8. P. J. Cameron, *Combinatorics: Topics, Techniques, Algorithms,* Cambridge University Press, Cambridge, 1994.

9. A. D. Campbell, *Advanced Analytic Geometry,* John Wiley, New York, 1938.
10. R. Casse, *Projective Geometry: An Introduction,* Oxford University Press, Oxford, 2006.
11. J. W. S. Cassels, *Rational Quadratic Forms,* London Mathematical Society Monographs **13**, Academic Press, London, 1978.
12. J. H. Conway with the assistance of F. Y. C. Fung, *The Sensual (Quadratic) Form,* Mathematical Association of America, Washington, DC, 2005.
13. I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky, *Discriminants, Resultants, and Multidimensional Determinants,* Birkhäuser, Boston, MA, 1994. http://dx.doi.org/10.1007/978-0-8176-4771-1
14. K. W. Gruenberg and A. J. Weir, *Linear Geometry,* Van Nostrand, Princeton, N.J., 1967.
15. J. W. P. Hirschfeld, *Projective Geometries over Finite Fields,* The Clarendon Press/Oxford University Press, New York, 1998.
16. W. V. D. Hodge and D. Pedoe, *Methods of Algebraic Geometry,* Vol. II, Cambridge Mathematical Library, Book III: General theory of algebraic varieties in projective space, Book IV: Quadrics and Grassmann varieties, Reprint of the 1952 original, Cambridge University Press, Cambridge, 1994.
17. K. Hoffman and R. Kunze, *Linear Algebra,* Prentice-Hall Inc., Englewood Cliffs, N.J., 1971.
18. D. R. Hughes and F. C. Piper, *Projective Planes,* Springer, New York, 1973.
19. W. E. Jenner, *Rudiments of Algebraic Geometry,* Oxford University Press, New York, 1963.
20. A. Kurosh, *Higher Algebra,* Mir, Moscow, 1988.
21. R. Lidl and H. Niederreiter, *Finite Fields,* Encyclopedia of Mathematics and its Applications, 20, Cambridge University Press, Cambridge, 1997.
22. Yu. I. Manin, *Cubic Forms,* North-Holland, Amsterdam, 1986.
23. E. A. Maxwell, *The Methods of Plane Projective Geometry, Based on the Use of General Homogeneous Coordinates,* Cambridge University Press, Cambridge, 1946.
24. R. Miranda, *Algebraic Curves and Riemann Surfaces,* American Mathematical Society, Providence, RI, 1995.
25. P. S. Modenov and A. S. Parkhomenko, *Geometric Transformations. Vol. 2: Projective Transformations,* Academic Press, New York, 1965.
26. M. Namba, *Geometry of Projective Algebraic Curves,* Marcel Dekker, New York, 1984.
27. O. T. O'Meara, *Introduction to Quadratic Forms,* Springer, Berlin, 2000. http://dx.doi.org/10.1007/978-3-642-62031-7
28. M. Reid, *Undergraduate Algebraic Geometry,* Cambridge University Press, Cambridge, 1988.
29. P. Samuel, *Projective Geometry,* Springer, New York, 1988.
30. I. R. Shafarevich, *Basic Algebric Geometry: Varieties in Projective Space,* Springer, New York, 1995.
31. G. E. Shilov, *Linear Algebra,* Dover, New York, 1977.
32. N. P. Sokolov, *Space Matrices and Their Application* (in Russian), Moscow, 1961.
33. D. M. Y. Sommerville, *Analytic Geometry of Three Dimensions,* Cambridge University Press, New York, 1951.
34. J. Stolfi, *Oriented Projective Geometry, A Framework for Geometric Computations,* Academic Press, Boston, MA, 1991.
35. R. J. Walker, *Algebraic Curves,* Princeton University Press, Princeton, NJ, 1950.

Summary Consider the problem of determining, without using a computer or calculator, whether a given quadratic form factors into the product of two linear forms. A solution derived by inspection is often highly nontrivial; however, we can take advantage of equivalent conditions. In this article, we prove the equivalence of five such conditions. Furthermore, we discuss vocabulary such as "reducible," "degenerate," and "singular" that are used in the literature to describe these conditions, highlighting the inconsistency with which this vocabulary is applied.

BRIAN G. KRONENTHAL is an assistant professor of mathematics at Kutztown University of Pennsylvania. He earned a B.S. from Lafayette College (Easton, Pennsylvania) and a Ph.D. from the University of Delaware, where his advisor was F. Lazebnik. His research problems are currently motivated by extremal combinatorics; however, he uses tools from a variety of mathematical areas. In addition to teaching and research, he enjoys playing tennis and rooting for the Philadelphia Phillies.

FELIX LAZEBNIK is a professor of mathematics at the University of Delaware. His research interests are in graph theory and combinatorics, and as tools, he likes to use number theory, algebra, geometry, and simple analysis. After receiving his M.S. from Kiev State University under L. A. Kalužnin and M. H. Klin, for four and a half years he worked as a mathematics teacher in a high school. He received his Ph.D. from the University of Pennsylvania, where his advisor was H. S. Wilf. He likes both teaching and doing research. He enjoys literature, history of mathematics and sciences, art of Leonard Cohen, movies, nature, travels and food, especially when these can be discussed or shared with friends and family.

NOTES

Defining Exponential and Trigonometric Functions Using Differential Equations

HARVEY DIAMOND
West Virginia University
Morgantown, WV 26506
diamond@math.wvu.edu

Students in science and mathematics are exposed to transcendental functions relatively early, in the sense that formal definitions of these functions must await years of further study. The formal definitions usually come in a standard calculus course.

The first encounter with an "official" definition sometimes comes after the definite integral is available, when $\ln x$ is defined by $\ln x = \int_1^x (1/t)dt$. Then $\exp(x)$ is defined as the inverse function of $\ln x$. This is the approach, for instance, in [**1**]. It works as a definition, and provides a good opportunity for deriving the properties of $\ln x$, but somehow it reverses the natural order of things and makes $\ln x$ depend on an inconvenient exception (the power that doesn't integrate as a power). We might, with a similar approach [**2**, p. 160], define $\sin x$ by its inverse function $\sin^{-1} x = \int_0^x (1/\sqrt{1-t^2})dt$, which at least has the virtue of representing the arc length of a unit circle from (0, 1) to $(x, \sqrt{1-x^2})$. But this approach is rare, perhaps because proving the standard identities is awkward.

A bit later in calculus, $\sin x$, $\cos x$, $\exp(x)$ can be defined in a "cleaner" way using power series. While power series have appealing simplicity, and provide a unifying framework for defining functions, they do not lend themselves directly to proving many fundamental properties, such as the periodicity of $\sin x$ and $\cos x$, the multiplicative property $\exp(t)\exp(x) = \exp(t+x)$, even the fact that $e^x \to 0$ as $x \to -\infty$.

Indeed, we can use all of these formal definitions to show students that in mathematics, precise, simple, and useful definitions for simple concepts can be hard to come by.

In this note we suggest that these functions—$\exp(x)$, $\sin x$, $\cos x$—can be defined most naturally through their governing differential equations. These equations are beautifully simple, and all of the functions' basic properties follow with almost no calculation, using the ideas of existence and uniqueness for initial value problems.

This material can be easily presented as part of an introductory differential equations course. (We would like to argue further, that with the computational and graphics tools now at our disposal, differential equations, initial value problems, and these ideas could be introduced much earlier in the calculus sequence, but that discussion is for another time.) This discussion would also show students that in mathematics, we never stop searching for the most natural ways to define objects, and that basic objects can look simpler through the use of more advanced concepts.

Nothing presented here is new. The multiplicative property of the exponential function, as well as analogous trigonometric identities, are special cases of the functional

Math. Mag. **87** (2014) 37–42. doi:10.4169/math.mag.87.1.37. © Mathematical Association of America

properties of the transition matrix for linear systems $\mathbf{x}' = A\mathbf{x}$, and the analysis of solutions of differential equations through their governing equations (e.g., Sturm–Liouville theory) is similarly well-established. Our intended contribution is the packaging and presentation for undergraduates.

The basic theorems

We begin by stating the basic existence and uniqueness theorems for differential equations, and some of their immediate consequences. These theorems hold in much greater generality, but we need only the simplest cases.

THEOREM 1. *If q, x_0, and a are real numbers, then there is a unique differentiable function $y : \mathbb{R} \to \mathbb{R}$ such that*

$$y'(x) + q\,y(x) = 0 \tag{1}$$

for all $x \in \mathbb{R}$ and

$$y(x_0) = a.$$

Any function $y : \mathbb{R} \to \mathbb{R}$ that satisfies (1) for all $x \in \mathbb{R}$ is called a *solution* of the differential equation $y' + qy = 0$. The equation has many solutions, but only one satisfying any particular *initial condition* such as $y(x_0) = a$.

THEOREM 2. *If p, q, x_0, a, and b are real numbers, then there is a unique differentiable function $y : \mathbb{R} \to \mathbb{R}$ such that*

$$y''(x) + p\,y'(x) + q\,y(x) = 0 \tag{2}$$

for all $x \in \mathbb{R}$ and

$$y(x_0) = a, \quad y'(x_0) = b.$$

A function satisfying (2) is called a *solution* of the differential equation $y'' + py' + qy = 0$.

We state the following immediate consequences in the case of second-order equations, as in Theorem 2. They hold as well for first-order equations.

THEOREM 3. *Let p and q be real numbers.*

(a) *If $u : \mathbb{R} \to \mathbb{R}$ and $v : \mathbb{R} \to \mathbb{R}$ are solutions of the differential equation $y'' + py' + qy = 0$ and α and β are any real constants, then $y = \alpha u + \beta v$ is also a solution.*

(b) *If ϕ is a solution of $y'' + py' + qy = 0$ and c is a real number, and $y : \mathbb{R} \to \mathbb{R}$ is defined by $y(x) = \phi(x - c)$ for all $x \in \mathbb{R}$, then y is a solution.*

(c) *If ϕ is a solution, then its derivative ϕ' is a solution.*

Each part follows by direct verification.

The exponential function: definition and properties

DEFINITION. The function $\exp : \mathbb{R} \to \mathbb{R}$ is the unique solution of the differential equation

$$y' = y$$

satisfying the initial condition $y(0) = 1$.

From the definition, we know that exp is its own derivative and that $\exp(0) = 1$.

If c is any real number, we can substitute directly to see that $y(x) = c\exp(x)$ is the solution of $y' = y$ with $y(0) = c$. For any $t \in \mathbb{R}$, we can let $c = \exp(t)$, so that $\exp(t)\exp(x)$ is a solution of $y' = y$ with $y(0) = \exp(t)$. From Theorem 3, part (b), the function $y(x) = \exp(x + t)$ is a solution of the same equation satisfying the same initial condition, and so the two solutions must be identical; that is, we must have $\exp(x + t) = \exp(t)\exp(x)$ for all real x and t—the multiplicative property of the exponential function. In particular, $\exp(x)\exp(-x) = \exp(0) = 1$ and for any integer n, we have $\exp(nx) = [\exp(x)]^n$. We define $e = \exp(1)$ and the multiplicative property then shows us that $[\exp(p/q)]^q = \exp(p) = e^p$ so that $\exp(p/q) = e^{p/q}$ for any integers p, q. For rational r, we therefore have $\exp(r) = e^r$. The function $\exp(x)$ then gives us the unique continuous extension of e^x from the rational numbers to the real numbers.

Since $\exp(x)$ is clearly positive and increasing (we cannot have $\exp(a) = 0$, lest we obtain the contradiction $\exp(x) \equiv 0$ as the solution of $y' = y$, $y(a) = 0$) we can then define the inverse function $\exp^{-1}(x) \equiv \ln x$ and develop its properties in the usual way (e.g., if $u = \exp(a)$ and $v = \exp(b)$, then $uv = \exp(a + b)$, from which $\ln(uv) = a + b = \ln u + \ln v$ follows).

From the differential equation, it is clear that the nth derivative of $\exp(x)$ is equal to $\exp(x)$, and we can expand in the usual Taylor series from there.

The limit formula for e

From the fact that $\exp(1) = e = (\exp(1/n))^n$, we would like to obtain the famous limit

$$\lim_{n\to\infty} (1 + 1/n)^n = e.$$

The traditional approach via the derivative of $\ln x$ requires the machinery of inverse functions. We take a more direct approach.

From the mean value theorem (MVT), we know that

$$\frac{\exp(1/n) - \exp(0)}{1/n} = \exp(c)$$

for some c satisfying $0 < c < 1/n$. Since $\exp(x)$ is increasing, we have

$$\exp(0) < \frac{\exp(1/n) - \exp(0)}{1/n} < \exp(1/n).$$

From the left inequality, we have $1 + 1/n < e^{1/n}$, and from the right, we have

$$e^{1/n}\left(1 - \frac{1}{n}\right) < 1,$$

so that

$$\left(1 + \frac{1}{n}\right)\left(1 - \frac{1}{n}\right)e^{1/n} < 1 + \frac{1}{n}.$$

Combining the two inequalities and raising both sides to the n power,

$$\left(1 - \frac{1}{n^2}\right)^n e < \left(1 + \frac{1}{n}\right)^n < e.$$

Now, using the MVT, when $0 < x < 1$ and $k > 1$, we have

$$\frac{(1-x)^k - 1}{x} = -k\,(1-c)^{k-1} > -k$$

so that $(1-x)^k > 1 - kx$ and in particular

$$\left(1 - \frac{1}{n^2}\right)^n > 1 - \frac{1}{n}.$$

We obtain the inequality $(1 - 1/n)e < (1 + 1/n)^n < e$ and the limit formula follows from the squeeze theorem.

Sines and cosines: definitions and properties

DEFINITIONS.

- The cosine function $\cos x$ is the unique solution of $y'' + y = 0$ satisfying the initial conditions $y(0) = 1$, $y'(0) = 0$.
- The sine function $\sin x$ is the unique solution of $y'' + y = 0$ satisfying the initial conditions $y(0) = 0$, $y'(0) = 1$.

The basic properties and identities for $\sin x$ and $\cos x$ can be derived as follows.

(a) $\frac{d}{dx}\sin x = \cos x$, $\frac{d}{dx}\cos x = -\sin x$

Define $\phi(x) = \sin x$. Since ϕ is a solution of $y'' + y = 0$, we know from Theorem 3, part (c) that $y = \phi'$ is also a solution. We have $y(0) = \phi'(0) = 1$, while $y'(0) = \phi''(0) = -\phi(0) = 0$. The conditions for $y = \phi'(x)$ are the same initial conditions as those satisfied by the solution $y = \cos x$ and so we obtain $(\sin x)' = \cos x$.

Similarly, if $\phi(x) = \cos x$, then $\phi'(0) = 0$ and $\phi''(0) = -\phi(0) = -1$, so $y = \phi'(x) = (\cos x)'$ is a solution of $y'' + y = 0$ with the same initial conditions as $-\sin x$, and we obtain $(\cos x)' = -\sin x$.

(b) $\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$, $\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$

Consider $\phi(x) = \sin(x + a)$. Then $\phi(x)$ is a solution of $y'' + y = 0$ with $\phi(0) = \sin a$ and $\phi'(0) = \cos a$ so that $\phi(x) = \sin(x + a) = \sin a \cos x + \cos a \sin x$ must hold, since the function on the right is a solution of $y'' + y = 0$ that satisfies the same initial conditions as $\sin(x + a)$.

If we consider $\phi(x) = \cos(x + a)$, then $\phi(x)$ is a solution of $y'' + y = 0$ with $\phi(0) = \cos a$ and $\phi'(0) = -\sin a$ so that $\phi(x) = \cos(x + a) = \cos a \cos x - \sin a \sin x$.

These identities, usually proved geometrically with considerable work, fall out effortlessly in our approach.

(c) $\sin^2 x + \cos^2 x = 1$

If we multiply both sides of $y'' + y = 0$ by y', then we obtain $y'y'' + yy' = \frac{1}{2}\frac{d}{dx}[(y')^2 + (y)^2] = 0$, so $(y')^2 + (y)^2 = C$ for any solution y. In particular, if $y = \sin x$, then we have $(\cos x)^2 + (\sin x)^2 = C$, and by substituting the values at $x = 0$, we find that $C = 1$.

This places the point $(\cos t, \sin t)$ on the unit circle for any t. The geometric significance of t as arc length, which finally connects $\sin t$ and $\cos t$ to their geometric origins, is shown further below.

(d) $\cos(-x) = \cos x$, $\sin(-x) = -\sin x$

We can verify directly that $\cos(-x)$ and $\sin(-x)$ are solutions of $y'' + y = 0$, and that they satisfy the same initial conditions as $\cos x$ and $-\sin x$, respectively. (In general, if $\phi(x)$ is any solution of $y'' + y = 0$, then $\phi(-x)$ is also a solution.)

(e) There exists a first positive zero of $\cos x$; and if this zero is denoted by p, then we have $\cos(x + p) = -\sin x$ and $\sin(x + p) = \cos x$, leading to

$$\cos(x+2p) = -\cos x, \qquad \sin(x+2p) = -\sin x,$$
$$\cos(x+4p) = \cos x, \qquad \sin(x+4p) = \sin x,$$
$$\cos(p-x) = \sin x, \qquad \sin(p-x) = \cos x.$$

To show that there is a first positive zero, note that on any interval $[0, a]$ for which $\cos x > 0$, we have $(\sin x)' = \cos x > 0$, so $\sin x$ is positive and increasing on $(0, a]$, while $(\cos x)' = -\sin x$ shows that $\cos x$ is positive and decreasing. Suppose that $\cos x$ were positive for all $x > 0$. Applying the MVT to $f(x) = \cos x$ at $x = t$ and $x = 2t$:

$$\frac{\cos t - \cos 2t}{t} = \sin c > \sin t,$$

for some c satisfying $t < c < 2t$. Since the left side has limit zero as $t \to \infty$ and $\sin t$ is (under our assumption) positive and increasing for all $t > 0$, a contradiction is obtained. So $\cos x$ must have a first positive zero. Later, we make the connection that $p = \pi/2$.

We have $\cos p = 0$ and (since $\sin x$ has been increasing on $[0, p]$) $\sin p = 1$. We obtain $\cos(x + p) = -\sin x$, $\sin(x + p) = \cos x$ by using the identities. We reapply these results twice to obtain

$$\cos(x+2p) = -\sin(x+p) = -\cos x,$$
$$\sin(x+2p) = \cos(x+p) = -\sin x,$$
$$\cos(x+4p) = -\cos(x+2p) = \cos x,$$
$$\sin(x+4p) = -\sin(x+2p) = \sin x.$$

The results $\cos(p - x) = \sin x$, $\sin(p - x) = \cos x$ follow from applying the identity.

(f) (Connection with geometry)

Let us parameterize the unit circle $x^2 + y^2 = 1$ with respect to arc-length s, by a parameterization $(x(s), y(s))$. We have

$$\frac{d}{ds}(x^2 + y^2) = 2x\frac{dx}{ds} + 2y\frac{dy}{ds} = 0.$$

The derivatives must then satisfy $(\frac{dx}{ds}, \frac{dy}{ds}) = C(-y, x)$ for some $C = C(s)$. However, $\|(\frac{dx}{ds}, \frac{dy}{ds})\| = 1$ in an arc length parameterization, so $|C| = 1$, and if we want the parameterization to be counterclockwise we need $C = 1$. Hence $(\frac{dx}{ds}, \frac{dy}{ds}) = (-y, x)$ so that $\frac{dx}{ds} = -y$ and $\frac{dy}{ds} = x$. It follows that $x'' + x = 0$ and $y'' + y = 0$. Starting the parameterization off at $(1, 0)$ when $s = 0$, we have $x = \cos s$, $y = \sin s$, the classical definitions of cos and sin. In particular, the first positive zero of $x = \cos s$ occurs at $s = \pi/2$, the value of p defined in paragraph (e) above.

(Note: A similar development of properties and identities can be performed on solutions of $y'' - y = 0$, calling the fundamental solutions $\cosh x$ and $\sinh x$, respectively. Then the connection can be made to exponential functions by observing that e^x, e^{-x} are also solutions and then, for instance, that $\phi(x) = (e^x + e^{-x})/2$ satisfies the same initial conditions as $\cosh x$, and so the two must be identical.)

Conclusion

From the simplest possible differential equations, namely $y' = y$ and $y'' + y = 0$, we can rigorously define $\exp(x)$, $\cos x$, and $\sin x$ and easily derive all of their standard properties. In the case of $\cos x$ and $\sin x$, we obtain all these properties without reference to geometry, except when we show that these are the very same trigonometric functions from the unit circle and the triangle.

REFERENCES

1. C. Edwards and D. Penney, *Calculus and Analytic Geometry*, 3rd ed., Prentice-Hall, Upper Saddle River, NJ, 1990.
2. J. Rosenblatt and S. Bell, *Mathematical Analysis for Modeling*, CRC Press, Boca Raton, FL, 1998.

Summary This note addresses the question of how to rigorously define the functions $\exp(x)$, $\sin(x)$, and $\cos(x)$, and develop their properties directly from that definition. We take a differential equations approach, defining each function as the solution of an initial value problem. Assuming only the basic existence/uniqueness theorem for solutions of linear differential equations, we derive the standard properties and identities associated with these functions. Our target audience is undergraduates with a calculus background.

To appear in *The College Mathematics Journal,* March 2014

Articles

Power Series for Up-Down Min-Max Permutations *By Fiacha Heneghan and Kyle Petersen*

Challenging Magic Squares for Magicians *By Arthur Benjamin and Ethan Brown*

Classifying Nilpotent Maps via Partitions Diagrams *By Nicholas Loehr*

The Fastest Way Not to Run a Four Minute Mile *By Michael Bolt, Anthony Meyer, and Nicholas Visser*

Truck Versus Human: Mathematics Under Pressure *By Elizabeth Field, Rachel Ivison, Amanda Reyher, and Steven Warner*

A Single Family of Semigroups with Every Positive Rational Commuting Probability *By Michelle Soule*

Cookie Monster Devours Naccis *By Leigh Braswell and Tanya Khovanova*

Matroids on Groups? *By Jeremy LeCrone and Nancy Neudauer*

Some Logarithmic Approximations for π and e

POO-SUNG PARK
Kyungnam University
Changwon 631-701, Republic of Korea
pspark@kyungnam.ac.kr

The author found an approximation for π. This value is slightly bigger than π by about 3×10^{-5}. It may be useful where other logarithms are involved:

$$\pi \approx \log_5 157 = \log_5(1 + 1 + 5 + 5^2 + 5^3) = 3.1416221\ldots.$$

If we insert more terms, we can obtain the better approximation

$$\pi \approx \log_5 \left(\frac{4}{5^3} + \frac{4}{5^2} + \frac{4}{5} + 1 + 1 + 5 + 5^2 + 5^3 \right) = 3.14159049\ldots$$

that has an absolute error $< 2.2 \times 10^{-6}$. These approximations are derived from the observation

$$5^\pi = 156.9925\ldots = 1111.44401\ldots_5 \approx 1112_5.$$

A similar method yields an approximation for e:

$$e \approx \log_8 285 = \frac{\log(5(1 + 7 + 7^2))}{\log(1 + 7)} = 2.7182727\ldots,$$

that has an absolute error $< 9.2 \times 10^{-6}$. This is derived from

$$8^e = 285.0054\ldots = 555.0015\ldots_7.$$

Acknowledgment This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2013R1A1A2010614).

Summary We offer the approximations $\pi \approx \log_5 157$ and $e \approx \log_8 285$, which may be useful where other logarithms are involved.

A 5-Circle Incidence Theorem

J. CHRIS FISHER
Regina, Canada S4S 6M1
fisher@math.uregina.ca

LARRY HOEHN
Austin Peay State University
Clarksville, TN 37044
hoehnl@apsu.edu

EBERHARD M. SCHRÖDER
FB Mathematik d. Universität
20146 Hamburg, Germany
eberhard.schroeder@gmx.net

The theorem

Our main result is a surprising theorem of Euclidean geometry that was discovered with the help of a computer graphics program. Fortunately, it comes with a proof that is as elementary as the theorem.

Let $A_1, A_2, \ldots, A_5$ be five points situated in the Euclidean plane so that no three of them are collinear, and A_iA_{i+2} is not parallel to $A_{i-1}A_{i+1}$ for any i (where indices are calculated modulo 5). Define five further points $B_1, B_2, \ldots, B_5$ by

$$B_i = A_iA_{i+2} \cap A_{i-1}A_{i+1}.$$

These ten distinct points define five circles $A_iB_iA_{i+1}$, which we denote by Γ_i (as in FIGURE 1). Finally, we denote by g_i the radical axis of Γ_{i-1} and Γ_i (that is, the common chord or, if the circles happen to be tangent at A_i, the common tangent).

THEOREM. *The five lines g_1, ..., g_5 are concurrent or parallel.*

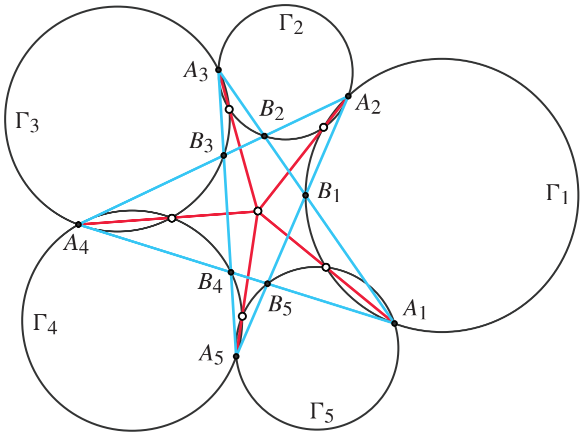


Figure 1 The 5-circle theorem

Math. Mag. **87** (2014) 44–49. doi:10.4169/math.mag.87.1.44. © Mathematical Association of America

The theorem implies that either (a) the g_i are concurrent and the five circles Γ_i have the same radical center, namely the common point of the g_i (which is the center of a circle orthogonal to the Γ_i when the common point is exterior to the five circles), or (b) the g_i are parallel, in which case the centers of the five circles are collinear. We shall use the standard terminology: Lines in the same plane that are concurrent or parallel are said to *lie in a pencil*.

Our proof follows from theorems of Euclid (III.35 and 36), Ceva, and Menelaus, topics found in geometry texts such as [**3**], whose notation we follow. Afterward, we will discuss what we believe to be the true nature of the theorem: The result belongs to the theory of affine metric planes over arbitrary fields, with circles being replaced by families of conics. We shall also touch on the possibility of an n-circle theorem.

Prerequisites

Directed distances We use the same symbol XY to indicate the line determined by the points X and Y and the directed distance from X to Y. The context will always make clear the intended meaning.

Because the distance XY is directed, we have in general $YX = -XY$. But since we will consider distances only in the form of products and ratios, we will never need to specify the direction of a line. It is enough to know that if X, Y, Z, W are points on a line, then the directed distances XY and ZW have the same sign if the direction from X to Y is the same as the direction from Z to W. It follows that if Z is a point on the line XY, then

$$XZ + ZY = XY,$$

whatever the relative positions of the points. Further, the ratio XZ/ZY is positive exactly when Z is between X and Y.

Ceva and Menelaus We will need two theorems that involve directed distances. Consider noncollinear points A, B, C, and additional points X, Y, Z on the lines BC, CA, AB, respectively.

CEVA'S THEOREM. [**3**, 53] *The lines AX, BY, and CZ lie in a pencil if and only if*

$$\frac{BX}{XC} \cdot \frac{CY}{YA} \cdot \frac{AZ}{ZB} = 1.$$

MENELAUS'S THEOREM. [**3**, 66] *The points X, Y, and Z are collinear if and only if*

$$\frac{BX}{XC} \cdot \frac{CY}{YA} \cdot \frac{AZ}{ZB} = -1.$$

Points at infinity Ceva's Theorem and Menelaus's Theorem remain true if some or all of X, Y, and Z are "points at infinity." To say that X is the point at infinity on the line BC, for example, is shorthand for the case that the line AX is parallel to BC, and in this case we interpret the ratio BX/XC as -1 and the ratio XB/XC as $+1$. In this paper, a point is finite unless we specifically allow the possibility that it is a point at infinity.

The radical axis If Φ_1 and Φ_2 are intersecting circles, their *radical axis* is the line through their two points of intersection. If the circles are tangent, their radical axis is the line through their point of tangency that is tangent to both circles. In FIGURE 1, each line g_i is the radical axis of the circles Γ_i and Γ_{i-1}.

We need one fact about the radical axis. Let X be any point on the radical axis of Φ_1 and Φ_2. If a line through X meets Φ_1 at P and Q and meets Φ_2 at R and S, then $XP \cdot XQ = XR \cdot XS$. A proof is given by Coxeter and Greitzer [**3**, Sec. 2.2], who define radical axes more generally and provide a rich treatment of their properties.

Focus on two circles

We begin the proof of the theorem by considering the pairs of circles shown in the configurations of FIGURE 2.

LEMMA 1. *Let Φ_1 and Φ_2 be two intersecting circles with radical axis g, and let h be a line that intersects the circle Φ_1 at P and Q, and the circle Φ_2 at R and S, where P, Q, R, S are four distinct points. If h intersects g at a point X, then*

$$\frac{XQ}{XR} = \frac{SQ}{PR} \tag{1}$$

and

$$\frac{XP}{XQ} = \frac{PR}{RQ} \cdot \frac{PS}{SQ}. \tag{2}$$

If h is parallel to g (that is, if X is the point at infinity on g), then the conclusions remain true if we take $XQ/XR = XP/XQ = 1$.

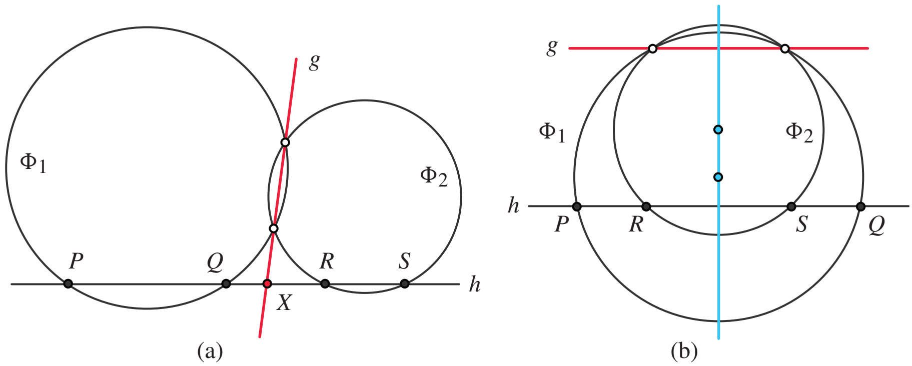


Figure 2 (a) Lemma 1. (b) When h is parallel to g.

Proof. First, assume that h intersects g in a finite point X (FIGURE 2a). According to [**3**, p. 30], we have $g \cap \Phi_1 = g \cap \Phi_2 = \Phi_1 \cap \Phi_2$. Since $P \in g$ would imply $P \in g \cap \Phi_1 = g \cap \Phi_2$ and therewith $P \in h \cap \Phi_2 = \{R, S\}$, we deduce that $P \notin g$ and $g \neq h$. Similarly, we get $Q, R, S \notin g$. From our fact about the radical axis (above), we have $XP \cdot XQ = XR \cdot XS$. Now,

$$XP \cdot XQ = XR \cdot XS \quad \Leftrightarrow \quad \frac{XP}{XR} = \frac{XS}{XQ} \quad \Leftrightarrow \quad \frac{XR+RP}{XR} = \frac{XQ+QS}{XQ}$$

$$\Leftrightarrow \quad \frac{RP}{XR} = \frac{QS}{XQ} \quad \Leftrightarrow \quad \frac{XQ}{XR} = \frac{SQ}{PR},$$

which gives us equation (1). Also,

$$XP \cdot XQ = XR \cdot XS \quad \Leftrightarrow \quad \frac{XQ}{XR} = \frac{XS}{XP} \quad \Leftrightarrow \quad \frac{XR+RQ}{XR} = \frac{XP+PS}{XP}$$
$$\Leftrightarrow \quad \frac{RQ}{XR} = \frac{PS}{XP} \quad \Leftrightarrow \quad \frac{XP}{XR} = \frac{PS}{RQ},$$

whence (using the final equality in each of the above lines)

$$\frac{XP}{XQ} = \frac{XP}{XR} \cdot \frac{XR}{XQ} = \frac{PS}{RQ} \cdot \frac{PR}{SQ} = \frac{PR}{RQ} \cdot \frac{PS}{SQ},$$

which gives us equation (2).

If h is parallel to g—that is, if X is the point at infinity—then we can verify from the symmetry of the situation (FIGURE 2b) that

$$1 = \frac{SQ}{PR}$$

and

$$1 = \frac{PR}{RQ} \cdot \frac{PS}{SQ},$$

which are the conclusions of Lemma 1, if we take $XQ/XR = XP/XQ = 1$. ■

Three lines meet

We next prove a lemma from which our main theorem immediately follows. To avoid being overwhelmed by subscripts, we find it convenient to introduce neutral letters (which will be matched up with the A_i's, B_i's, and C_i's subsequently); furthermore, the configuration featured in the lemma is somewhat more general than the 5-circle configuration that is our main concern.

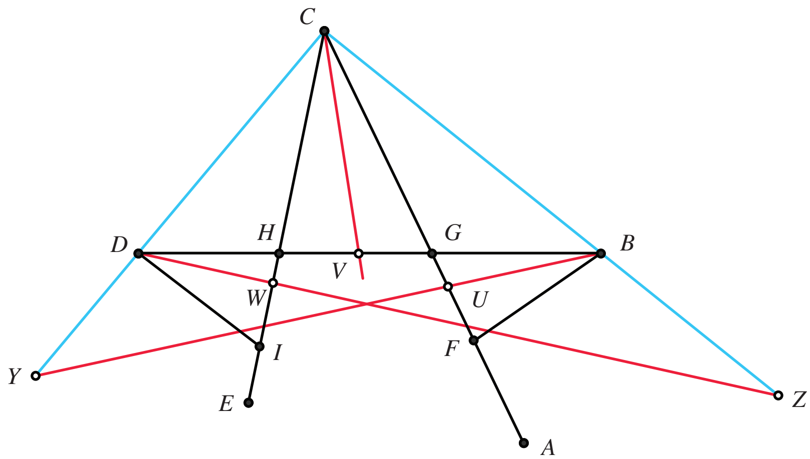


Figure 3

LEMMA 2. *Nine distinct points $A, \ldots, I$ are arranged as in* FIGURE 3 *with A, C, E noncollinear, while A, C, F, G are collinear, as are C, E, H, I and B, D, G, H. Assume that U, V, W are three further points, finite or infinite, which belong to AC, BD, CE, respectively, and satisfy*

$$\frac{UC}{UG} = \frac{CA}{AG} \cdot \frac{CF}{FG}, \quad \frac{VD}{VB} = \frac{GD}{HB}, \quad \frac{WH}{WC} = \frac{HI}{IC} \cdot \frac{HE}{EC}. \tag{3}$$

Then the lines BU, CV, DW lie in a pencil if and only if

$$\frac{CF}{FG} \cdot \frac{GB}{BH} \cdot \frac{HE}{EC} = \frac{GA}{AC} \cdot \frac{HD}{DG} \cdot \frac{CI}{IH}. \tag{4}$$

Proof. We require two further points, finite or infinite; call them Y and Z, where Y is common to lines BU and CD, and Z to lines DW and BC.

We apply Ceva's theorem to triangle *BCD* and its cevians BY, CV, and DZ to deduce that the lines BU, CV, and DW lie in a pencil if and only if

$$\frac{CY}{YD} \cdot \frac{DV}{VB} \cdot \frac{BZ}{ZC} = 1. \tag{5}$$

Next, apply Menelaus's theorem to triangle *CDG* and the triad of collinear points B, U, Y, and to triangle *BCH* and the collinear points D, W, Z, to obtain

$$\frac{CY}{YD} = \frac{UC}{UG} \cdot \frac{GB}{BD} \quad \text{and} \quad \frac{BZ}{ZC} = \frac{BD}{DH} \cdot \frac{WH}{WC}. \tag{6}$$

Now insert the right and left equations from (3) into their appropriate places in (6), and then insert the resulting equations from (6), together with the middle equation of (3) into (5), to deduce that the lines BU, CV, DW lie in a pencil if and only if

$$\frac{CA}{AG} \cdot \frac{CF}{FG} \cdot \frac{GB}{BD} \cdot \frac{GD}{BH} \cdot \frac{BD}{DH} \cdot \frac{HI}{IC} \cdot \frac{HE}{EC} = 1. \tag{7}$$

Rearrange equation (7) to obtain equation (4), and the lemma is proved. ■

Proof of the Theorem. We are now ready for the proof of the 5-Circle Theorem. For the points $A, B, \ldots, I$ of Lemma 2, we choose $A_1, \ldots, A_5, B_1, \ldots, B_4$, respectively; and for U, V, W, choose C_2, C_3, C_4, where C_i is the common point—finite or infinite—of the lines g_i and $A_{i-1}A_{i+1}$ ($i = 1, \ldots, 5$). Note that $F \in BE$ (because $B_1 \in A_2A_5$) and $I \in AD$ (because $B_4 \in A_4A_1$). Thus, Menelaus's theorem applied to triangle *CGH* and the collinear triad B, E, F asserts that the expression on the left of equation (4) equals -1, while applied to the collinear triad A, D, I asserts that the expression on the right also equals -1. Lemma 1 tells us that Lemma 2 applies, whence the lines $g_2 = A_2C_2$, $g_3 = A_3C_3$, $g_4 = A_4C_4$ lie in a pencil. Similar reasoning tells us that g_1, g_2, g_3 and g_3, g_4, g_5 must lie in that same pencil, which completes the proof of the 5-Circle Theorem [**6**, Ch. 17, pp. 945–1013]. ■

Further comments

Because our proof is based on three theorems that hold in affine metric planes over fields of arbitrary characteristic, the 5-Circle Theorem holds in a more general setting. In particular, the five circles of the theorem can be replaced by ellipses that share two conjugate imaginary points at infinity, or by hyperbolas whose respective asymptotes are parallel, or by parabolas with parallel axes [**6**]. Of course, the metric used in each case depends on the family of conics serving as circles. The theorem also holds on quadric surfaces where lines are replaced by the plane sections through a designated "North Pole", while the circles are the remaining conic plane sections. The details will be published in [**4**].

Our theorem bears a striking resemblance to the 19th-century 5-circle theorem associated with Auguste Miquel. Given the ten points $A_1, \ldots, A_5, B_1, \ldots, B_5$ as in our 5-circle theorem, define five circles $\Omega_i = B_{i-1}A_iB_i$ for $i = 1, \ldots, 5$. Then the second points of intersection of Ω_i with Ω_{i+1} lie on a sixth circle. This theorem plays a crucial role in circle geometries; various aspects, relatives, and reincarnations are explored in [**1**], [**2**], and [**5**]. As far as we can determine, the relationship of Miquel's theorem to our theorem is entirely superficial. For one thing, the classical theorem deals with five lines and six circles, while ours deals with 10 lines and five circles; furthermore, the sides of the triangles $B_{i-1}A_iB_i$ inscribed in Miquel's circles are lines of the configuration. In addition, the techniques used in proofs of Miquel's theorem do not help at all in proving our theorem.

On the other hand, our theorem, like the Miquel theorem, seems to have an extension to configurations involving n circles. Computer experiments for $n = 6$ and 7 suggest that if we start with n points $A_1, A_2, \ldots, A_n$, no three of which are collinear, and situated so that A_iA_{i+2} is not parallel to $A_{i-1}A_{i+1}$ for any i (where indices are calculated modulo n), and we define n further points $B_1, B_2, \ldots, B_n$ by $B_i = A_iA_{i+2} \cap A_{i-1}A_{i+1}$, then these $2n$ distinct points define n circles $\Gamma_i = A_iB_iA_{i+1}$. As before, we denote by g_i the radical axis of Γ_{i-1} and Γ_i.

CONJECTURE. *If the lines $g_1, g_2, \ldots, g_{n-3}$ lie in a pencil, then the remaining three g_i lie in the same pencil.*

We have managed to prove the conjecture when $n = 6$ [**4**]. Most of our argument for $n = 5$ remains valid for larger n, but so far we are unable to extend the proof beyond $n = 6$.

Acknowledgment We thank the referee for his advice that we compare our result to Miquel's 5-circle theorem.

REFERENCES

1. J. W. Clawson, A chain of circles associated with the 5-line, *Amer. Math. Monthly* **61** (1954) 161–166. http://dx.doi.org/10.2307/2307214
2. J. W. Clawson, A chain of circles associated with the n-line, *Amer. Math. Monthly* **63** (1956) 306–315. http://dx.doi.org/10.2307/2310512
3. H. S. M. Coxeter and S. L. Greitzer, *Geometry Revisited*, Mathematical Association of America, Washington, DC, 1967.
4. J. Chris Fisher, Larry Hoehn, and Eberhard Schröder, Zwei Kreistheoreme für affin-metrische Ebenen, *Mitt. Math. Ges. Hamburg* **33** (2013) 167–180.
5. M. S. Longuet-Higgins, Inversive properties of the plane n-line, and a symmetric figure of 2×5 points on a quadric, *J. London Math. Soc.* **12** (1976) 206–212.
6. E. M. Schröder, Metric Geometry. In F. Buekenhout, *Handbook of Incidence Geometry*, Elsevier Science B.V., Amsterdam, 1995.

Summary We state and prove a surprising incidence theorem that was discovered with the help of a computer graphics program. The theorem involves sixteen points on ten lines and five circles; our proof relies on theorems of Euclid, Menelaus, and Ceva. The result bears a striking resemblance to Miquel's 5-circle theorem, but as far as we can determine, the relationship of our result to known incidence theorems is superficial.

A Simple Proof that $e^{p/q}$ Is Irrational

THOMAS BEATTY
Florida Gulf Coast University
Fort Myers, Florida 33965-6565
tbeatty@fgcu.edu

TIMOTHY W. JONES
Florida Gulf Coast University
Fort Myers, Florida 33965-6565
tjones@fgcu.edu

Hermite proved that e is transcendental in 1873 [**3**]. His proof has been improved over the years by several mathematicians. A similar evolution has not taken place for proofs that show the irrationality of rational powers of e. In this note, we use relatively recent transcendence techniques [**4, 6**] to prove that the powers of e are irrational.

This approach may have pedagogical advantages in that it allows for the understanding of recent transcendental techniques, for both e and π, in the simpler context of an irrationality proof. It also gives a nice use of the mean value theorem that is suitable for first-year calculus students.

e^p is irrational

Assume, to the contrary, that $e^p = a/b$ with a, b, and p positive integers.

Since factorial growth exceeds polynomial, we can choose a positive integer n large enough that

$$be^p p^{2n+1} < n!. \tag{1}$$

Choose a value of n satisfying (1) and define $f(x) = x^n(p-x)^n$. Define $F(x)$ as the sum of $f(x)$ and its derivatives; that is,

$$F(x) = f(x) + f'(x) + \cdots + f^{(2n)}(x).$$

Next, let $G(x) = -e^{-x}F(x)$. Then $G'(x) = e^{-x}f(x)$. Using the mean value theorem on the interval $[0, p]$, we know that there exists $\zeta \in (0, p)$ such that

$$\frac{G(p) - G(0)}{p} = G'(\zeta),$$

or

$$\frac{-e^{-p}F(p) + F(0)}{p} = e^{-\zeta}f(\zeta). \tag{2}$$

Now, multiplying both sides of (2) by pe^p gives

$$-F(p) + e^p F(0) = pe^{p-\zeta}f(\zeta),$$

and then substituting $e^p = a/b$ and multiplying by b gives

$$-bF(p) + aF(0) = bpe^{p-\zeta}f(\zeta). \tag{3}$$

Math. Mag. **87** (2014) 50–51. doi:10.4169/math.mag.87.1.50. © Mathematical Association of America

We claim that the left side of (3) is an integer multiple of $n!$. When we repeatedly differentiate $f(x)$, we find that every term of every derivative includes either a factor of x or a factor of $n!$. Similarly, each term includes either a factor of $(p - x)$ or a factor of $n!$. It follows that both $F(0)$ and $F(p)$ are integer multiples of $n!$, and so the left side of (3) is also an integer multiple of $n!$. A Leibniz table, developed in [**7**], shows these properties succinctly.

Meanwhile, the right-hand side of (3) is strictly positive, and it is at most $bp^{2n+1}e^p$. This follows as the maximum values of x^n and $(p - x)^n$ on $(0, p)$ are both p^n, so that $f(\zeta)$ is bounded above by p^{2n}. The additional p factor in $pbe^{p-\zeta} f(\zeta)$ gives the $2n + 1$ exponent. Therefore, by (1), the right side of (3) is strictly less than $n!$.

We have, then, a contradiction: An integer multiple of $n!$ is positive, but less than $n!$.

$e^{p/q}$ is irrational

To show that rational powers of e are irrational, suppose to the contrary that $e^{p/q} = a/b$, where p, q, a, and b are positive integers. Then

$$(e^{p/q})^q = e^p = (a/b)^q,$$

and, as $(a/b)^q$ is rational, this contradicts the irrationality of e^p.

Further reading

To see how the techniques used in this article can be applied, with some modifications, to show the irrationality of π, see [**7**]. Readers interested in a transcendence proof for e should give Herstein's proof a try [**4**]. After mastering the transcendence of e, we are ready to approach the big brother and big sister of all these irrationality and transcendence proofs: the transcendence of π, which shows that you can't square the circle. Hobson gives the history of attempts to square the circle from antiquity up to the proof of its impossibility [**5**]. Niven's 1939 transcendence of π proof [**8**] adds some further historical perspectives while giving a simplification of Lindemann's original 1882 proof. Original transcendence proofs for e and π can be found in [**1**].

REFERENCES

1. L. Berggren, J. Borwein, and P. Borwein, *Pi: A Source Book,* 3rd ed., Springer, New York, 2004. http://dx.doi.org/10.1007/978-1-4757-4217-6
2. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers,* 6th ed., Oxford, 2008.
3. C. Hermite, Sur la fonction exponientielle, *Compt. Rend. Acad. Sci. Paris* **77** (1873) 18–24, 74–79, 226–233, 285–293.
4. I. N. Herstein, *Topics in Algebra*, 2nd ed., John Wiley, New York, 1975.
5. E. W. Hobson, *Squaring the Circle: A History of the Problem,* Cambridge University Press, Cambridge, 1913; Reprinted by Merchant Books, New York, 2007.
6. A. Hurwitz, Beweis der Transendenz der Zahl *e*, *Mathematische Annalen* **43** (1893) 220–221. http://dx.doi.org/10.1007/BF01443646
7. T. W. Jones, Euler's identity, Leibniz tables, and the irrationality of pi, *College Math. J.* **43** (2012) 361–364. http://dx.doi.org/10.4169/college.math.j.43.5.361
8. I. Niven, The transcendence of π, *Amer. Math. Monthly* **46** (1939) 469–471. http://dx.doi.org/10.2307/2302515

Summary Using a simple application of the mean value theorem, we show that rational powers of *e* are irrational.

A Unique Area Property of the Quadratic Function

CONNIE XU
Montgomery Blair High School
Silver Spring, MD 20901-2451
xuconni@gmail.com

A problem that I encountered on a calculus test was to find the area of the region $\mathcal{R}$ bounded by the graph of $f(x) = x^2$, the x-axis, and the tangent line at the point $(1, 1)$, as shown in FIGURE 1.

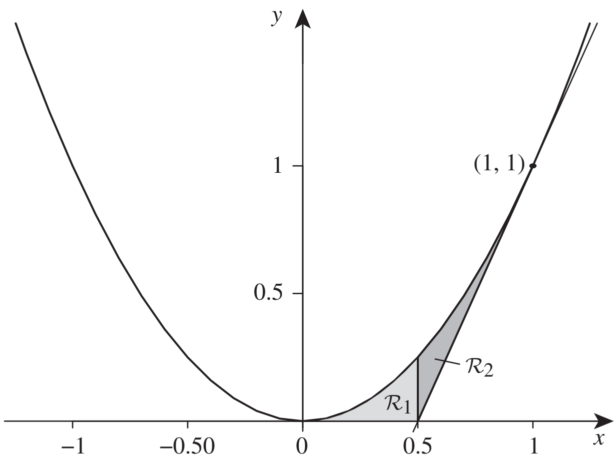


Figure 1 $y = x^2$ with tangent line at the point $(1, 1)$

The equation of the tangent line at the point $(1, 1)$ is $y = 2x - 1$, which intersects the x-axis at the point $(1/2, 0)$. To find the desired area, we draw a line segment between $(1/2, 0)$ and $(1/2, 1/4)$ to divide the region $\mathcal{R}$ into two subregions $\mathcal{R}_1$ and $\mathcal{R}_2$. Then

$$A_1 = \text{area of the subregion } \mathcal{R}_1 = \int_0^{1/2} x^2\,dx = \frac{1}{24},$$

$$A_2 = \text{area of the subregion } \mathcal{R}_2 = \int_{1/2}^{1} [x^2 - (2x - 1)]\,dx = \frac{1}{24}.$$

The desired area is $A_1 + A_2 = 1/12$.

The x-axis is itself a tangent line, so this problem is really about two tangent lines. How far does it generalize? We became curious about two questions. First, if f is any quadratic function and two regions are formed by its graph, two tangent lines, and the vertical line through the intersection of the tangent lines, do the two regions have the same area? Second, are there any other functions with this property?

Math. Mag. **87** (2014) 52–56. doi:10.4169/math.mag.87.1.52. © Mathematical Association of America

Let f be a convex function, and let x_1 and x_2 be any two numbers with $x_1 < x_2$. We use the following notation:

L_1 is the tangent line to the graph of f at $(x_1, f(x_1))$;
L_2 is the tangent line to the graph of f at $(x_2, f(x_2))$;
$\hat{x}$ is the x-coordinate of the point where L_1 and L_2 intersect;
A_1 is the area of the region $\mathcal{R}_1$ bounded by L_1, the graph of f, and the line $x = \hat{x}$;
A_2 is the area of the region $\mathcal{R}_2$ bounded by L_2, the graph of f, and the line $x = \hat{x}$.

THEOREM. *Let f be a twice-differentiable function such that $f''(x)$ is continuous and positive for all x. Then the following are equivalent:*

(a) *f is quadratic.*
(b) *$A_1 = A_2$ for every choice of x_1 and x_2 with $x_1 < x_2$.*

Proof. Let $T_i(x)$ represent the function whose graph is the tangent line L_i, that is,

$$T_i(x) = f(x_i) + f'(x_i)(x - x_i) \tag{1}$$

for $i = 1, 2$.

We first prove that (a) implies (b). Let $f(x) = ax^2 + bx + c$ with $a > 0$. Substituting $f(x_i) = ax_i^2 + bx_i + c$ and $f'(x_i) = 2ax_i + b$ into (1) yields that $T_i(x) = ax_i^2 + bx_i + c + (2ax_i + b)(x - x_i)$ for $i = 1, 2$. The solution of the equation $T_1(x) = T_2(x)$ for x is $\hat{x} = (x_1 + x_2)/2$, the x-coordinate of the intersection of L_1 and L_2. Thus,

$$\begin{aligned} A_1 &= \int_{x_1}^{\hat{x}} [f(x) - T_1(x)]\, dx \\ &= \int_{x_1}^{\hat{x}} \Big(ax^2 + bx + c - \big[ax_1^2 + bx_1 + c + (2ax_1 + b)(x - x_1)\big]\Big)\, dx \\ &= a\int_{x_1}^{\hat{x}} (x - x_1)^2\, dx = \frac{a}{3}(\hat{x} - x_1)^3 = \frac{a}{24}(x_2 - x_1)^3 \end{aligned}$$

and

$$\begin{aligned} A_2 &= \int_{\hat{x}}^{x_2} [f(x) - T_2(x)]\, dx \\ &= \int_{\hat{x}}^{x_2} \Big(ax^2 + bx + c - \big[ax_2^2 + bx_2 + c + (2ax_2 + b)(x - x_2)\big]\Big)\, dx \\ &= a\int_{\hat{x}}^{x_2} (x - x_2)^2\, dx = -\frac{a}{3}(\hat{x} - x_2)^3 = \frac{a}{24}(x_2 - x_1)^3. \end{aligned}$$

Clearly, $A_1 = A_2$. This completes the proof that (a) implies (b).

Now we prove that (b) implies (a). We begin by evaluating A_1 and A_2, this time without the assumption that f is quadratic:

$$\begin{aligned} A_1 &= \int_{x_1}^{\hat{x}} [f(x) - T_1(x)]\, dx = \int_{x_1}^{\hat{x}} [f(x) - f(x_1) - f'(x_1)(x - x_1)]\, dx \\ &= \int_{x_1}^{\hat{x}} f(x)\, dx - f(x_1)(\hat{x} - x_1) - \frac{f'(x_1)}{2}(\hat{x} - x_1)^2, \end{aligned}$$

and

$$A_2 = \int_{\hat{x}}^{x_2} [f(x) - T_2(x)]\,dx = \int_{\hat{x}}^{x_2} [f(x) - f(x_2) - f'(x_2)(x - x_2)]\,dx$$
$$= \int_{\hat{x}}^{x_2} f(x)\,dx + f(x_2)(\hat{x} - x_2) + \frac{f'(x_2)}{2}(\hat{x} - x_2)^2.$$

The condition that $A_1 = A_2$ is equivalent to $A_1 - A_2 = 0$, which means that

$$\int_{x_1}^{\hat{x}} f(x)dx - f(x_1)(\hat{x} - x_1) - \frac{f'(x_1)}{2}(\hat{x} - x_1)^2$$
$$- \int_{\hat{x}}^{x_2} f(x)dx - f(x_2)(\hat{x} - x_2) - \frac{f'(x_2)}{2}(\hat{x} - x_2)^2 = 0 \tag{2}$$

for all x_1 and x_2 with $x_1 < x_2$.

To simplify (2), we fix x_2 and differentiate the left side with respect to x_1. This gives

$$f(\hat{x})\frac{\partial \hat{x}}{\partial x_1} - f(x_1) - f'(x_1)(\hat{x} - x_1) - f(x_1)\left(\frac{\partial \hat{x}}{\partial x_1} - 1\right)$$
$$-\frac{f''(x_1)}{2}(\hat{x} - x_1)^2 - f'(x_1)(\hat{x} - x_1)\left(\frac{\partial \hat{x}}{\partial x_1} - 1\right)$$
$$+f(\hat{x})\frac{\partial \hat{x}}{\partial x_1} - f(x_2)\frac{\partial \hat{x}}{\partial x_1} - f'(x_2)(\hat{x} - x_2)\frac{\partial \hat{x}}{\partial x_1} = 0, \tag{3}$$

where $\partial\hat{x}/\partial x_1$ represents the derivative of $\hat{x}$ with respect to x_1, keeping x_2 fixed.

Now let's compute $\hat{x}$ and $\partial\hat{x}/\partial x_1$. To compute $\hat{x}$, we solve $T_1(x) = T_2(x)$, giving

$$f(x_1) + f'(x_1)(x - x_1) = f(x_2) + f'(x_2)(x - x_2),$$

whose solution for x is

$$\hat{x} = \frac{f(x_2) - f'(x_2)x_2 - f(x_1) + f'(x_1)x_1}{f'(x_1) - f'(x_2)}. \tag{4}$$

The denominator of (4) is negative because $f'(x)$ is strictly increasing in x.

Now, differentiating (4) using the quotient and product rules, we have

$$\frac{\partial \hat{x}}{\partial x_1} = \frac{\partial}{\partial x_1}\left(\frac{f(x_2) - f'(x_2)x_2 - f(x_1) + f'(x_1)x_1}{f'(x_1) - f'(x_2)}\right)$$
$$= \frac{f''(x_1)x_1[f'(x_1) - f'(x_2)] - f''(x_1)[f(x_2) - f'(x_2)x_2 - f(x_1) + f'(x_1)x_1]}{[f'(x_1) - f'(x_2)]^2}$$
$$= \frac{f''(x_1)x_1}{f'(x_1) - f'(x_2)} - \frac{f''(x_1)\,\hat{x}}{f'(x_1) - f'(x_2)}$$
$$= \frac{x_1 - \hat{x}}{f'(x_1) - f'(x_2)} f''(x_1). \tag{5}$$

Collecting like terms and using (5) to simplify, we conclude that (3) is equivalent to

$$2f(\hat{x}) - f(x_1) - f'(x_1)(\hat{x} - x_1) - f(x_2) - f'(x_2)(\hat{x} - x_2)$$
$$= \frac{f''(x_1)}{2}(\hat{x} - x_1)^2\left(\frac{\partial \hat{x}}{\partial x_1}\right)^{-1} = \frac{1}{2}(x_1 - \hat{x})[f'(x_1) - f'(x_2)]. \tag{6}$$

We have used the fact that $x_1 - \hat{x} \neq 0$ in (6) (since the two tangent lines cannot meet at x_1).

We can use the same process with x_2 that we did with x_1; that is, fixing x_1, differentiating both sides of (2) with respect to x_2, and substituting expressions for $\hat{x}$ and $\partial\hat{x}/\partial x_2$ to yield

$$2f(\hat{x}) - f(x_1) - f'(x_1)(\hat{x} - x_1) - f(x_2) - f'(x_2)(\hat{x} - x_2)$$
$$= \frac{f''(x_2)}{2}(\hat{x} - x_2)^2 \left(\frac{\partial \hat{x}}{\partial x_2}\right)^{-1} = \frac{1}{2}(\hat{x} - x_2)[f'(x_1) - f'(x_2)]. \tag{7}$$

Since the left-hand sides of (6) and (7) are equal, their right-hand sides must also be the same, i.e.,

$$\frac{1}{2}(x_1 - \hat{x})[f'(x_1) - f'(x_2)] = \frac{1}{2}(\hat{x} - x_2))[f'(x_1) - f'(x_2)],$$

which is equivalent to

$$\hat{x} = \frac{x_1 + x_2}{2}. \tag{8}$$

By (4), we see that (8) is equivalent to

$$\frac{f(x_2) - f'(x_2)x_2 - f(x_1) + f'(x_1)x_1}{f'(x_1) - f'(x_2)} = \frac{x_1 + x_2}{2},$$

or

$$f(x_2) - f(x_1) = \frac{x_2 - x_1}{2}\left[f'(x_1) + f'(x_2)\right] \tag{9}$$

for any x_1 and x_2 with $x_1 < x_2$.

Now we show that f must be a quadratic function under condition (9). For $x \in (-\infty, \infty)$, let $g(x) = f(x) - f(0) - f'(0)x$. Then $g'(x) = f'(x) - f'(0)$ is strictly increasing in x because $g''(x) = f''(x) > 0$ for all x. Using this fact with $g'(0) = 0$, we can conclude that $g(x)$ is strictly decreasing on $(-\infty, 0)$, strictly increasing on $(0, \infty)$, and attaining its minimum 0 at $x = 0$. Therefore, $g(x) > 0$ for $x \neq 0$. Furthermore, simple algebra shows that (9) can be expressed in terms of the function g below:

$$g(x_2) - g(x_1) = \frac{x_2 - x_1}{2}\left[g'(x_2) + g'(x_1)\right] \tag{10}$$

for any x_1 and x_2 with $x_1 < x_2$.

When $x > 0$, setting $x_1 = 0$ and $x_2 = x$ in (10) and using the facts that $g(0) = 0$ and $g'(0) = 0$ yield $g(x) = (x/2)g'(x)$ for all $x > 0$, which is equivalent to

$$\left(\ln[g(x)]\right)' = \left(\ln(x^2)\right)', \tag{11}$$

whose solution is $g(x) = c\,x^2$, where

$$c = \lim_{x\to 0+} \frac{g''(x)}{2} = \lim_{x\to 0+} \frac{f''(x)}{2} = \frac{f''(0)}{2}$$

because $f''(x)$ is assumed to be continuous. Thus,

$$g(x) = \frac{f''(0)}{2}x^2. \tag{12}$$

When $x < 0$, setting $x_1 = x$ and $x_2 = 0$ in (10) will obtain the same equations as (11) and (12). Furthermore, since $g(0) = 0$, (12) holds when $x = 0$. Therefore, (12) holds for all $x \in (-\infty, \infty)$. Since $g(x) = f(x) - f(0) - f'(0)x$, we have

$$f(x) = f(0) + f'(0)x + g(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2$$

for $x \in (-\infty, \infty)$; that is, f is a quadratic function of x. This completes the proof that (b) implies (a). ■

We need to mention another criterion for a function to be quadratic obtained by Stenlund [**1**], who proves the equivalence of the following two conditions:

(a) f is quadratic.

(b) For any x_1 and x_2 with $x_1 < x_2$, the tangent lines at $(x_1, f(x_1))$ and $(x_2, f(x_2))$ meet at a point with x-coordinate $\hat{x} = (x_1 + x_2)/2$.

In his proof that (b) implies (a), Stenlund assumes that f has derivatives of all orders because he uses the Taylor series of f. After obtaining (8), which is the same as his condition (b), we cannot directly apply Stenlund's result because we only assume that $f''(x) > 0$. Our method of using (8) to show f to be quadratic is simpler than Stenlund's method and does not assume the existence of $f^{(k)}(x)$ for $k \geq 3$. Finally, it is worth mentioning that our theorem is also true when a twice-differentiable function f defined on the real line is concave such that $f''(x)$ is continuous and $f''(x) < 0$ for all x.

REFERENCE

1. M. Stenlund, On the tangent lines of a parabola, *College Math. J.* **32** (2001) 194–196. http://dx.doi.org/10.2307/2687469

Summary Suppose that a function f defined on the real line is convex or concave with $f''(x)$ continuous and nonzero for all x. Let $(x_1, f(x_1))$ and $(x_2, f(x_2))$ be two arbitrary points on the graph of f with $x_1 < x_2$. For $i = 1, 2$, let L_i denote the tangent line to f at the point $(x_i, f(x_i))$ and let A_i be the area of the region $\mathcal{R}_i$ bounded by the graph of f, the tangent line L_i, and the line $x = \hat{x}$, the x-coordinate of the intersection of L_1 and L_2. It is proved that f is a quadratic function if and only if $A_1 = A_2$ for every choice of x_1 and x_2.

de Bruijn Arrays for L-Fillings

LARA PUDWELL
Valparaiso University
Valparaiso, Indiana 46383
Lara.Pudwell@valpo.edu

RACHEL ROCKEY
Valparaiso University
Valparaiso, Indiana 46383
Rachel.Rockey@valpo.edu

Consider the sequence 00010111. If we take its values three at a time, we get 000, 001, 010, 101, 011, 111, 110, and 100—all eight binary patterns of length 3. (We consider the end of the sequence to be glued to the beginning.) A sequence like this is called a *de Bruijn sequence.* More generally, we may ask for a sequence made from the k digits $\{0, 1, 2, \ldots, k-1\}$ that contains all possible patterns of length n, each exactly once; this is called a (k, n)-*de Bruijn sequence.*

de Bruijn sequences are well studied and have been used in applications ranging from robotics to developing card tricks. Diaconis and Graham [**2**] give a delightful overview of the applications. Nicolaas Govert de Bruijn [**1**]—for whom such sequences are named—and I. J. Good [**3**] independently proved that (k, n)-de Bruijn sequences exist for every $k \geq 2$ and $n \geq 2$, and there are $(k!)^{k^{n-1}}/k^n$ of them.

More recently, mathematicians have analyzed a 2-dimensional generalization. If $k \geq 2$ and $m, n \in \mathbb{Z}^+$, then there are k^{mn} rectangular patterns of size $m \times n$. For $r, s \in \mathbb{Z}^+$, a (k, r, s, m, n)-*de Bruijn torus* is an $r \times s$ array that contains each of these patterns exactly once. (Here, the left side of the array is glued to the right side, and the top of the array is glued to the bottom.) For this to occur, the parameters must satisfy $rs = k^{mn}$. Much is still unknown about general de Bruijn tori. Hurlbert and Isaak [**4**] showed that (k, r, s, m, n)-de Bruijn tori always exist when $r = s$, $m = n$, and $rs = k^{mn}$. An example of a (2, 4, 4, 2, 2)-de Bruijn torus is shown in FIGURE 1.

0	0	1	0
1	1	1	0
0	1	1	1
0	1	0	0

Figure 1 A (2, 4, 4, 2, 2)-de Bruijn torus

In this note, we consider a similar problem. Rather than compactly arranging all possible sequences or all possible rectangles, we consider an L-shape; that is, a 2×2 array with the upper right corner removed. Since the L-shape has three entries, there are k^3 fillings of the L with digits from $\{0, \ldots, k-1\}$. The $2^3 = 8$ fillings of an L with a binary alphabet are shown in FIGURE 2. We wish to find a $k \times k^2$ array (with the left side glued to the right side and the top glued to the bottom) that contains each of the k^3 different L's exactly once. To distinguish from the tori in the previous paragraph, we call such an array a *k-de Bruijn L-array*. An example of a 2-de Bruijn L-array is shown in FIGURE 3.

Math. Mag. **87** (2014) 57–60. doi:10.4169/math.mag.87.1.57.

0			0			0			0	
0	0		0	1		1	0		1	1

1			1			1			1	
0	0		0	1		1	0		1	1

Figure 2 All possible binary L fillings

0	0	1	0
0	1	1	1

Figure 3 A 2-de Bruijn L-array

It turns out that such arrays always exist. The rest of this note is aimed at constructing one.

Suppose the alphabet size is k, and consider a $k \times k^2$ array. We index the rows by $r = 0, \ldots, k-1$, and the columns by $c = 0, \ldots, k^2 - 1$. We write the column index as

$$c = sk + e$$

with $0 \le s \le k-1$ and $0 \le e \le k-1$, so that s and e are the digits of c in base k. Now, toward filling the array, we define the function f by

$$f(r, c) = s + re \mod k,$$

meaning the remainder when $s + re$ is divided by k. For example, suppose that $k = 4$ and we wish to compute $f(2, 13)$. Since $13 = 3 \cdot 4 + 1$, we have $r = 2$, $c = 13$, $s = 3$, $e = 1$, and $f(2, 13) = (3 + 2 \cdot 1) \bmod 4 = 1$.

Our main result is this theorem.

THEOREM. *Placing the value $f(r, c)$ into row r, column c in a $k \times k^2$ array produces a k-de Bruijn L-array.*

The 2-de Bruijn L-array, the 3-de Bruijn L-array, and the 4-de Bruijn L-array produced by this formula are shown in FIGURE 4.

0	0	1	1
0	1	1	0

0	0	0	1	1	1	2	2	2
0	1	2	1	2	0	2	0	1
0	2	1	1	0	2	2	1	0

0	0	0	0	1	1	1	1	2	2	2	2	3	3	3	3
0	1	2	3	1	2	3	0	2	3	0	1	3	0	1	2
0	2	0	2	1	3	1	3	2	0	2	0	3	1	3	1
0	3	2	1	1	0	3	2	2	1	0	3	3	2	1	0

Figure 4 A 2-de Bruijn L-array, a 3-de Bruijn L-array, and a 4-de Bruijn L-array

As an equivalent way to describe the entries, we index rows and columns starting with 0, but now we partition the k^2 columns of the array into k *squares*. Columns

$0, 1, \dots, k-1$ make up square 0; columns $k, \dots, 2k-1$ make up square 1; in general, columns $sk, \dots, (s+1)k-1$ make up square s. Like the row number, an entry's square number s must have $0 \le s \le k-1$. An entry in column $c = sk + e$ is actually in the eth column of square s.

As an example, consider the 3×9 array in FIGURE 5. The cell marked ♠ has coordinates $r = 0$, $e = 0$, $s = 0$; cell ♥ has $r = 0$, $e = 0$, $s = 1$; cell ◆ has $r = 1$, $e = 0$, $s = 2$; and cell ♣ has $r = 2$, $e = 2$, $s = 2$. This coordinate system uniquely identifies each of the k^3 entries in a $k \times k^2$ array with a 3-tuple in $\{0, \dots, k-1\}^3$, and the definition of $f(r, c)$ given above is equivalent to placing $(s + re)$ (reduced mod k) in row r, column e of square s.

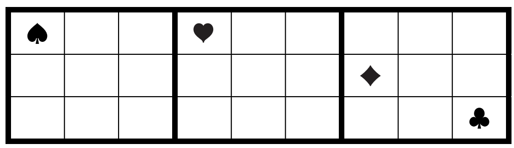


Figure 5 A 3×9 array

Proof of the theorem

Consider the $k \times k^2$ array with $f(r, c)$ in row r, column c. To prove the theorem, we need to show that any two L-fillings in different positions in the array are distinct.

So suppose that

$$\begin{array}{|c|c}\hline a_1 & \\ \hline b_1 & \multicolumn{1}{|c|}{d_1} \\ \hline\end{array} \quad \text{and} \quad \begin{array}{|c|c}\hline a_2 & \\ \hline b_2 & \multicolumn{1}{|c|}{d_2} \\ \hline\end{array}$$

are two L-fillings in different locations. Clearly if $a_1 \ne a_2$ or $b_1 \ne b_2$, then these fillings are distinct, so suppose that $a_1 = a_2 = a$ and $b_1 = b_2 = b$. We must show that $d_1 \ne d_2$. To be specific, say a_1 is in row r, column c with $c = sk + e$ and a_2 is in row R, column C with $C = Sk + E$. By construction, if b appears immediately below a in column $c = sk + e$, then $e \equiv (b - a) \pmod{k}$. Thus, $e = E$.

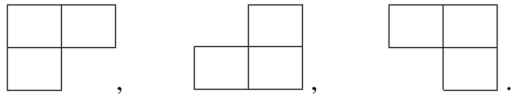
If $r = R$, we see that $(s + re) \equiv (S + Re) \pmod{k}$ and, after subtracting re from both sides, $s \equiv S \pmod{k}$. This means that both fillings have the same row and column numbers, which contradicts these two fillings being in different locations. It must be the case that $r \ne R$.

We now have two cases. If $e = k - 1$, we see by construction that $s \ne S$. (If $s = S$, then $a_1 = s + r(k-1) \bmod k$ and $a_2 = s + R(k-1) \bmod k$, which means that $s + r(k-1) \bmod k = s + R(k-1) \bmod k$. This implies $r \equiv R \pmod{k}$, which contradicts the fact that $r \ne R$.) Hence, $d_1 = (s+1) \bmod k$ and $d_2 = (S+1) \bmod k$ cannot be equal and the fillings are distinct.

So suppose that $e \ne k - 1$. Expanding $d_1 = (s + (r+1)(e+1)) \bmod k$ and substituting $a = (s + re) \bmod k$ shows $d_1 = (a + e + r + 1) \bmod k$. We similarly see that $d_2 = (a + e + R + 1) \bmod k$. Because $r \ne R$, it must be that $d_1 \ne d_2$, and so again we have distinct fillings.

For further exploration

A similar computation shows that the array with $f(r, c)$ in row r and column c is a k-de Bruijn L-array for each of the other three orientations of the L, namely:



Although we have constructed a k-de Bruijn L-array for every $k \geq 2$, many interesting questions remain for de Bruijn L-arrays.

For example, when $k = 2$, a case analysis shows that (up to rotation) there are precisely two 2-de Bruijn L-arrays; one is shown in FIGURE 3 and the other is shown in FIGURE 4. A computer search for other L-arrays when $k = 3$ yields dozens more solutions. Some have evident symmetry while others do not; two examples are given in FIGURE 6. A computer search for other L-arrays when $k \geq 4$ is prohibitive, so less naive approaches are needed. In particular, for $k > 2$ it is unknown how many k-de Bruijn L-arrays exist.

0	0	0	1	1	1	2	2	2
1	0	0	2	1	1	0	2	2
1	2	1	2	0	2	0	1	0

0	1	1	1	0	1	2	2	1
0	0	1	1	2	1	0	0	2
0	2	0	0	2	2	1	2	2

Figure 6 Two 3-de Bruijn L-arrays

Other two-dimensional shapes merit investigation. We could explore de Bruijn arrays for fillings of staircase shapes or for fillings of the union of a longer column with a longer row. The construction of this paper is particular to the 3-square L discussed above, though, and it is not easily generalized.

Acknowledgment This project was partially supported by the National Science Foundation grant DUE-1068346.

REFERENCES

1. Nicolaas Govert de Bruijn, A combinatorial problem, *Koninklijke Nederlandse Akademie v. Wetenschappen* **49** (1946) 758–764.
2. Persi Diaconis and Ron Graham, *Magical Mathematics: The Mathematics that Animate Great Magic Tricks*, Princeton University Press, Princeton, NJ, 2011.
3. I. J. Good, Normal recurring decimals, *Journal of the London Mathematical Society* **21** (1946) 167–169.
4. Glenn Hurlbert and Garth Isaak, On the de Bruijn torus problem, *Journal of Combinatorial Theory Series A* **64** (1993) 50–62. http://dx.doi.org/10.1016/0097-3165(93)90087-O

Summary We use modular arithmetic to construct a de Bruijn L-array, which is a $k \times k^2$ array consisting of exactly one copy of each L-shaped pattern (a 2×2 array with the upper right corner removed) with digits chosen from $\{0, \ldots, k-1\}$.

PROBLEMS

BERNARDO M. ÁBREGO, *Editor*
California State University, Northridge

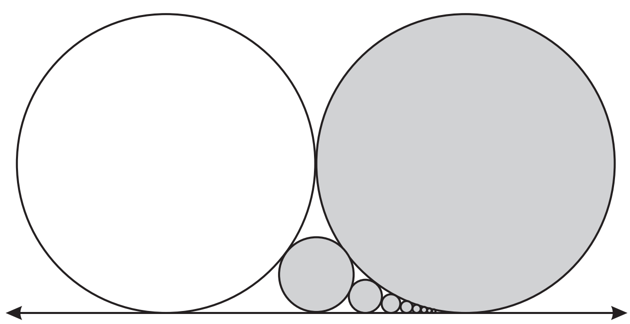
Assistant Editors: SILVIA FERNÁNDEZ-MERCHANT, California State University, Northridge; JOSÉ A. GÓMEZ, Facultad de Ciencias, UNAM, México; EUGEN J. IONASCU, Columbus State University; ROGELIO VALDEZ, Facultad de Ciencias, UAEM, México; WILLIAM WATKINS, California State University, Northridge

PROPOSALS

To be considered for publication, solutions should be received by July 1, 2014.

1936. *Proposed by Allen Schwenk, Western Michigan University, Kalamazoo, MI.*

Two mutually tangent circles of radius 1 lie on a common tangent line as shown. The circle on the left is colored white and the circle on the right is colored gray. A third, smaller circle, is tangent to both of the larger circles and the line, and it is also colored gray. An infinite sequence of gray circles all tangent to the line are inserted as follows: Each subsequent circle is tangent to the preceding circle, and it is also tangent to the largest gray circle. What is the total area bounded by the gray circles?



1937. *Proposed by Herb Bailey, Rose-Hulman Institute of Technology, Terre Haute, IN, Underwood Dudley, Tallahassee, FL, and W. C. Gosnell, Amherst, MA.*

Let P and r, respectively, denote the perimeter and the inradius of a triangle. The value $P - r$ is a square for the right triangle with sides 20, 21, and 29. The value $P + r$ is a square for the right triangle with sides 51, 140, and 149. Show that there

Math. Mag. **87** (2014) 61–68. doi:10.4169/math.mag.87.1.61.

We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals must, in general, be accompanied by solutions and by any bibliographical information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution. Submitted problems should not be under consideration for publication elsewhere.

Solutions should be written in a style appropriate for this MAGAZINE.

Solutions and new proposals should be mailed to Bernardo M. Ábrego, Problems Editor, Department of Mathematics, California State University, Northridge, 18111 Nordhoff St, Northridge, CA 91330-8313, or mailed electronically (ideally as a LaTeX or pdf file) to mathmagproblems@csun.edu. All communications, written or electronic, should include **on each page** the reader's name, full address, and an e-mail address and/or FAX number.

are no primitive integer right triangles such that both $P - r$ and $P + r$ are squares. (A right triangle with integer sides is primitive if the greatest common divisor of the side lengths is one.)

1938. *Proposed by G. W. Indika Shameera Amarasinghe, The Open University of Sri Lanka (OUSL), Nawala, Sri Lanka.*

Let ABC be a triangle with $a = BC$, $b = AC$, and $c = AB$. Suppose that there is a point D on $\overline{BC}$ such that $AD = a$. Prove that

$$\frac{5a^2}{bc} + \frac{b^3}{a^2c} + \frac{c^3}{a^2b} \geq \frac{13}{2}.$$

1939. *Proposed by Marcel Chirita, Bucarest, Romania.*

Find all the functions $f : \mathbb{R} \to \mathbb{R}$ such that $f(0) = 1$, f is continuous at $x = 0$, and there is a fixed number a, $0 < a < 1$, such that

$$3f(x) - 5f(ax) + 2f(a^2x) = x^2 + x,$$

for all real numbers x.

1940. *Proposed by Michael W. Botsko, Saint Vincent College, Latrobe, PA.*

(a) Let $(X, \|\cdot\|)$ be a real finite dimensional normed linear space and let $B = \{x \in X : \|x\| \leq 1\}$. Let $T : B \to X$ be a continuous function such that for each x, a unit vector in B, there is no $k > 1$ such that $T(x) = kx$. Is it necessary that T have a fixed point in B?

(b) If $(X, \|\cdot\|)$ is not finite dimensional, is it necessary that T have a fixed point in B?

Quickies

Answers to the Quickies are on page 68.

Q1037. *Proposed by Mowaffaq Hajja, Yarmouk University, Irbid, Jordan.*

Let ABC be a triangle with $a = BC$, $b = AC$, $c = AB$, and P a point inside $\triangle ABC$. Let AA', BB', and CC' be the cevians through P. Prove that if

$$AB' + AC' = \frac{b+c}{2}, \quad BC' + BA' = \frac{c+a}{2}, \quad \text{and} \quad CA' + CB' = \frac{a+b}{2},$$

then A', B', and C' are the midpoints of the sides $\overline{BC}$, $\overline{CA}$, and $\overline{AB}$, respectively.

Q1038. *Proposed by H. A. ShahAli, Tehran, Iran.*

Let n be a positive integer and let $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ be positive real numbers such that $a_1 + a_2 + \cdots + a_n \geq a_1b_1 + a_2b_2 + \cdots + a_nb_n$. Prove that

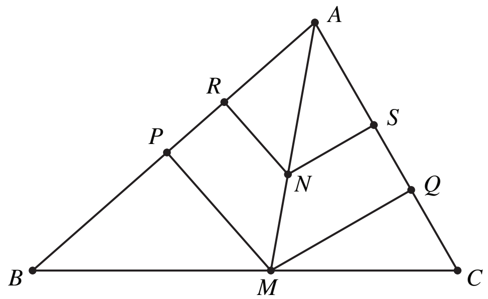
$$a_1 + a_2 + \cdots + a_n \leq \frac{a_1}{b_1} + \frac{a_2}{b_2} + \cdots + \frac{a_n}{b_n}.$$

Solutions

Implications of same area trapezoids **February 2013**

1911. *Proposed by Sadi Abu-Saymeh and Mowaffaq Hajja, Mathematics Department, Yarmouk University, Irbid, Jordan.*

Let ABC be a triangle and M a point on $\overline{BC}$ such that both angles $\angle BAM$ and $\angle CAM$ are acute. Let N be a point on $\overline{AM}$ and let P and R (respectively Q and S) be the feet of the perpendiculars from M and N onto $\overline{AB}$ (respectively $\overline{AC}$). Prove that if Area(BMP) = Area(MCQ) and Area($MNRP$) = Area($NMQS$), then M is the midpoint of $\overline{BC}$. Moreover, if $AB \neq AC$, then $\angle BAC = 90°$.



Solution by Robert L. Doucette, McNeese State University, Lake Charles, LA.
If h is the length of the hypotenuse and θ is the measure of one of the acute angles of a right triangle, the area of the triangle equals $\frac{1}{2}(h\cos\theta)(h\sin\theta) = \frac{1}{4}h^2\sin(2\theta)$. Let $\alpha = \angle BAM$ and let $\beta = \angle MAC$. Note that

$$\text{Area}(MNRP) = \text{Area}(MAP) - \text{Area}(NAR) = \frac{1}{4}(AM^2 - AN^2)\sin(2\alpha).$$

Similarly, Area($NMQS$) $= \frac{1}{4}(AM^2 - AN^2)\sin(2\beta)$. It follows that $\sin(2\alpha) = \sin(2\beta)$. Since both α and β are acute angles, either $\alpha = \beta$ or $\alpha + \beta = 90°$.

If $\alpha = \beta$, then $\triangle MPA \cong \triangle MQA$, and since Area($BMP$) = Area($MCQ$), it follows that Area($MBA$) = Area($MCA$). Thus,

$$\frac{1}{2}AB \cdot AM \cdot \sin\alpha = \text{Area}(MBA) = \text{Area}(MCA) = \frac{1}{2}AC \cdot AM \cdot \sin\beta.$$

Therefore, $AB = AC$, and so $\triangle MBA \cong \triangle MCA$ and $MB = MC$.

If $\alpha \neq \beta$, then $\alpha + \beta = 90°$. In this case, $\triangle PBM$ and $\triangle QMC$ are each similar to $\triangle ABC$. Since similar triangles with equal areas are congruent, it follows that $\triangle PBM$ and $\triangle QMC$ are congruent, and so once again $MB = MC$.

We have shown that M is the midpoint of BC and that $AB \neq AC$ implies that $\angle BAC = \alpha + \beta = 90°$.

Also solved by George Apostolopoulos (Greece), Michel Bataille (France), Bruce S. Burdick, Robert Calcaterra, John Chase, Timothy V. Craine, Chip Curtis, Michelle Daher (Lebanon), Prithwijit De (India), John Fitch, Dmitry Fleischman, Shohruh Ibragimov (Uzbekistan), Omran Kouba (Syria), Victor Kutsenok, Elias Lampakis (Greece), Kee-Wai Lau (China), Graham Lord, Peter Nüesch (Switzerland), Victor Pambuccian, C. G. Petalas (Greece), Ángel Plaza (Spain) and Francisco Perdomo (Spain), Kevin A. Roper, Joel Schlosberg, Achilleas Sinefakopoulos (Greece), Vasile G. Teodorovici (Canada), and the proposers. There was one incorrect submission.

A sequence of null integrals **February 2013**

1912. *Proposed by George Apostolopoulos, Messolonghi, Greece.*

Let $n \geq 3$ be an integer. Evaluate

$$\int_0^\infty \frac{x^n - 2x + 1}{x^{2n} - 1} dx.$$

Solution by CMC 328, Carleton College, Northfield, MN.
Note that despite the denominator, the integral is not actually improper at $x = 1$, because the numerator also has a factor $x - 1$. Let I be the value of the integral; we will show that $I = 0$ regardless of n.

To begin, we make the substitution $x = u^{-1}$. This yields

$$\begin{aligned} I &= \int_0^\infty \frac{x^n - 2x + 1}{x^{2n} - 1} dx = -\int_\infty^0 \frac{u^{-n} - 2u^{-1} + 1}{u^{-2n} - 1} u^{-2} du \\ &= \int_0^\infty \frac{u^{n-2} - 2u^{2n-3} + u^{2n-2}}{1 - u^{2n}} du = \int_0^\infty \frac{-x^{n-2} + 2x^{2n-3} - x^{2n-2}}{x^{2n} - 1} dx. \end{aligned}$$

Therefore,

$$\begin{aligned} 2I &= \int_0^\infty \frac{x^n - 2x + 1}{x^{2n} - 1} dx + \int_0^\infty \frac{-x^{n-2} + 2x^{2n-3} - x^{2n-2}}{x^{2n} - 1} dx \\ &= \int_0^\infty \frac{x^n - 2x + 1 - x^{n-2} + 2x^{2n-3} - x^{2n-2}}{x^{2n} - 1} dx \\ &= \int_0^\infty \frac{(x^n + 1)(1 - x^{n-2})}{x^{2n} - 1} + \frac{-2x + 2x^{2n-3}}{x^{2n} - 1} dx \\ &= \int_0^\infty \frac{1 - x^{n-2}}{x^n - 1} dx + \int_0^\infty \frac{-2x + 2x^{2n-3}}{x^{2n} - 1} dx, \end{aligned}$$

where neither integral is improper at $x = 1$. However, if we make the substitution $u = x^2$ in the second integral on the right, it becomes

$$\int_0^\infty \frac{-1 + u^{n-2}}{u^n - 1} du,$$

which clearly cancels the first integral on the right. Therefore, $I = 0$.

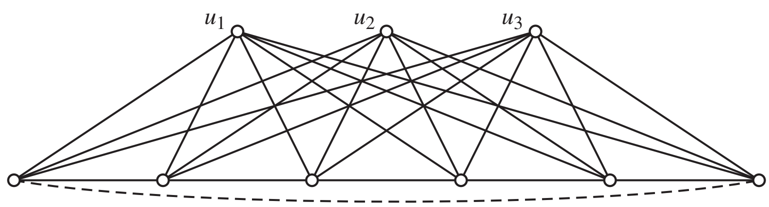
Editor's Note. Our solvers found many different ways to evaluate the integral. Some of these were by using partial fractions, series expansions, and contour integrals. Several readers pointed out that the integral is also zero for $n = 2$. Finbarr Holland noted that in fact the integral is zero for any real value $n > 1$.

Also solved by Arkady Alt, Michel Bataille (France), Bruce S. Burdick, Robert Calcaterra, Eugene S. Eyeson, John Fitch, Ovidiu Furdui (Romania), J. A. Grzesik, GWstat Problem Solving Group, Eugene A. Herman, Finbarr Holland (Ireland), Omran Kouba (Syria), Elias Lampakis (Greece), Kee-Wai Lau (China), Northwestern University Math Problem Solving Group, Moubinool Omarjee (France), Paolo Perfetti (Italy), Tomas Persson (Sweden) and Mikael P. Sundqvist (Sweden), Ángel Plaza (Spain), Ioannis M. Roussos, Iason Rusodimos, Daniel Silver and Frank Jellett, J. G. Simmonds, Achilleas Sinefakopoulos (Greece), Nicholas C. Singer, Bob Tomper, Nora Thornber, Traian Viteam (Chile), Michael Vowe (Switzerland), Haohao Wang and Yanping Xia, Connie Xu, and the proposer.

Suspenseful colorings **February 2013**

1913. *Proposed by Stan Wagon, Macalester College, St. Paul, MN.*

Let m and n be positive integers. The *m-suspension* of a graph G is the graph obtained by adding to G a set of m vertices $\{u_1, u_2, \ldots, u_m\}$ and all edges of the form vu_i, with $1 \le i \le m$ and v a vertex of G. The *fan graph $F_{m,n}$* is the m-suspension of an n-vertex path, while the *cone graph $C_{m,n}$* is the m-suspension of an n-cycle. The figure shows $F_{3,6}$ and, with the dashed edge, $C_{3,6}$. Let Δ denote a graph's maximum vertex degree.



(a) Determine all pairs (m, n) for which $F_{m,n}$ has a proper edge coloring using Δ colors (i.e., a coloring of the edges so that edges that share a vertex have different colors).

(b) Determine all pairs (m, n) for which $C_{m,n}$ has a proper edge coloring using Δ colors.

Solution by Joel Iiams, University of North Dakota, Grand Forks, ND.
We show that the only ordered pair for which $F_{m,n}$ does not have a proper edge coloring using Δ colors is $(1, 2)$. Meanwhile for $3 \le n$, $C_{m,n}$ has a proper edge coloring using Δ colors if and only if $n \ne m + 1$.

Denote the colors to be used by $\{0, 1, 2, \ldots, \Delta - 1\}$. Label the upper vertices $u_1, u_2, \ldots, u_m$ from left to right. Label the lower vertices by $v_1, v_2, \ldots, v_n$ from left to right. Use the notation $C(vu) = k$ to mean that the edge connecting the vertices v and u receives color k. Finally, let $\pi = (1, 2, \ldots, m)$ denote the cyclic permutation.

When $n \le 2$, technically there is no cone graph, only a fan graph. When $n = 1$, $F_{m,1}$ has $\Delta = m$ and we can use $C(v_1u_i) = i - 1$. When $n = 2$ and $m = 1$, $F_{1,2}$ is a 3-cycle, which cannot be properly edge-colored with $\Delta = 2$ colors. However, for $m \ge 2$, $\Delta = m + 1$ and the assignment $C(v_1v_2) = 0$, $C(v_iu_j) = \pi^{i-1}(j)$ gives a proper edge coloring.

Henceforth we take $3 \le n$, and observe that if $C_{m,n}$ has a proper Δ-coloring, then so does $F_{m,n}$.

When $n \ge m + 2$, $C_{m,n}$ has $\Delta = n$. Any proper edge coloring of $C_{n-2,n}$ will naturally give rise to a proper edge coloring of $C_{m,n}$ with $m < n - 2$, so we will take $m = n - 2$. We color as follows: $C(v_iv_{i+1}) = i$ for $i = 1, 2, \ldots, n - 1$, $C(v_nv_1) = 0$, and $C(v_iv_j) = i + j \pmod n$. The additive cancellation property for the integers gives that no two edges adjacent to an upper vertex are colored the same. Meanwhile, the colors of the edges adjacent to v_i are the residues modulo n of the consecutive integers $i - 1, i, i + 1, \ldots, i + m$. These colors are distinct by virtue of $m = n - 2$.

Next, consider $C_{m,n}$ with $3 \le n \le m$. In this case, $\Delta = m + 2$. When n is even, we make the assignments $C(v_{2k-1}v_{2k}) = 0$, $C(v_{2k}v_{2k+1}) = m + 1$, reading subscripts modulo n. Also set $C(v_iu_j) = \pi^{i-1}(j)$. The edges v_ku_j and v_iv_j receive the same color only when $\pi^{ik}(j) = j$, which implies that $m = \text{ord}(\pi)$ divides ik. This implies that $i = k$, since $1 \le i, k \le n \le m$. For n odd, make the assignments as n even, except we change $C(v_1u_1)$ from 1 to $m + 1$, $C(v_nu_{m+2-n})$ from 1 to 0, and $C(v_nv_1)$ from $m + 1$ to 1. The reassignments to v_1u_1 and v_nu_{m+2-n} do not affect the properness of the coloring for edges incident to u_1 and u_{m+2-n}. The swaps are made only to allow a free color for $v_n v_1$.

When $3 \le n = m+1$, $C_{m,n}$ has $\frac{1}{2}[m(m+1)+(m+1)(m+2)] = m(m+2)+1$ edges. Meanwhile, $\Delta = m+2$. By the Pigeonhole Principle, some color must be used $m+1$ times. But each time a color is used, it requires 2 distinct vertices. So our graph would need to have $2(m+1)$ vertices, a contradiction. Therefore, there is not a proper Δ edge coloring for $C_{m,m+1}$. On the other hand, it is easy to check that $F_{m,m+1}$ has a proper edge coloring using $C_{m,n}$ colors via the assignment $C(v_i v_{i+1}) = i$, $i = 1, 2, \ldots, m$, and $C(v_i u_j) = i + j \pmod{m+2}$.

Also solved by Robert Calcaterra, Con Amore Problem Group (Denmark), and the proposer. There was one incorrect submission.

An inequality of concave functions **February 2013**

1914. *Proposed by Michael W. Botsko, Saint Vincent College, Latrobe, PA.*

Let $h : \mathbb{R} \to \mathbb{R}$ be a differentiable function with $h(0) = h'(0) = 0$, and suppose that h' is increasing on $\mathbb{R}$. Prove that if $f : [0,1] \to \mathbb{R}$ is Riemann integrable, then

$$\int_0^1 [h(x)f(x) - xh(f(x))]dx \le \int_0^1 xh(x)dx.$$

Solution by John Zacharias, Arlington, VA.
Note that h' increasing and $h'(0) = 0$ implies that $h' > 0$ on $(0,\infty)$ and $h' < 0$ on $(-\infty, 0)$. Thus, $h(x) > h(0) = 0$ for all $x \in \mathbb{R} \setminus \{0\}$. Fix $x \in (0,1)$ and $y > x$. Since h is concave up and $h(0) = 0$, we have that $h(x)/x \le h(y)/y$. Furthermore, the inequality continues to be true if we subtract the positive quantity $h(x)/y$ from the left-hand side, thereby making it smaller. That is,

$$\frac{h(x)}{x} - \frac{h(x)}{y} \le \frac{h(y)}{y}.$$

Multiplying throughout by the positive quantity xy gives

$$h(x)(y-x) \le xh(y). \tag{1}$$

Observe that for $y < x$, the left-hand side of (1) is negative and the right-hand side is positive, so (1) holds for all $y \in \mathbb{R}$; not merely for $y > x$. By setting $y = f(x)$, the inequality in (1) can be rearranged to read

$$h(x)f(x) - xh(f(x)) \le xh(x),$$

which is valid for all $x \in (0,1)$. The result now follows from integrating both sides from 0 to 1.

Also solved by George Apostolopoulos (Greece), Robert Calcaterra, Robert L. Doucette, Don L. Hancock, Eugene A. Herman, Moubinool Omarjee (France), Sanjay K. Patel (India), and the proposer.

The numerical range of a rotation **February 2013**

1915. *Proposed by Dietrich Trenkler, Department of Economics, University of Osnabrück, Osnabrück, Germany, and Götz Trenkler, Department of Statistics, University of Dortmund, Dortmund, Germany.*

Let $n = (n_1, n_2, n_3)^T \in \mathbb{R}^3$ be a unit vector and $\theta \in [0, 2\pi]$ a real number. In $\mathbb{R}^3$, a *proper rotation* by the angle θ about an axis in the direction of n can be carried out by the matrix

$$R = (\cos\theta)I + (1-\cos\theta)nn^T + (\sin\theta)N,$$

where I denotes the identity matrix in $\mathbb{R}^3$ and

$$N = \begin{pmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{pmatrix}.$$

Determine the set

$$\mathfrak{F}(R) = \left\{z^* R z : z \in \mathbb{C}^3 \text{ and } z^* z = 1\right\},$$

where $z^* = \overline{z}^T$ denotes the conjugate transpose of z.

Solution by Omran Kouba, Higher Institute for Applied Sciences and Technology, Damascus, Syria.

We will show that $\mathfrak{F}(R)$ is the triangle of vertices 1, $e^{i\theta}$, and $e^{-i\theta}$ in $\mathbb{C}$.

Indeed, consider an arbitrary unit vector $p \in \mathbb{R}^3$ such that $p^T n = 0$, i.e., a unit vector p orthogonal to n, and let $q = n \times p$ (the cross product of n and p in this order). Now, (n, p, q) is a direct orthonormal basis of the canonical scalar product space $\mathbb{R}^3$ equipped with the usual scalar product. It is straightforward to check that, for $x \in \mathbb{R}^3$, we have

$$Rx = (\cos\theta)x + (1 - \cos\theta)(n^T x)n + (\sin\theta)(n \times x).$$

Thus, $Rn = n$, $Rp = (\cos\theta)p + (\sin\theta)q$, and $Rq = (-\sin\theta)p + (\cos\theta)q$.

Now, let us consider the canonical scalar product space $\mathbb{C}^3$. Let $u = (1/\sqrt{2})(p - iq)$ and $v = \overline{u} = (1/\sqrt{2})(p + iq)$. Since (n, p, q) is an orthonormal basis, it follows that $nn^* = 1$, $nu^* = nv^* = 0$,

$$uu^* = vv^* = \frac{1}{2}(p - iq)(p^T + iq^T) = \frac{1}{2}(pp^T + qq^T + i(pq^T - qp^T)) = 1,$$

and

$$uv^* = \frac{1}{2}(p - iq)(p^T - iq^T) = \frac{1}{2}(pp^T - qq^T - i(qp^T + pq^T)) = 0.$$

Thus, (n, u, v) is an orthonormal basis of $\mathbb{C}^3$. Moreover, $Rn = n$, $Ru = (1/\sqrt{2})(Rp - iRq) = e^{i\theta}u$, and $Rv = (1/\sqrt{2})(Rp + iRq) = e^{-i\theta}v$. Because (n, u, v) is a basis, it follows that any $z \in \mathbb{C}^3$ such that $z^*z = 1$ can be written as $z = \alpha n + \beta u + \gamma v$, for some α, β, and γ in $\mathbb{C}$. Moreover, $1 = z^*z = |\alpha|^2 + |\beta|^2 + |\gamma|^2$ and

$$\begin{aligned} z^* R z &= z^* R(\alpha n + \beta u + \gamma v) = z^*(\alpha Rn + \beta Ru + \gamma Rv) \\ &= (\alpha^* n^* + \beta^* u^* + \gamma^* v^*)(\alpha n + \beta e^{i\theta} u + \gamma e^{-i\theta} v) \\ &= |\alpha|^2 + |\beta|^2 e^{i\theta} + |\gamma|^2 e^{-i\theta}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathfrak{F}(R) &= \{|\alpha|^2 + |\beta|^2 e^{i\theta} + |\gamma|^2 e^{-i\theta} : |\alpha|^2 + |\beta^2| + |\gamma^2| = 1\} \\ &= \{r + s e^{i\theta} + t e^{-i\theta} : r, s, t \geq 0, \text{ and } r + s + t = 1\}. \end{aligned}$$

That is, $\mathfrak{F}(R)$ is the convex hull in $\mathbb{C}$ of the set consisting of the points 1, $e^{i\theta}$, and $e^{-i\theta}$.

Remark. More generally, let A be an $m \times m$ complex square matrix, and suppose that A is normal, i.e., $A^*A = AA^*$. Then $\mathfrak{F}(A) = \{z^*Az : z^*z = 1\} = \text{Convex Hull}(\sigma(A))$, that is, $\mathfrak{F}(A)$ is the smallest convex set that contains all the eigenvalues of A. The proof follows from the fact that a normal matrix A has an orthonormal basis of eigenvectors.

Also solved by Michel Bataille (France), Robert Calcaterra, Peter McPolin (Northern Ireland), and the proposers. There were two incorrect submissions.

Answers

Solutions to the Quickies from page 62.

A1037. Adding the first two equations and subtracting the third gives

$$AB' + AC' + BC' + BA' - CA' - CB' = c = AC' + BC'.$$

Thus, $AB' + BA' = CA' + CB'$, and by symmetry $BC' + CB' = AB' + AC'$. If A' is not the midpoint of $\overline{BC}$, then we may assume that $BA' < CA'$. It follows that $CB' < AB'$ and $AC' < BC'$. Therefore, $BA' \cdot CB' \cdot AC' < CA' \cdot AB' \cdot BC'$. In view of Ceva's theorem, this contradicts the assumption that AA', BB', and CC' are concurrent.

A1038. Adding the nonnegative terms $a_j(\sqrt{b_j} - 1/\sqrt{b_j})^2 = a_j b_j - 2a_j + a_j/b_j$ for $1 \le j \le n$ gives

$$a_1b_1 + a_2b_2 + \cdots + a_nb_n - 2(a_1 + a_2 + \cdots + a_n) + \frac{a_1}{b_1} + \frac{a_2}{b_2} + \cdots + \frac{a_n}{b_n} \ge 0.$$

Adding the given inequality $a_1 + a_2 + \cdots + a_n \ge a_1b_1 + a_2b_2 + \cdots + a_nb_n$ gives the desired result.

Escape the Square

Suppose you stand in the center of the $[-1, 1] \times [-1, 1]$ square and walk off in a random direction. What is the average, or expected, distance you will walk until you hit the edge of the square? By using the fact that

$$\int \sec(x)\, dx = \log_e |\sec(x) + \tan(x)|,$$

we find that the average distance is

$$\frac{2}{\pi} \int_{-\pi/4}^{\pi/4} \sec(x)\, dx = \frac{4}{\pi} \log_e(\sqrt{2} + 1) \approx 1.1222.$$

It is interesting that from a seemingly simple problem, we gets three of the most popular irrational numbers: π, e, and $\sqrt{2}$.

Another interesting fact is that if we choose to walk to a random point on the square's edge, the average distance is $\frac{1}{\sqrt{2}} + \frac{1}{2}\log_e(\sqrt{2} + 1) \approx 1.1478$, so the constant $\log_e(\sqrt{2} + 1)$ shows up in both averages.

Mark Dalthorp (Age 14)
Zion Lutheran School
Corvallis, OR

Math. Mag. **87** (2014) 68. doi:10.4169/math.mag.87.1.68.

REVIEWS

PAUL J. CAMPBELL, *Editor*
Beloit College

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

Baker, Nicholson, Wrong answer: The case against Algebra II, *Harper's Magazine* (September 2013).

Loveless, The algebra imperative: Assessing algebra in a national and international context, http://www.brookings.edu/research/papers/2013/09/04-algebra-imperative-education-loveless. Summary at http://www.brookings.edu/blogs/brown-center-chalkboard/posts/2013/09/04-algebra-coursetaking-loveless.

If you are like me, at this year's holiday parties, friends or relatives—especially those opposed to "government regulation"—asked your opinion about the Common Core Standards and hoped you would agree that their children should not be forced to learn algebra. So you should have an informed opinion about the Standards (http://www.corestandards.org/math), but you are likely already well versed in the arguments of the algebra war. Author Baker, a novelist, mainly quotes other sources, largely to the effect that algebra requires students "to do something they can't do." He urges that Algebra II no longer be required by colleges, and that we should create a new "one-year teaser course for ninth graders"—after which all math would be elective. Meanwhile, my college requires a single science course (it used to be three) and my colleagues in the sciences lament that students cannot add fractions, calculate percentages, etc.—i.e., can't do arithmetic. (Of course, those colleagues hope for a role from my department in arithmetic remediation.) Author Loveless gives an explanation in terms of grade inflation being combined with "course inflation": Not only grades but also course titles ("Algebra I," "honors," "advanced") are useless as a guide to what students learned (much less what they retain). He advocates creation of an "open algebra exam," a publicly-available test bank of thousands of questions over agreed-upon topics (and he urges consequent "teaching to the test"). He quotes a 2010 report of the National Center for Public Policy and Higher Education (NCPPHE), which has potential bearing on the fate of the Common Core Standards: "Over the last 15 years, many states have emphasized mastery of specific content and performance standards, as shown through grades and statewide assessments; however, this shift . . . generally has not been extended to higher levels of achievement associated with college readiness. . . . The flawed assumption has been that if students take the right courses and earn the right grades, they will be ready for college."

Wilson, Robin, and John J. Watkins (eds.), *Combinatorics: Ancient and Modern*, Oxford University Press, 2013; x + 381 pp, $99.95. ISBN 978-0-19-965659-2.

This book surveys the history of combinatorics through 16 essays, beginning with Donald Knuth's "Two thousand years of combinatorics." About half of the remaining essays are devoted to pre-modern combinatorics, mainly sorted by culture (Indian, Chinese, Islamic, Jewish, Renaissance), and half to modern combinatorics, mainly divided by topic (graph theory, partitions, block designs, Latin squares, and more). A final essay delves into the role and links of combinatorics in mathematics (all branches), in science (the genetic code, fundamental discreteness in physics), and society (sudoku).

Math. Mag. **87** (2014) 69–70. doi:10.4169/math.mag.87.1.69. © Mathematical Association of America

Charette, Robert N., The STEM crisis is a myth, http://spectrum.ieee.org/at-work/education/the-stem-crisis-is-a-myth.

Is there a current crisis, or a looming crisis, of too few workers in scientific/technical (STEM) fields? Author Charette examined hundreds of sources, finding conflicting opinions, data, and definitions. Charette notes who benefits from the alarm ("many people's fortunes are now tied to the STEM crisis, real or manufactured") and suggests that the current push of students into STEM fields will produce a glut and an ensuing boom/bust cycle. Accompanying discussions at the Website delve into STEM job security in engineering; comparisons of salaries over time; whether to encourage a student to pursue a STEM career; and the policy that, based on global competitiveness and income equality, "STEM-worker salaries should look a lot more like non-STEM worker salaries." Crisis or not, Charette notes that "many students aren't interested in STEM careers because they see that the academic work needed . . . is just too hard in comparison to non-STEM degrees."

Kaper, Hans, and Hans Engler, *Mathematics and Climate*, Society for Industrial and Applied Mathematics, 2013; xx + 295 pp, $59 (P). ISBN 978-1-611972-60-3.

Aimed at an audience of advanced undergraduates and master's students in mathematics, this book applies lots of mathematics in the context of the climate of the Earth. The book is not meant to be a general introduction to climate science, so students should have some prior background in that area. Chapters treat insolation, ocean circulation, dynamical systems, box models, the Lorenz equations, data analysis and statistical modeling, Fourier analysis, Kalman filters, extreme events, and more. There are exercises (no solutions given), some of which are open-ended. The reader is presumed to be proficient in calculus, linear algebra, ordinary differential equations, and statistics; and an instructor for a course using this book, in addition to being on top of all the mathematics, needs to be very knowledgeable about climate science.

Stewart, Ian, *Symmetry: A Very Short Introduction*, Oxford University Press, 2013; xii + 144 pp, $11.95 (P). ISBN 978-0-19-965198-6.

Falconer, Kenneth, *Fractals: A Very Short Introduction*, Oxford University Press, 2013; xv + 132 pp, $11.95 (P). ISBN 978-0-19-967598–2.

These two pocket-sized volumes are in a series of several hundred "Very Short Introductions" to a wide variety of topics. Author Stewart brings the reader past the usual topics (rosettes, friezes, wallpaper groups) to touch lightly on elliptic functions, permutations of roots of equations, and groups (conjugacy, normal subgroups, Rubik's cube, sand dune shapes, Lie groups and algebras, Dynkin diagrams, and simple groups). The reader needs to be comfortable with mathematical notation (more so than with other books by Stewart) but is rewarded with a broad vista of the role of groups in symmetry of all kinds. Author Falconer starts from the von Koch snowflake and uses only analytic geometry, logarithms, and complex numbers in weaving a comprehensive story about the fascination of fractals (I did not realize that the set of numbers with any given proportion of zeros in its decimal expansion is a fractal). I am also curious about the content of an older volume that I have not seen: *Mathematics: A Very Short Introduction* by Fields Medalist Timothy Gowers.

Nahin, Paul J., *Will You Be Alive 10 Years from Now? And Numerous Other Questions in Probability*, Princeton University Press, 2014; xxii + 232 pp, $27.95. ISBN 978-0-691-15680-4.

Well, the title of this book should draw in readers from the general public! In this further collection of probability puzzles, following his *Digital Dice* (2008) and *Duelling Idiots* (2002), author Nahin eschews well-known results (e.g., the birthday problem) and "theoretical" proofs (he prefers simulations in Matlab). To be included in this book, a problem had to be both "amazing" (in terms of surprise of the solution) and not well known. The problems are varied and indeed intriguing, and the solutions are delightful.

NEWS AND LETTERS

74th Annual William Lowell Putnam Mathematical Competition

Editor's Note: Additional solutions will be printed in the *Monthly* later in the year.

PROBLEMS

A1. Recall that a regular icosahedron is a convex polyhedron having 12 vertices and 20 faces; the faces are congruent equilateral triangles. On each face of a regular icosahedron is written a nonnegative integer such that the sum of all 20 integers is 39. Show that there are two faces that share a vertex and have the same integer written on them.

A2. Let S be the set of all positive integers that are *not* perfect squares. For n in S, consider choices of integers $a_1, a_2, \ldots, a_r$ such that $n < a_1 < a_2 < \cdots < a_r$ and $n \cdot a_1 \cdot a_2 \cdots a_r$ is a perfect square, and let $f(n)$ be the minimum of a_r over all such choices. For example, $2 \cdot 3 \cdot 6$ is a perfect square, while $2 \cdot 3,\ 2 \cdot 4,\ 2 \cdot 5,\ 2 \cdot 3 \cdot 4,\ 2 \cdot 3 \cdot 5,\ 2 \cdot 4 \cdot 5$, and $2 \cdot 3 \cdot 4 \cdot 5$ are not, and so $f(2) = 6$. Show that the function f from S to the integers is one-to-one.

A3. Suppose that the real numbers $a_0, a_1, \ldots, a_n$ and x, with $0 < x < 1$, satisfy

$$\frac{a_0}{1-x} + \frac{a_1}{1-x^2} + \cdots + \frac{a_n}{1-x^{n+1}} = 0.$$

Prove that there exists a real number y with $0 < y < 1$ such that

$$a_0 + a_1 y + \cdots + a_n y^n = 0.$$

A4. A finite collection of digits 0 and 1 is written around a circle. An *arc* of length $L \geq 0$ consists of L consecutive digits around the circle. For each arc w, let $Z(w)$ and $N(w)$ denote the number of 0's in w and the number of 1's in w, respectively. Assume that $|Z(w) - Z(w')| \leq 1$ for any two arcs w, w' of the same length. Suppose that some arcs $w_1, \ldots, w_k$ have the property that

$$Z = \frac{1}{k}\sum_{j=1}^{k} Z(w_j) \quad \text{and} \quad N = \frac{1}{k}\sum_{j=1}^{k} N(w_j)$$

are both integers. Prove that there exists an arc w with $Z(w) = Z$ and $N(w) = N$.

A5. For $m \geq 3$, a list of $\binom{m}{3}$ real numbers a_{ijk} $(1 \leq i < j < k \leq m)$ is said to be *area definite for* $\mathbb{R}^n$ if the inequality

$$\sum_{1 \leq i < j < k \leq m} a_{ijk} \cdot \text{Area}(\triangle A_i A_j A_k) \geq 0$$

holds for every choice of m points $A_1, \ldots, A_m$ in $\mathbb{R}^n$. For example, the list of four numbers $a_{123} = a_{124} = a_{134} = 1$, $a_{234} = -1$ is area definite for $\mathbb{R}^2$. Prove that if a list of $\binom{m}{3}$ numbers is area definite for $\mathbb{R}^2$, then it is area definite for $\mathbb{R}^3$.

Math. Mag. **87** (2014) 71–78. doi:10.4169/math.mag.87.1.71. © Mathematical Association of America

A6. Define a function $w : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ as follows. For $|a|, |b| \le 2$, let $w(a, b)$ be as in the table shown; otherwise, let $w(a, b) = 0$.

$w(a,b)$		$b=-2$	-1	0	1	2
a	-2	-1	-2	2	-2	-1
	-1	-2	4	-4	4	-2
	0	2	-4	12	-4	2
	1	-2	4	-4	4	-2
	2	-1	-2	2	-2	-1

For every finite subset S of $\mathbb{Z} \times \mathbb{Z}$, define

$$A(S) = \sum_{(\mathbf{s},\mathbf{s}')\in S\times S} w(\mathbf{s} - \mathbf{s}').$$

Prove that if S is any finite nonempty subset of $\mathbb{Z} \times \mathbb{Z}$, then $A(S) > 0$. (For example, if $S = \{(0, 1), (0, 2), (2, 0), (3, 1)\}$, then the terms in $A(S)$ are 12, 12, 12, 12, 4, 4, 0, 0, 0, 0, -1, -1, -2, -2, -4, -4.)

B1. For positive integers n, let the numbers $c(n)$ be determined by the rules $c(1) = 1$, $c(2n) = c(n)$, and $c(2n + 1) = (-1)^n c(n)$. Find the value of

$$\sum_{n=1}^{2013} c(n)c(n + 2).$$

B2. Let $\mathcal{C} = \bigcup_{N=1}^{\infty} \mathcal{C}_N$, where $\mathcal{C}_N$ denotes the set of those 'cosine polynomials' of the form

$$f(x) = 1 + \sum_{n=1}^{N} a_n \cos(2\pi nx)$$

for which:

(i) $f(x) \ge 0$ for all real x, and
(ii) $a_n = 0$ whenever n is a multiple of 3.

Determine the maximum value of $f(0)$ as f ranges through $\mathcal{C}$, and prove that this maximum is attained.

B3. Let $\mathcal{P}$ be a nonempty collection of subsets of $\{1, \ldots, n\}$ such that:

(i) if $S, S' \in \mathcal{P}$, then $S \cup S' \in \mathcal{P}$ and $S \cap S' \in \mathcal{P}$, and
(ii) if $S \in \mathcal{P}$ and $S \ne \varnothing$, then there is a subset $T \subset S$ such that $T \in \mathcal{P}$ and T contains exactly one fewer element than S.

Suppose that $f : \mathcal{P} \to \mathbb{R}$ is a function such that $f(\varnothing) = 0$ and

$$f(S \cup S') = f(S) + f(S') - f(S \cap S') \quad \text{for all } S, S' \in \mathcal{P}.$$

Must there exist real numbers $f_1, \ldots, f_n$ such that

$$f(S) = \sum_{i\in S} f_i$$

for every $S \in \mathcal{P}$?

B4. For any continuous real-valued function f defined on the interval [0, 1], let

$$\mu(f) = \int_0^1 f(x)\,dx, \quad \mathrm{Var}(f) = \int_0^1 \big(f(x) - \mu(f)\big)^2\,dx, \quad M(f) = \max_{0 \le x \le 1} |f(x)|.$$

Show that if f and g are continuous real-valued functions defined on the interval [0, 1], then

$$\mathrm{Var}(fg) \le 2\,\mathrm{Var}(f)M(g)^2 + 2\,\mathrm{Var}(g)M(f)^2.$$

B5. Let $X = \{1, 2, \ldots, n\}$, and let $k \in X$. Show that there are exactly $k \cdot n^{n-1}$ functions $f : X \to X$ such that for every $x \in X$ there is a $j \ge 0$ such that $f^{(j)}(x) \le k$. [Here $f^{(j)}$ denotes the jth iterate of f, so that $f^{(0)}(x) = x$ and $f^{(j+1)}(x) = f(f^{(j)}(x))$.]

B6. Let $n \ge 1$ be an odd integer. Alice and Bob play the following game, taking alternating turns, with Alice playing first. The playing area consists of n spaces, arranged in a line. Initially all spaces are empty. At each turn, a player either

- places a stone in an empty space, or
- removes a stone from a nonempty space s, places a stone in the nearest empty space to the left of s (if such a space exists), and places a stone in the nearest empty space to the right of s (if such a space exists).

Furthermore, a move is permitted only if the resulting position has not occurred previously in the game. A player loses if he or she is unable to move. Assuming that both players play optimally throughout the game, what moves may Alice make on her first turn?

SOLUTIONS

Solution to A1. Label the vertices $1, 2, \ldots, 12$. Let s_i denote the sum of the numbers on the five faces that meet at vertex i. Because each face contributes its number to exactly three of these vertex sums, we have $\sum_{i=1}^{12} s_i = 3 \cdot 39 < 120$. Thus, the average of the s_i is less than 10, and so there is an i for which $s_i < 10$. However, the sum of any five distinct nonnegative integers is at least $0 + 1 + 2 + 3 + 4 = 10$, so the five numbers that contribute to s_i are not distinct.

Solution to A2. Suppose not. Then there exist m and n in S with $m < n$ and $f(m) = f(n)$; let $n \cdot a_1 \cdot a_2 \cdots a_r$ and $m \cdot b_1 \cdot b_2 \cdots b_s$ be corresponding products as in the problem statement (in particular, they are perfect squares), with a_r, b_s minimal and $a_r = b_s$. If we multiply together these products, put the factors in nondecreasing order, and then omit all factors that appear twice (including $a_r = b_s$), we will have a new product for m as in the problem statement, and the largest remaining factor in this product will be less than b_s, a contradiction.

Solution to A3. The claim is obviously true if $a_0 = a_1 = \cdots = a_n = 0$, so assume this is not the case. Define the nonzero polynomial

$$f(x) = a_0x + a_1x^2 + a_2x^3 + \cdots + a_nx^{n+1}.$$

If we use the expansion $(1 - x^k)^{-1} = 1 + x^k + x^{2k} + x^{3k} + \cdots$, which is absolutely convergent because $0 < x < 1$, we can rewrite the given condition as

$$f(1) + f(x) + f(x^2) + f(x^3) + \cdots = 0.$$

Because f has finitely many roots, not all the terms in this sum can vanish, so there must be two nonzero terms of opposite sign. If these are $f(x^i)$ and $f(x^j)$ with $i < j$, then by the Intermediate Value Theorem there exists a number y in the interval (x^j, x^i) such that $f(y) = 0$. But then $0 < y < 1$ and $f(y)/y = a_0 + a_1 y + \cdots + a_n y^n = 0$, so we are done.

Solution to A4. Let n be the total number of digits around the circle, and let W denote the whole circle, that is, the unique arc of length n.

Suppose $L > 0$ is any integer. Consider the n arcs of length L beginning at each of the n positions. The total number of zeros they contain is $L\,Z(W)$, and so the average number of zeros they contain is $L\,Z(W)/n$. If this is not an integer, it follows that if w is an arc of length L, then $Z(w)$ is either $\lceil L\,Z(W)/n \rceil$ or $\lfloor L\,Z(W)/n \rfloor$, and both possibilities occur. If $L\,Z(W)/n$ is an integer, it follows that if w is an arc of length L, then $Z(w)$ is either $L\,Z(W)/n$ or one of $(L\,Z(W)/n) \pm 1$; but if one of the latter actually occurred, they couldn't both occur and so the average could not be $L\,Z(W)/n$. Thus $Z(w) = L\,Z(W)/n$ for all such w. Write $\alpha = Z(W)/n$, and note that in either case $Z(w) = \lceil \alpha L \rceil$ or $\lfloor \alpha L \rfloor$ if w has length L, with both possibilities occurring for some w.

Under the assumption in the problem, let $L = N + Z$, so that $L \in \mathbb{Z}$. It is necessary and sufficient to show that there exists an arc of length L with Z zeros. Now consider our family of arcs $w_1, \ldots, w_k$, of lengths $L_1, \ldots, L_k$ respectively, with average length $L \in \mathbb{Z}$ and average number of zeros $Z \in \mathbb{Z}$. Then $Z(w_j) = \lceil \alpha L_j \rceil$ or $Z(w_j) = \lfloor \alpha L_j \rfloor$ for all j. In either case $|Z(w_j) - \alpha L_j| < 1$; thus $|Z - \alpha L| < 1$, and Z is either $\lceil \alpha L \rceil$ or $\lfloor \alpha L \rfloor$. Since both possibilities must occur as $Z(w)$ for some arc w of length L, the result follows.

Solution to A5. Let $T = \triangle DEF$ be an arbitrary triangle in $\mathbb{R}^3$, and let $\mathbf{t} = \overrightarrow{DE} \times \overrightarrow{DF}$; then $\text{Area}(T) = |\mathbf{t}|/2$. If we project T onto a plane P with unit normal vector $\mathbf{n}$, then the area of that projection equals $|\cos\theta| \cdot \text{Area}(T)$, where $0 \le \theta \le \pi$ is the angle between P and the plane that T lies in.

Now think of a variable vector $\mathbf{n}$ ranging over the unit sphere, and the corresponding normal plane P. We can integrate the areas of the projections $\pi_{\mathbf{n}}(T)$ of T onto P over all such $\mathbf{n}$ by using spherical coordinates with $\mathbf{t}$ in the direction of one of the poles. This yields

$$\iint_{S^2} \text{Area}(\pi_{\mathbf{n}}(T))\, dS = \int_0^{\pi} \int_0^{2\pi} |\cos\theta| \cdot \text{Area}(T) \cdot \sin\theta \, d\phi \, d\theta = 2\pi \cdot \text{Area}(T),$$

and, conversely, we can express the area of T using the integral of the areas of the projections.

Now suppose that a given list of numbers a_{ijk} is area definite for $\mathbb{R}^2$. Then, for any points $A_1, \ldots, A_m$ in $\mathbb{R}^3$, we have

$$\begin{aligned}
&\sum_{1 \le i < j < k \le m} a_{ijk} \cdot \text{Area}(\triangle A_i A_j A_k) \\
&\quad = \frac{1}{2\pi} \sum_{1 \le i < j < k \le m} a_{ijk} \cdot \iint_{S^2} \text{Area}(\pi_{\mathbf{n}}(\triangle A_i A_j A_k))\, dS \\
&\quad = \frac{1}{2\pi} \iint_{S^2} \sum_{1 \le i < j < k \le m} a_{ijk}\, \text{Area}(\triangle \pi_{\mathbf{n}}(A_i) \pi_{\mathbf{n}}(A_j) \pi_{\mathbf{n}}(A_k))\, dS \ge 0,
\end{aligned}$$

where the last step uses the fact that the list of a_{ijk} is area definite for $\mathbb{R}^2$, applied to the points $\pi_{\mathbf{n}}(A_1), \ldots, \pi_{\mathbf{n}}(A_m)$.

Solution to A6. For $0 \le x \le 1$, $0 \le y \le 1$ let

$$T(x, y) = \sum_{(a,b)\in\mathbb{Z}\times\mathbb{Z}} w(a, b)\, e^{2\pi i(ax+by)} \quad \text{and} \quad A(x, y) = \sum_{\mathbf{s}=(s_1,s_2)\in S} e^{2\pi i(s_1x+s_2y)}.$$

(Note that $T(x, y)$ is actually a finite sum, by the definition of $w(a, b)$.) Then

$$\begin{aligned} |A(x, y)|^2 &= \sum_{\mathbf{s}\in S} e^{2\pi i(s_1x+s_2y)} \sum_{\mathbf{s}'\in S} e^{-2\pi i(s_1'x+s_2'y)} \\ &= \sum_{(\mathbf{s},\mathbf{s}')\in S\times S} e^{2\pi i((s_1-s_1')x+(s_2-s_2')y)}, \end{aligned}$$

so

$$T(x, y)\,|A(x, y)|^2 = \sum_{(a,b)\in\mathbb{Z}\times\mathbb{Z}} \sum_{(\mathbf{s},\mathbf{s}')\in S\times S} w(a, b)\, e^{2\pi i((a+s_1-s_1')x+(b+s_2-s_2')y)}.$$

If we integrate this over the unit square, each exponential with $a + s_1 - s_1' = b + s_2 - s_2' = 0$ will contribute 1, and all other contributions will be zero. Using the symmetry $w(a, b) = w(-a, -b)$, we can therefore write

$$\int_0^1 \int_0^1 T(x, y)\,|A(x, y)|^2\, dx\, dy = \sum_{(\mathbf{s},\mathbf{s}')\in S\times S} w(\mathbf{s} - \mathbf{s}') = A(S).$$

Thus, to show that $A(S) > 0$, it is enough to show that $T(x, y) \ge 0$ for all x, y and that $T(x, y)\, A(x, y)$ is not identically zero. Grouping the summands in $T(x, y)$ with $(a, b) \ne (0, 0)$ in conjugate pairs, we get

$$\begin{aligned} T(x, y) &= 12 - 8\cos 2\pi x - 8\cos 2\pi y + 4\cos 4\pi x + 4\cos 4\pi y \\ &\quad + 8\cos 2\pi(x+y) + 8\cos 2\pi(x-y) - 4\cos 2\pi(2x+y) - 4\cos 2\pi(2x-y) \\ &\quad - 4\cos 2\pi(x+2y) - 4\cos 2\pi(x-2y) - 2\cos 4\pi(x+y) - 2\cos 4\pi(x-y) \\ &= 12 - 8\cos 2\pi x - 8\cos 2\pi y + 4(2\cos^2 2\pi x - 1) + 4(2\cos^2 2\pi y - 1) \\ &\quad + 16(\cos 2\pi x)(\cos 2\pi y) - 8(2\cos^2 2\pi x - 1)\cos 2\pi y \\ &\quad - 8(2\cos^2 2\pi y - 1)\cos 2\pi x - 4(2\cos^2 2\pi x - 1)(2\cos^2 2\pi y - 1) \\ &= 16(a^2 + b^2 + ab - a^2b - ab^2 - a^2b^2), \end{aligned}$$

where $a = \cos 2\pi x$ and $b = \cos 2\pi y$. Now $-1 \le a, b \le 1$; if also $ab \ge 0$, then

$$a^2 + b^2 + ab - a^2b - ab^2 - a^2b^2 = a^2(1 - b) + b^2(1 - a) + ab(1 - ab) \ge 0.$$

On the other hand, if $ab \le 0$, then

$$a^2 + b^2 + ab - a^2b - ab^2 - a^2b^2 = (a + b)^2 - ab(1 + a)(1 + b) \ge 0.$$

So we have shown that $T(x, y) \ge 0$ for all x, y. Also, we see from the above that $T(x, y)$ can only be zero if a and b are in the set $\{-1, 0, 1\}$, that is, if x and y are integer multiples of $1/4$. On the other hand, $A(0, 0)$ is the cardinality of S, so by continuity, $A(x, y)$ is nonzero for x, y sufficiently close to zero. Hence $T(x, y)\, A(x, y)$ is not identically zero, and we are done.

Solution to B1. From the rules, we have $c(2m)c(2m+2) = c(m)c(m+1)$ and

$$c(2m+1)c(2m+3) = (-1)^m c(m)(-1)^{m+1}c(m+1) = -c(m)c(m+1).$$

Therefore, the terms in the sum cancel in pairs, starting with the term for $n = 2$. Thus the sum is equal to $c(1)c(3) = -1$.

Solution to B2. First note that $\cos(2\pi n/3) = -1/2$ for all integers n not divisible by 3. Therefore, for any f in $\mathcal{C}_N$,

$$f(1/3) = 1 - \frac{1}{2}\sum_{n=1}^{N} a_n \geq 0,$$

and so

$$\sum_{n=1}^{N} a_n \leq 2 \quad \text{and} \quad f(0) = 1 + \sum_{n=1}^{N} a_n \leq 3.$$

To show that 3 is in fact attained as a maximum value, consider

$$f(x) = 1 + \frac{4}{3}\cos(2\pi x) + \frac{2}{3}\cos(4\pi x),$$

which clearly satisfies (ii). As for (i), this can be checked by finding the critical points of f, or by noting that

$$f(x) = \frac{1}{3}(2\cos(2\pi x) + 1)^2 = \frac{1}{3}\left(\frac{\sin(3\pi x)}{\sin(\pi x)}\right)^2.$$

Solution to B3. Yes. First we show that if $T_1, T_2 \in \mathcal{P}$, $i \notin T_1, T_2$, and $T_1 \cup \{i\}$, $T_2 \cup \{i\} \in \mathcal{P}$, then

$$f(T_1 \cup \{i\}) - f(T_1) = f(T_2 \cup \{i\}) - f(T_2).$$

To see this, apply the condition on f to $S = T_1 \cup \{i\}$, $S' = T_2$ to get

$$f(T_1 \cup T_2 \cup \{i\}) = f(T_1 \cup \{i\}) + f(T_2) - f(T_1 \cap T_2),$$

and also apply it to $S = T_1$, $S' = T_2 \cup \{i\}$ to get

$$f(T_1 \cup T_2 \cup \{i\}) = f(T_1) + f(T_2 \cup \{i\}) - f(T_1 \cap T_2);$$

subtract the two equations from each other. Now we can define f_i to be the value of $f(T \cup \{i\}) - f(T)$ for any set $T \in \mathcal{P}$ for which $i \notin T$ and $T \cup \{i\} \in \mathcal{P}$; if there is no such set T, f_i can be chosen arbitrarily. Then the desired equation

$$f(S) = \sum_{i \in S} f_i$$

follows by induction on the cardinality of S, using (ii).

Solution to B4. For any f, the quantity

$$\int_0^1 \left(f(x) - a\right)^2 dx = \int_0^1 (f(x))^2\, dx - 2\mu(f)a + a^2$$

is a quadratic polynomial in a, which has its minimum when $a = \mu(f)$. Applying this to fg, we have

$$\operatorname{Var}(fg) = \int_0^1 \big(f(x)g(x) - \mu(fg)\big)^2\,dx \le \int_0^1 \big(f(x)g(x) - \mu(f)\mu(g)\big)^2\,dx.$$

Now note that

$$f(x)g(x) - \mu(f)\mu(g) = (f(x) - \mu(f))g(x) + (g(x) - \mu(g))\mu(f),$$

and that

$$\begin{aligned}&\big((f(x) - \mu(f))g(x) + (g(x) - \mu(g))\mu(f)\big)^2\\ &\qquad \le 2\Big(\big((f(x) - \mu(f))g(x)\big)^2 + \big((g(x) - \mu(g))\mu(f)\big)^2\Big).\end{aligned}$$

Thus,

$$\begin{aligned}&\int_0^1 \big(f(x)g(x) - \mu(f)\mu(g)\big)^2\,dx\\ &\qquad \le 2\int_0^1 \big((f(x) - \mu(f))g(x)\big)^2\,dx + 2\int_0^1 \big((g(x) - \mu(g))\mu(f)\big)^2\,dx\\ &\qquad \le 2\int_0^1 \big(f(x) - \mu(f)\big)^2\,dx\, M(g)^2 + 2\int_0^1 \big(g(x) - \mu(g)\big)^2\,dx\, M(f)^2,\end{aligned}$$

proving the stated result.

Solution to B5 (based on a student paper). First note that to each of the functions to be counted we can associate a rooted forest (disjoint union of trees) with k connected components (the trees) with roots labeled $1, 2, \ldots, k$ and vertices labeled $1, 2, \ldots, n$, namely the graph F with vertex set X and edges from x to $f(x)$ whenever $x \in X$ and $x > k$. The condition on the functions f is precisely what is needed to ensure that every vertex x does occur in exactly one of the trees (the one with root $f^{(j)}(x)$, where $j \ge 0$ is minimal such that $f^{(j)}(x) \le k$) and that there are no cycles. In fact, there is a bijection between the set of functions to be counted and the set of ordered pairs (F, g), where F is such a forest and g is any function from $\{1, 2, \ldots, k\}$ to $\{1, 2, \ldots, n\}$. In the forward direction, given f, the forest F is defined as above and the function g is the restriction of f to $\{1, 2, \ldots, k\}$. In the reverse direction, given (F, g), the function f is defined by $f(x) = g(x)$ for $x \in \{1, 2, \ldots, k\}$ (that is, when x is one of the roots) and $f(x) =$ the parent of x for $x > k$.

The number of possible functions g is n^k, so it remains to count the number of possible rooted forests F. Note that because the number of edges of any tree is one fewer than the number of vertices, the number of edges of such a forest is $n - k$. We now temporarily drop the restriction that the set of roots be $1, 2, \ldots, k$ and, on the other hand, add a labeling on the *edges* of the forest.

Claim: The number of rooted forests with k connected components, vertices labeled $1, 2, \ldots, n$, and edges labeled $1, 2, \ldots, n - k$ is

$$A(n, k) = n^{n-k}(n-1)(n-2)\cdots(k+1)k = n^{n-k}\binom{n-1}{k-1}\cdot(n-k)!\,.$$

The proof of the claim is by induction on $n - k$; for the base case $n - k = 0$, there is just one possible forest, and all is well. Given a rooted forest with $k + 1$ components, vertices labeled $1, 2, \ldots, n$, and edges labeled $1, 2, \ldots, n - k - 1$, we can make a rooted forest with one fewer component by choosing any vertex v of a tree and connecting it to one of the k roots of the *other* trees by a new edge labeled $n-k$, thus creating a "merged" tree (with the same root as the original tree containing v). This process will create every rooted forest with k connected components, vertices labeled $1, 2, \ldots, n$, and edges labeled $1, 2, \ldots, n-k$ exactly once; because there are n choices for the vertex v and k choices for the root to connect it to, we have

$$A(n,k) = A(n,k+1)\cdot n\cdot k\,,$$

and the claim follows.

Having established the claim, we discard the labeling on the edges, which results in the number of forests being divided by $(n-k)!$, and we reintroduce the restriction that the roots be $1, 2, \ldots, k$, which results in a further division by $\binom{n}{k}$ (the number of possible sets of k roots, only one of which is acceptable). Therefore, the number of possible rooted forests F above is

$$\frac{A(n,k)}{(n-k)!\binom{n}{k}} = n^{n-k-1}k,$$

so the number of ordered pairs (F, g) is

$$n^k \cdot n^{n-k-1}k = k\cdot n^{n-1},$$

and we are done.

Solution to B6. Write $n = 2k-1$. Alice must play in space k (the middle space).

We claim that a position in which fewer than $n-1$ spaces are occupied is a winning position for the current player if and only if the sum of the occupied spaces is not an odd multiple of k.

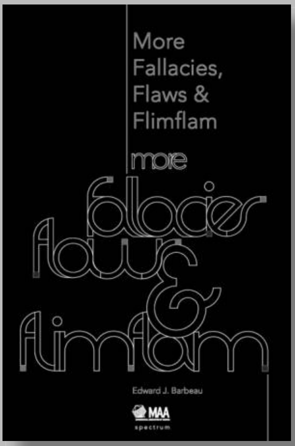
Note that the number of stones cannot decrease unless all spaces are occupied. Moreover, suppose there are at least two empty spaces. Then at each move, the number of stones increases by one unless a player removes a stone from a sequence of stones of the form $1, \ldots, j$ or $j, \ldots, n$; and in that case the sum $\min\{i : i \text{ empty}\} + (n - \max\{i : i \text{ empty}\})$ always decreases. It follows that no sequence of moves could cause a position to be repeated (even if that were permitted), except for the positions with at least $n-1$ stones.

Observe that any position with at least $n-1$ stones, in other words with 0 or 1 empty spaces, can be turned into any other (different) such position with a single move. It follows that once there are $n-1$ stones on the board, there are exactly n moves left in the game; since n is odd, the player with the first opportunity to play in that situation will win. Note that if the only unoccupied space is space i, then the sum of the occupied spaces is $k(2k-1)-i$, which is not an odd multiple of k.

Finally, note that in any position with at least two empty spaces, the $2k-1$ available moves can change the sum of the occupied spaces by any of $1, \ldots, 2k-1 \pmod{2k}$. First, playing in an unoccupied space i changes the sum by i. Second, removing a stone from space $j < i < J$ with j, J unoccupied but $j+1, \ldots, J-1$ occupied (allowing $j = 0$ and/or $J = 2k = n+1$) changes the sum by $j + J - i$; as i ranges from $j+1$ to $J-1$, the change in sum ranges from $J-1$ to $j+1$. So to change the sum by $i \pmod{2k}$, put a stone in space i if it is unoccupied; otherwise, find j, J as above and remove the stone from space $J + j - i$.

Thus, given a position with at least two empty spaces, if the sum is initially not an odd multiple of k, it can always be made so; while if the sum is initially an odd multiple of k, it can never remain so. The claim follows.

New from the MAA



More Fallacies, Flaws, & Flimflam
By Edward J. Barbeau
Spectrum Series

More Fallacies, Flaws, and Flimflam is the second volume of selections drawn mostly from *The College Mathematics Journal* column "Fallacies, Flaws, and Flimflam" from 2000 through 2008. The MAA published the first collection, *Mathematical Flaws, Fallacies, and Flimflam*, in 2000.

As in the first volume, *More Fallacies, Flaws, and Flimflam* contains items ranging from howlers (outlandish procedures that nonetheless lead to a correct answer) to deep or subtle errors often made by strong students. Although some are provided for entertainment, others challenge the reader to determine exactly where things go wrong.

Items are sorted by subject matter. Elementary teachers will find chapter 1 of most use, while middle and high schoolteachers will find chapters 1, 2, 3, 7, and 8 applicable to their levels. College instructors can delve for material in every part of the book.

There are frequent references to *The College Mathematics Journal*; these are denoted by *CMJ*.

Catalog Code: MFFL
Paperbound, 188 pp., 2013
ISBN 978-0-88385-580-5
List: $35.00
MAA Member: $28.00

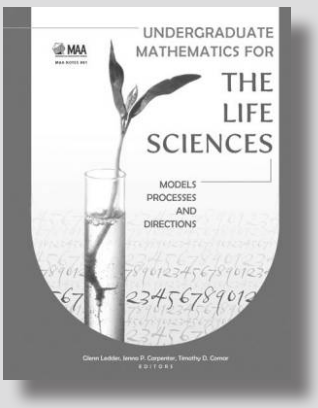
To order, visit maa-store.hostedbywebstore.com
or call 800-331-1622.

MAA
MATHEMATICAL ASSOCIATION OF AMERICA

Latest in the MAA Notes Series

Undergraduate Mathematics for the Life Sciences
Models, Processes, and Directions

Glenn Ledder, Jenna P. Carpenter, Timothy D. Comar, Editors



In this Notes volume, authors from a variety of institutions address some of the problems involved in reforming mathematics curricula for biology students. The problems are sorted into three themes: Models, Processes, and Directions. It is difficult for mathematicians to generate curriculum ideas for the training of biologists so a number of the curriculum models that have been introduced at various institutions comprise the Models section. Processes deals with taking that great course and making sure it is institutionalized in both the biology department (as a requirement) and in the mathematics department (as a course that will live on even if the creator of the course is no longer on the faculty). Directions looks to the future, with each paper laying out a case for pedagogical developments that the authors would like to see.

The authors represent a wide variety of academic institutions, from universities to community colleges, and all of the articles begin with information about the institutional context. Many of the articles also include links to resources that can be found on the internet, and some have associated books in print as well. All emphasize features that could be applied to similar projects at other institutions and offer useful advice for the newcomer to mathematics curriculum development for life science students.

2013, 226 pages
eISBN: 9781614443162
Print ISBN: 9780883851913

ebook: $25.00
Print on demand (paperbound): $43.00

To order go to www.maa.org/ebooks/NTE81

MAA
MATHEMATICAL ASSOCIATION OF AMERICA

MAA
MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth St., NW • Washington, DC 20036

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS