LETTER FROM THE EDITOR

We start off April's offerings of mathematical goodness by revisiting the venerable topic of Voronoi tessellations. To construct one, we start with a set of "seed" points in the plane. The region in the tessellation associated with a given seed is then the set of points closer to that seed than to any other. Diagrams of this sort have found applications in both pure and applied mathematics. In our lead article, Kathrin Gillespie, Sophia Kotok, and Albert Schueller discuss centroidal Voronoi tessellations, in which each seed is the centroid of its associated region. The result is a lively mixture of calculus, geometry, and analysis, building up to some exceedingly clever constructions.

Antonella Cupillari and Saaman Khalilollahi discuss a classic problem about circles and tangent lines. The problem first appeared in Maria Gaetana Agnesi's classic book *Instituzioni Analitichie*, published in 1748. The authors find some errors in Agnesi's work and present a complete solution to this fascinating geometry problem. As the author of a major biography of Agnesi, Dr. Cupillari has done much to increase awareness of this unjustly forgotten figure in the history of mathematics. Most mathematicians will be familiar with the so-called "Witch of Agnesi," but there was far more to her work than that.

From there we move on to another venerable topic: cake-cutting algorithms. This subject has come a long way from its origins with the "I cut, you choose," method for dividing a piece of cake between two people. That procedure is said to be "envy-free," meaning roughly that both people will feel they have been treated fairly. Researchers have devised envy-free procedures for dividing cakes among larger numbers of people, though it must be admitted these procedures are, generally, wildly impractical if you are imagining actual cakes being sliced with physical knives. Steven Brams and Peter Landweber serve up a novel, envy-free procedure for dividing a cake among three people. Their procedure is a significant advance over currently known algorithms, as they explain so lucidly in their article.

This issue's shorter pieces also provide much food for thought. Artem Logachev, Olga Logachova, and Stanislav Russijan consider a deceptively simple combinatorial problem inspired by ticket numbers. Their investigation quickly leads them to generating functions and the Cauchy integral formula. Michael Albert, Bjarki Gudmundsson, and Henning Ulfarsson present a surprising connection between the Collatz conjecture and the Fibonacci numbers. With the article by Fábio Lima and Pedro Jordão we turn to multivariable calculus, as they present a matrix-based approach to the problem of deriving the gradient and divergence operators in cylindrical and spherical coordinates. Their approach has several advantages over what is typically found in textbooks. Wenchang Chu then brings us home with a clever, and apparently novel, proof of the Chu-Vandermonde formula for convolutions of binomial coefficients.

We round out the issue with Problems, Reviews, a pair of proofs without words, and a report on the 62nd annual International Mathematical Olympiad. That should be enough to keep you busy until we do this all again in the June issue!

Jason Rosenhouse, Editor

Math. Mag. **95** (2022) 87. doi:10.1080/0025570X.2022.2046409 © Mathematical Association of America

ARTICLES

Two-Point Centroidal Voronoi Tessellations

KATHRIN GILLESPIE
Redfin
Seattle, WA 98101
kathringillespie@gmail.com

SOPHIA KOTOK
Rochester Institute of Technology
Rochester, NY 14623
sophie.kotok@gmail.com

ALBERT W. SCHUELLER
Whitman College
Walla Walla, WA 99362
schuelaw@whitman.edu

The Voronoi tessellation (VT) of a plane figure is a partition of the figure into a collection of subsets relative to a designated set of "special" points in the figure called *generators*. For example, the partition of the square figure shown in the upper-left of Figure 1 is a VT for the 10 generators (black dots) inside the square. While there are a lot of possible partitions of a plane figure like a square, the VT is the partition in which each subset contains exactly one generator and all of the points nearer that generator than any other. Points that are equidistant from two or more generators lie on the boundaries of these subsets.

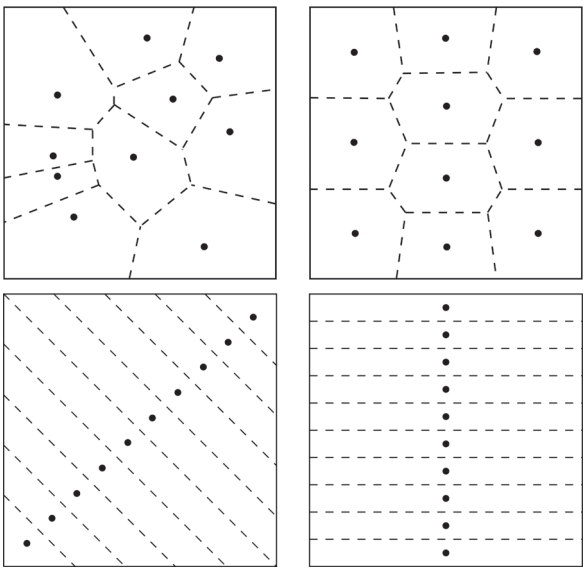


Figure 1 *Upper-left:* A VT of a square with 10 randomly placed generators. *Upper-right:* A CVT of a square with 10 generators. *Lower-left:* A 10-point CVT with the centers along a diagonal. *Lower-right:* A 10-point CVT with the centers along the vertical line of symmetry.

Math. Mag. **95** (2022) 88–98. doi:10.1080/0025570X.2022.2025721 © Mathematical Association of America

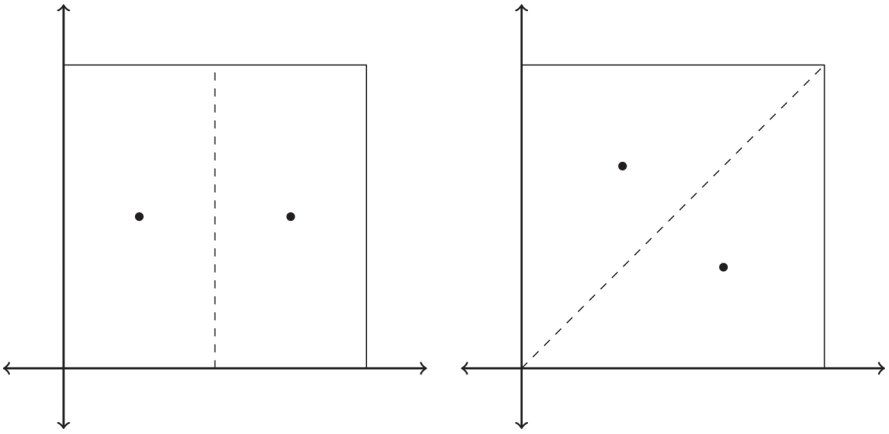


Figure 2 Two-point CVTs of a square. This is a reproduction of a similar figure from Du, Faber, and Gunzberger [**2**]. The CVTs coincide with lines of symmetry.

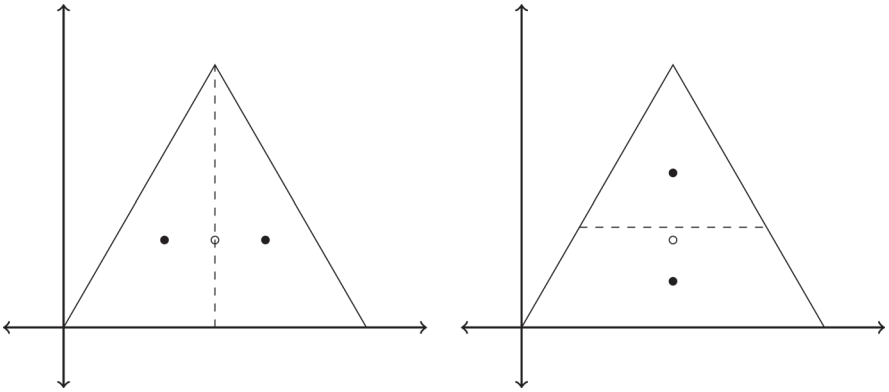


Figure 3 Two-point CVTs of an equilateral triangle. The one on the left lies on a line of symmetry, the one on the right does not. The hollow points are the centroids of the entire triangles.

A *centroidal* Voronoi tessellation (CVT) is a VT whose generator points are *also* the centroids of their respective regions. The remaining VTs in Figure 1 are also CVTs.

Figure 1 shows that CVTs are not unique for a particular figure and number of generators. Indeed, we see there are at least three different 10-point CVTs of a square. A natural question to ask is: *What are the possible CVTs of a plane figure?* We explore this question in the more tractable case of only two generators in plane figures with even rotational symmetry and a line of reflective symmetry.

The two figures in Figure 2 show two different two-point CVTs of the square. A square is a figure with rotational symmetry of even order and several lines of symmetry. In both cases, we see the boundary between the Voronoi regions lies along a line of symmetry. For a long time, the authors held to the conjecture that: *The two-point CVTs of a bounded, convex, symmetric figure with even rotational symmetry must coincide with the lines of symmetry of the figure*. Surprisingly, this conjecture is false. A counterexample is shown in Figure 7.

The conjecture is more obviously false for symmetric plane figures with odd rotational symmetry. For example, consider the CVTs of the equilateral triangle in Figure 3. The left-hand CVT aligns with a line of symmetry of the triangle, but the right-hand CVT does not.

As a very applied area of mathematics, there is relatively little geometric exploration of CVTs in the literature. An excellent review of the applications of CVTs is given by Du, Faber, and Gunzberger [2]. The current article was almost entirely motivated by a figure from that article and is reproduced here in Figure 2. We will show that the conjecture is true for a set of "ellipse-like" figures which we define in Section . Finally, in Section , we show how to construct counterexamples to the conjecture.

Two-point CVTs

VTs can be formulated in quite general contexts in exotic spaces with generic metrics. For simplicity, we focus here on VTs in the plane with the Euclidean metric. A general review of VTs is provided in Okabe, Boots, Sugihara, and Chiu [8].

Voronoi tessellations Unless otherwise stated, we assume Ω to be a non-empty, convex, bounded, closed set in the plane. Let's first give a formal definition of a VT, taken from Du, Faber, and Gunzberger [2].

Definition 1 (Voronoi tessellation)**.** *Given a finite collection of k distinct points, $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k \in \Omega$, the VT of Ω is the set of sets $\{V_i\}_{i=1}^k$ defined by*

$$V_i = \{\mathbf{x} \in \Omega : |\mathbf{x} - \mathbf{z}_i| \leq |\mathbf{x} - \mathbf{z}_j| \text{ for } j = 1, \ldots, k,\ j \neq i\}.$$

Here, we refer to $\mathbf{z}_i$ as a *generator* and V_i as the *Voronoi region* corresponding to $\mathbf{z}_i$. The *Voronoi boundary* between V_i and V_j is denoted and defined by $\partial V_{i,j} = V_i \cap V_j$. The VT of a square with 10 randomly placed generators is shown in Figure 1.

While there does not, at first, appear to be much structure in the VT in upper-left of Figure 1, there are a number of important statements we can make about this VT and VTs in general. Proofs of these statements can be found in Okabe, Boots, Sugihara, and Chiu [8].

Theorem 1. *If Ω is convex, then the Voronoi regions, V_i, are convex.*

Theorem 2. *A non-empty Voronoi boundary, $\partial V_{i,j}$, is a perpendicular bisector of the line segment $\overline{\mathbf{z}_i\mathbf{z}_j}$.*

The VT in the upper-right of Figure 1 is more regular in appearance than the one in the upper-left. The generators here are also the centroids of their regions. In this case, the VT is called a CVT. (Note: This figure is the result of applying Lloyd's algorithm [6] to the VT on the left. Because Lloyd's algorithm converges very slowly, the figure is only an approximation of a CVT.) The remaining VTs in Figure 1 are also CVTs. We give the formal definition of a CVT.

Definition 2 (CVT)**.** *A CVT is a VT for which*

$$\mathbf{z}_i = \frac{\int_{V_i} \mathbf{x}\, d\mathbf{x}}{\int_{V_i} d\mathbf{x}}$$

for $i = 1, 2, \ldots k$. (In other words, each generator, $\mathbf{z}_i$, is the centroid of its corresponding region, V_i.)

Two-point CVTs As we see in Figure 1, CVTs are quite regular in appearance, but characterizing all of the possible 10 point CVTs of a square is a challenge. We therefore limit our attention to cases of exactly two generators, as in Figure 2.

We will call the two generators $\mathbf{z}_1$ and $\mathbf{z}_2$ and their corresponding regions V_1 and V_2. Since there is now only one Voronoi boundary, we drop the subscript notation and simply let $\partial V = V_1 \cap V_2$. In what follows, it is useful to borrow the projection operator notation, π_i, from, for example, Spivak [9]. That is, if $\mathbf{z} = (z_1, z_2) \in \mathbb{R}^2$, then $\pi_i(\mathbf{z}) = z_i,\ i = 1, 2$.

Since we are interested in CVTs of sets with symmetry, we introduce rotational and reflective symmetry. Borrowing from Jennings [5], we say a set Ω has *even rotational symmetry* about a point P if $\mathbf{z} \in \Omega$ implies that the rotation of $\mathbf{z}$ about P by 180° also lies in Ω. Also from Jennings, we say that Ω is symmetric about the line ℓ if $\mathbf{z} \in \Omega$ implies that the reflection of $\mathbf{z}$ across ℓ also lies in Ω.

Ultimately, we will need to place Ω on a coordinate plane with the origin at the point of rotational symmetry. It is not difficult to show that the point of rotational symmetry of a bounded set is unique. For a figure of even rotational symmetry about P, we also have that P is the centroid of the figure.

Theorem 3. *If Ω is a bounded convex set with order two rotational symmetry about a point P, then P is the centroid of Ω.*

Proof. Position Ω so that P lies at the origin. Let $\mathbf{z}_\Omega$ be the centroid of Ω. Let $f(\theta)$ be the distance from P to the boundary of Ω along the ray of angle θ with the horizontal coordinate axis originating at P (the origin).

We note that f is well-defined because Ω is convex—that is, there cannot be two boundary points of Ω on the ray.

Since Ω has order two rotational symmetry about the origin, f is periodic with period π. That is, $f(\theta) = f(\theta + \pi)$. We show that $\mathbf{z}_\Omega = \mathbf{0}$ by showing that the moments

$$M_y = \int_\Omega x\, dA \text{ and } M_x = \int_\Omega y\, dA$$

of Ω about the y- and x-axes are both zero.

$$\begin{aligned} M_y &= \int_\Omega x\, dA \\ &= \int_{-\pi}^{\pi} \int_0^{f(\theta)} r^2 \cos\theta\, dr\, d\theta \\ &= \frac{1}{3}\int_0^{\pi} f^3(\theta)\cos\theta\, d\theta + \frac{1}{3}\int_{-\pi}^{0} f^3(\theta)\cos\theta\, d\theta \\ &= \frac{1}{3}\int_0^{\pi} f^3(\theta)\cos\theta\, d\theta - \frac{1}{3}\int_0^{\pi} f^3(t)\cos(t)\, dt = 0. \end{aligned}$$

Similarly, $M_x = 0$. Therefore, $\mathbf{z}_\Omega = P$. ∎

We conclude this section by showing that for a CVT of Ω, the centroid of Ω, $\mathbf{z}_\Omega$, is collinear with $\mathbf{z}_1$ and $\mathbf{z}_2$ and between them.

Theorem 4. *Let $\{(\mathbf{z}_1, V_1), (\mathbf{z}_2, V_2)\}$ be a CVT of a bounded set Ω. The centroid of Ω, $\mathbf{z}_\Omega$, lies strictly between $\mathbf{z}_1$ and $\mathbf{z}_2$ on the line segment $\overline{\mathbf{z}_1\mathbf{z}_2}$.*

Proof. Because $\mathbf{z}_\Omega$ is just a weighted mean of $\mathbf{z}_1$ and $\mathbf{z}_2$, we know that it lies between $\mathbf{z}_1$ and $\mathbf{z}_2$ on the line segment $\overline{\mathbf{z}_1\mathbf{z}_2}$. Since the weights of that weighted mean, namely $|V_1|$ and $|V_2|$ (the areas of V_1 and V_2 respectively), are each non-zero, we can conclude that $\mathbf{z}$ is strictly between $\mathbf{z}_1$ and $\mathbf{z}_2$. ∎

Finally, as we have alluded to earlier, for a bounded, convex set, Ω, it is useful to note that ∂V is a line segment in Ω and lies on the perpendicular bisector of $\overline{\mathbf{z}_1\mathbf{z}_2}$.

Figures with even order rotational symmetry

For two-point CVTs of figures with even rotational symmetry, we now know it is possible to orient the figure so that the centroid of the entire figure is at the origin, and so that the centroids of V_1 and V_2 lie on the x-axis on either side of the origin. We will show that ∂V lies on the y-axis. We establish this in Lemma 1.

Consider a convex set Ω in the plane with rotational symmetry of order two as in Figure 4. Start with a partition $\{V_1, V_2\}$ of Ω in which the boundary between V_1 and V_2, ∂V, is a line segment.

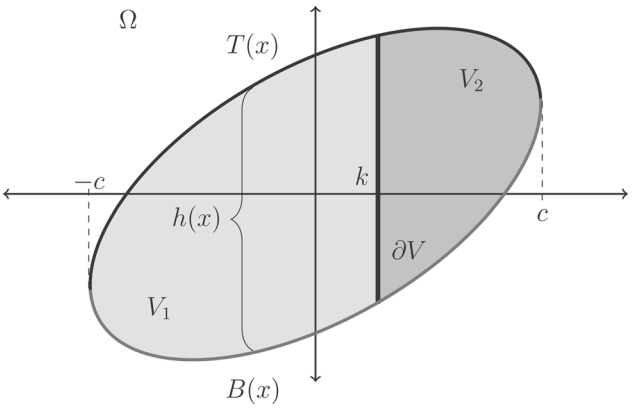


Figure 4 The region Ω with a generic partition $\{V_1, V_2\}$ in which ∂V is a vertical line segment at $x = k$. The set is oriented with its point of rotational symmetry at the origin and ∂V parallel to the y-axis. The top and bottom boundary curves are described by the functions $T(x)$ and $B(x)$, respectively. The height of Ω at a point x is given by $h(x) = T(x) - B(x)$.

Orient Ω in the coordinate plane with its point of rotational symmetry at the origin and ∂V parallel to the y-axis. We now prove a lemma that will be needed in the proof of the main result of this paper. We note that the lemma does not require that Ω have a *line* of symmetry—only that it have a *point* of rotational symmetry of order two. A generic parallelogram is an example of a set with no line of symmetry, but with rotational symmetry of order two. Hence, Lemma 1, as stated, would apply to two-point CVTs of parallelograms.

Lemma 1. *Let $\Omega \subset \mathbb{R}^2$ be a closed, bounded, convex set with rotational symmetry of order 2, and let $\{(\mathbf{z}_1, V_1), (\mathbf{z}_2, V_2)\}$ be a CVT of Ω. If the point of rotational symmetry of Ω is located at the origin and ∂V is parallel to the y-axis, then ∂V lies on the y-axis.*

Proof. In the context of Figure 4, with k indicating the position of ∂V, we assume that $\{(\mathbf{z}_1, V_1), (\mathbf{z}_2, V_2)\}$ is a CVT of Ω. For simplicity, we assume that $\pi_1(\Omega) = [-c, c]$ and let T and B be the top and bottom boundary curves of Ω. We note T and B are continuous on $[-c, c]$ (see Gordon [**3**, Theorem 4.28]). Let $h(x) = T(x) - B(x)$ be the "height" of the figure at x.

It follows from Theorems 3 and 4 that ∂V is the perpendicular bisector of $\overline{\mathbf{z}_1\mathbf{z}_2}$ and that $\mathbf{z}_1$ and $\mathbf{z}_2$ lie on the x-axis to the left and right of ∂V respectively.

We now focus on the x-coordinates of $\mathbf{z}_1$ and $\mathbf{z}_2$ given by $\pi_1(\mathbf{z}_1)$ and $\pi_1(\mathbf{z}_2)$. We will show that $\mathbf{z}_1$ and $\mathbf{z}_2$ are equidistant from the origin. Since ∂V is a bisector of $\overline{\mathbf{z}_1\mathbf{z}_2}$, we have $\pi_1(\mathbf{z}_1) + \pi_1(\mathbf{z}_2) = 2k$.

Recalling the centroid formula, we have

$$\pi_1(\mathbf{z}_1) = \frac{\int_{V_1} x\, dA}{\int_{V_1} dA} \qquad \text{and} \qquad \pi_1(\mathbf{z}_2) = \frac{\int_{V_2} x\, dA}{\int_{V_2} dA}.$$

Assume that $k \leq 0$ and observe that

$$\begin{aligned}
0 &\geq \pi_1(\mathbf{z}_1) + \pi_1(\mathbf{z}_2) \\
&= \frac{\int_{V_1} x\, dA}{\int_{V_1} dA} + \frac{\int_{V_2} x\, dA}{\int_{V_2} dA} \\
&= \frac{\int_{-c}^{k} \int_{B(x)}^{T(x)} x\, dy\, dx}{\int_{-c}^{k} \int_{B(x)}^{T(x)} dy\, dx} + \frac{\int_{k}^{c} \int_{B(x)}^{T(x)} x\, dy\, dx}{\int_{k}^{c} \int_{B(x)}^{T(x)} dy\, dx} \\
&= \frac{\int_{-c}^{k} xh(x)\, dx}{\int_{-c}^{k} h(x)\, dx} + \frac{\int_{k}^{c} xh(x)\, dx}{\int_{k}^{c} h(x)\, dx} \\
&= \frac{\int_{-c}^{-k} xh(x)\, dx + \int_{-k}^{k} xh(x)\, dx}{\int_{-c}^{-k} h(x)\, dx + \int_{-k}^{k} h(x)\, dx} - \frac{\int_{-c}^{-k} xh(x)\, dx}{\int_{-c}^{-k} h(x)\, dx}
\end{aligned}$$

where in the last line we broke up the integral in the numerator of the first term and performed the change of variable $x \to -x$ in both the numerator and denominator of the second term. We used the fact that $h(x)$ is even (and $xh(x)$ is odd). Continuing, we clear the common integral in the numerator and reverse the inequality because the integral is negative.

$$\begin{aligned}
\frac{\int_{-c}^{-k} xh(x)\, dx}{\int_{-c}^{-k} h(x)\, dx} &\geq \frac{\int_{-c}^{-k} xh(x)\, dx}{\int_{-c}^{-k} h(x)\, dx + \int_{-k}^{k} h(x)\, dx} \\
\int_{-c}^{-k} h(x)\, dx + \int_{-k}^{k} h(x)\, dx &\leq \int_{-c}^{-k} h(x)\, dx \\
\int_{-k}^{k} h(x)\, dx &\leq 0.
\end{aligned}$$

After canceling the common terms on the left and right of the second line above, we obtain the final line. Since h is nonnegative, we conclude that $k = 0$.

We can repeat this calculation under the assumption that $k \geq 0$, being careful to do the change of variable so that the integrals are on the interval $[k, c]$ instead of $[-c, -k]$ as above, to conclude again that $k = 0$. ■

Lemma 1 is about as much as we can say concerning the possible two-point CVTs of closed, bounded, convex sets of even order rotational symmetry. In the next section, we determine a sufficient condition for when ∂V must lie on a line of symmetry of such sets.

Ellipse-like figures

We now consider Ω a closed, bounded, convex set of even rotational symmetry with a line of symmetry. Given Lemma 1, we might expect that ∂V will lie on a line of symmetry. As noted at the beginning of the article, this conjecture is false. However, we have discovered a class of figures, which we will call "ellipse-like," for which the conjecture is true.

To define this class of figures, first let $f : [0, \pi/2] \to (0, \infty)$ describe a curve, $r = f(\theta)$, in the first quadrant in polar coordinates, (r, θ), as in Figure 5. If we further require that $x(\theta) = f(\theta)\cos\theta$ is nonincreasing, and $y(\theta) = f(\theta)\sin(\theta)$ is nondecreasing, then reflecting the curve across the y-axis and the result across the x-axis will generate the boundary of Ω as in Figure 5. The monotonicity conditions are necessary to insure convexity of Ω. All bounded, convex shapes with even rotational symmetry and a line of symmetry can be constructed in this way.

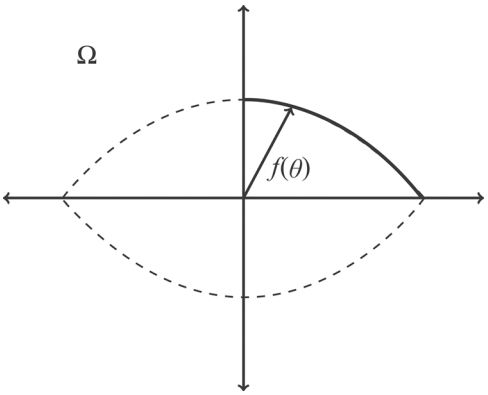


Figure 5 Construction of a generic Ω with a line of symmetry and rotational symmetry of order 2 by the curve $r = f(\theta)$ in the first quadrant. This set is ellipse-like, i.e. it satisfies the conditions of Definition 3.

To be ellipse-like, we will also require that f be injective. This subtler condition is necessary to make the geometric argument found in the proof of the main result, Theorem 5.

Definition 3 (Ellipse-like)**.** *Let $f : [0, \pi/2] \to (0, \infty)$ be a continuous function such that:*

1. *$x(\theta) = f(\theta)\cos\theta$ is nonincreasing, and $y(\theta) = f(\theta)\sin(\theta)$ is nondecreasing,*
2. *f is injective.*

Let A be a closed and bounded set whose boundary is generated by reflecting the curve $r = f(\theta)$ across the y-axis and reflecting the result across the x-axis. We say that a set Ω is ellipse-like *if it is a rigid transformation of such an A.*

In the case of an ellipse (non-circular) with axes of length $2a$ and $2b$, $a \neq b$, aligned along the x- and y-axes, respectively, we find $f(\theta) = ab/\sqrt{a^2\cos^2\theta + b^2\sin^2\theta}$. It is not difficult to show that f is strictly monotone on $[0, \pi/2]$, $f(\theta)\cos(\theta)$ is non-increasing, and $f(\theta)\sin(\theta)$ is non-decreasing. Hence, ellipses are ellipse-like. It is worth noting that circles are *not* ellipse-like because f is constant and hence not injective. It is also the case that rectangles are not ellipse-like because f fails to be injective.

We note that there are ellipse-like figures that are not ellipses. For example, in Figure 5, f describes the part of the parabola $y = 1 - x^2/4$ in the first quadrant. For this curve, $f(\theta) = 2(1 - \sin(\theta))\sec^2(\theta)$, $\theta \neq \pi/2$, $f(\pi/2) = 1$. We leave it to the reader to confirm that this figure is ellipse-like.

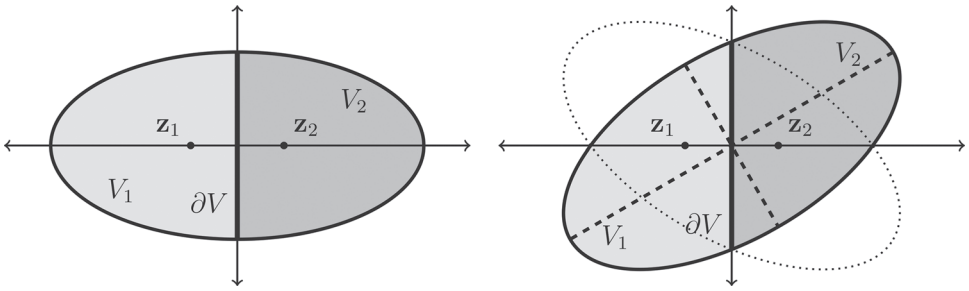


Figure 6 On the left, the ellipse-like region Ω and a CVT, $\{(\mathbf{z}_1, V_1), (\mathbf{z}_2, V_2)\}$, coinciding with a line of symmetry. On the right, the rotated region Ω with a fictive CVT partition $\{(\mathbf{z}_1, V_1), (\mathbf{z}_2, V_2)\}$. The dashed lines are the lines of symmetry of Ω.

Theorem 5. *If $\Omega \subset \mathbb{R}^2$ is an ellipse-like figure and $\{(\mathbf{z}_1, V_1), (\mathbf{z}_2, V_2)\}$ is a CVT of Ω, then ∂V lies on a line of symmetry of Ω.*

Proof. Let Ω be an ellipse-like figure. By Theorems 3 and 4 and Lemma 1, all two-point CVTs of the form $\{(\mathbf{z}_1, V_1), (\mathbf{z}_2, V_2)\}$ on Ω can be oriented so that the center of mass of Ω lies at the origin, ∂V lies on the y-axis, and $\mathbf{z}_1$ and $\mathbf{z}_2$ lie on the x-axis to either side of ∂V and equidistant from the y-axis.

Further, one can show that there is a unique two-point CVT of Ω corresponding to each line of symmetry. They are unique because the centroid of a set is unique.

Because Ω is ellipse-like, one can show that there are exactly two perpendicular lines of symmetry of Ω that intersect at the point of rotational symmetry. Orient Ω with its point of rotational symmetry at the origin and its lines of symmetry on the coordinate axes as on the left of Figure 6.

Suppose there is a CVT of Ω $\{(\mathbf{z}_1, V_1), (\mathbf{z}_2, V_2)\}$ with ∂V not on a line of symmetry as on the right of Figure 6. We know that we can orient Ω so that ∂V lies on the y-axis with $\mathbf{z}_1$ and $\mathbf{z}_2$ lying to either side on the x-axis. This is effectively just a rotation of Ω about its center of mass so that its lines of symmetry are not parallel to the coordinate axes.

Consider a copy of Ω rotated by the same amount in the opposite direction like the dotted figure on the right of Figure 6. Let f be the function that generates Ω as in Definition 3. Since f is injective and continuous, it is strictly monotone. Assume f is strictly decreasing as in Figure 6. Using monotonicity, one can show that the dotted curve divides V_2 into two parts. The part containing the origin is symmetric about the x-axis so its centroid lies on the x-axis. The other part lies above the x-axis, so its centroid does too. It follows that $\mathbf{z}_2$, as the weighted average of these centroids, must lie above the x-axis. This is a contradiction.

The case in which f is strictly increasing is similar. ■

A counterexample and some further study

We opened this article noting that the conjecture that *the two-point CVTs of a bounded, convex, symmetric figure with even rotational symmetry must coincide with the lines*

of symmetry of the figure is false. Figure 7 is a counterexample to this conjecture. The centroids of the left- and right-hand sides of the gray figure lie on the x-axis and are equidistant from the y-axis. Such counterexamples are not rare. We conclude with a discussion of these figures and how to construct them.

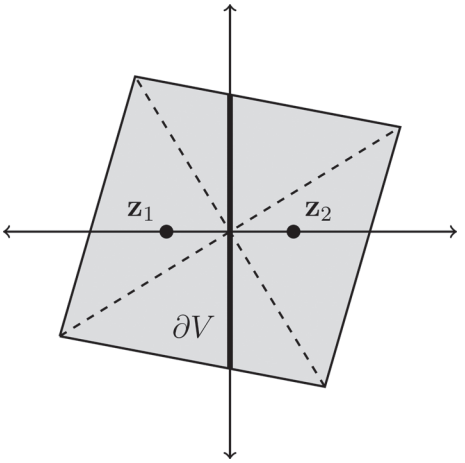


Figure 7 Counterexample to the original conjecture. The gray diamond with major axis 4.4 and minor axis 4 has a CVT with ∂V lying on the y-axis and not on a line of symmetry. It is symmetric and has rotational symmetry of order 2, but is not ellipse-like.

The diamond-shaped set in Figure 7 can be generated in the way described in Figure 5 where the base function, f, follows from the polar coordinate parameterization of the line segment from $(0, b)$ to $(c, 0)$,

$$f(\theta) = \frac{bc}{b\cos\theta + c\sin\theta}.$$

Although f satisfies the first condition of ellipse-like, it is not injective and hence the figure is not ellipse-like.

With this f and taking into account the reflective and rotational symmetries, we can express the full boundary function as

$$F(\theta) = \begin{cases} f(\pi + \theta) & \text{if } \theta \in [-\pi, -\pi/2) \\ f(-\theta) & \text{if } \theta \in [-\pi/2, 0) \\ f(\theta) & \text{if } \theta \in [0, \pi/2] \\ f(\pi - \theta) & \text{if } \theta \in [\pi/2, \pi] \end{cases}$$

The polar curve $r = F(\theta)$, $\theta \in [0, 2\pi)$, describes the boundary of a diamond-shaped figure with axes of length $2c$ in the x direction and $2b$ in the y direction. For $\phi \in [0, 2\pi)$, the curve $r = F(\theta - \phi)$ describes the boundary of the same diamond rotated counterclockwise by an angle ϕ.

By Lemma 1 and the symmetry of the diamond shaped figure, CVTs occur at angles ϕ such that $M_x(\phi) = 0$. This condition guarantees that the centroids of each Voronoi region lie on the x-axis. We can compute the moment of the right half of the diamond about the x-axis as a function of the angle, ϕ, by

$$M_x(\phi) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^{F(\theta-\phi)} r^2 \sin\theta \, dr \, d\theta.$$

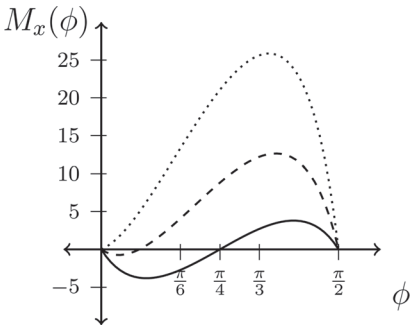


Figure 8 Moment curves for three different diamond shaped figures. The solid curve is the moment of the right half of a square diamond with diagonals of four units each about the x-axis as it rotates counterclockwise from 0 to $\pi/2$. The moment starts at 0, becomes negative, returns to 0 at $\pi/4$, becomes positive and returns to 0 at $\pi/2$. The dashed line is the moment curve of a diamond with vertical diagonal of 4 units and horizontal diagonal of 5 units. The dotted line is the moment curve of a diamond with vertical diagonal of 4 units and horizontal diagonal of 6 units.

In Figure 8, we see plots of M_x as a function of $\phi \in [0, \pi/2]$ for three different diamond shaped figures. The solid curve is for a diamond with $b = 4$ and $c = 4$ (i.e. a square). We see zeros at $\phi = 0, \pi/4, \pi/2$ which all correspond to lines of symmetry of the square.

The dashed line corresponds to a figure with $b = 4$ and $c = 5$ The zero between 0 and $\pi/6$ represents a CVT orientation, but the boundary of the CVT does not lie on a line of symmetry of the figure. The lines of symmetry of this figure correspond only to rotations of $\phi = 0$ and $\phi = \pi/2$. This is a counterexample to the original conjecture. Without this analysis, it is not at all obvious that this "extra" CVT is present at an angle of rotation of around 14.3°.

Interestingly, the dotted curve is for a diamond with $b = 4$ and $c = 6$. In this case, there are no intermediate roots and again CVTs only occur at lines of symmetry of the figure. We note that even though this third example is not ellipse-like, the conjecture holds.

The conclusion of Theorem 5 clearly extends to figures that are not ellipse-like. Stretching the diamond figure in Figure 7 further along the x-axis causes it to eventually satisfy the original conjecture. Shrinking it along the x-axis, towards a square configuration also causes it to satisfy the original conjecture. This result was born of a conjecture that seemed reasonable, but was false. We have presented a class of figures, ellipse-like, for which the conjecture is true. More work is necessary to fully characterize the class of figures for which the conclusion of Theorem 5 holds.

REFERENCES

[1] Aurenhammer, F. (1991). Centroidal Voronoi tessellations: Applications and algorithms. *CSUR ACM Comput. Surv. ACM Comput. Surv*. 23(3): 345–405. doi.org/10.1145/116873.116880.

[2] Du, Q., Faber, V., Gunzburger, M. (1999). Centroidal Voronoi tessellations: Applications and algorithms. *SIAM Rev. SIAM Rev*. 41(4): 637–676. doi.org/10.1137/s0036144599352836.

[3] Gordon, R. A. (2002). *Real Analysis: A First Course*. Boston, MA: Addison Wesley.

[4] Hamerly, G. and Elkan, C. (2002). Alternatives to the k-means algorithm that find better clusterings. *Proceedings of the Eleventh International Conference on Information and Knowledge Management - CIKM '02*. doi.org/10.1145/584792.584890.

[5] Jennings, G. A. (1994). *Modern Geometry with Applications*. New York: Springer.

[6] Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Trans. Inform. Theory IEEE Trans. Inf. Theory*. 28(2): 129–137. doi.org/10.1109/tit.1982.1056489.
[7] MacQueen, J. (1967). *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. Berkeley: University of California Press, pp. 281–297. http://projecteuclid.org/euclid.bsmsp/120051299.
[8] Okabe, A., Boots, B., Sugihara, K., Chiu, S. (2009). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd ed. Tokyo: John Wiley & Sons.
[9] Spivak, M. (1965). *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*. New York: W. A. Benjamin.
[10] Strauss, D. J., Hartigan, J. A. (1975). Clustering algorithms. *Biometrics*. 31(3): 793. doi.org/10.2307/2529577.
[11] Tiel, Jan van. (1984). *Convex Analysis: An Introductory Text*. Chichester: Wiley.
[12] Parker, A. E. (2013). Who solved the Bernoulli differential equation and how did they do it? *Coll. Math. J.* 44(2): 89. doi.org/10.4169/college.math.j.44.2.089.
[13] Hopkins, B. (2014). *Resources for Teaching Discrete Mathematics Classroom Projects, History Modules, and Articles*. Washington, DC: Mathematical Association of America.

Summary. We give a brief review of 2D, Euclidean, centroidal Voronoi tessellations (CVTs). Dropping to the specific case of two-point CVTs, a rich interaction between calculus, geometry and analysis is uncovered. Narrowing further to "ellipse-like" shapes with order 2 rotational symmetry, a full characterization of possible two-point CVTs is given for these figures. One might expect the CVTs of symmetric figures with even rotational symmetry to always correspond to lines of symmetry of the figure. Surprisingly, this conjecture is false. We close the article showing how to generate counterexamples to this conjecture and provide some direction for further inquiry.

KATHRIN GILLESPIE earned her Bachelor of Arts in mathematics at Whitman College. She is now working as a product manager at Redfin, a real estate technology company based in Seattle, Washington, leading a team developing software for consumers to find homes online.

SOPHIA KOTOK earned her Bachelor of Arts in mathematics at Whitman College. She is currently working toward a Ph.D. in mathematical modeling at Rochester Institute of Technology, where her research relates to transfer learning and optimization for deep neural networks.

ALBERT W. SCHUELLER (MR Author ID: 632785) is a mathematics professor at Whitman College in Walla Walla, WA. He is interested in centroids of all kinds, open educational resources, philosophy of mathematics, and computer generated art. He knows that everyone can do and enjoy mathematical exploration. And, that mathematics is the ultimate environmentally sustainable activity.

Agnesi's Problem III . . . Could the Textbook Be Wrong?

ANTONELLA CUPILLARI
Pennsylvania State University Erie
Erie, PA 16563
axc5@psu.edu

SAAMAN D. KHALILOLLAHI
Pennsylvania State University
State College, PA 16801
sdk5176@psu.edu

While looking for a geometry problem accessible to students in pre-calculus, but still challenging enough for seniors with a science background, and whose solution could not be easily found on the internet, we stumbled onto an old problem posed by Maria Gaetana Agnesi (1718–1799) [3, 5] in her *Instituzioni Analitiche* [1]. The plan was to use a special case of the problem to make students aware of the 300 year anniversary of Agnesi's birth (since it was 2018) and her role in the development of calculus textbooks. Her opus, published in 1748, was designed to be an extensive textbook for the Italian youth, written in Italian, and for readers already familiar with Euclid's *Elements*. She tackled much of the mathematics known at that time, from basic arithmetic and geometry to calculus, and her work was later translated into English by the Reverend John Colson [2].

The problem presented here is from the precalculus section of the book specifically from "VOLUME FIRST - BOOK I. The Analysis of Finite Quantities. Sect. III. Of the Construction of *Loci*, and the indeterminate Problems, which do not exceed the second degree" [1, p. 201]. The following is an almost literal translation of some pages from the *Instituzioni*. Reverend John Colson's translation [2] is very similar to the one included here and to Agnesi's work, and it has no additional commentary. (Some of the formatting has been changed to improve readability).

Problem III.

134. *Given the two circles EGF, BNO, and given their centers C, A; (Fig. 72.) if from any point G on the periphery of the circle EGF one draws the tangent GNO that meets the other circle BNO in the points N, O, and from these two points draws two tangents NM, OM, find the locus of all the points M in which said tangents meet.*

From a point M that is one of those required, draw the perpendicular MP to CA, and from the center A draw the line AM; since the triangles ANM, AOM are equal because of the right angles N, O, and the sides AN, NM equal to the sides AO, OM; it will also make the angle $NMA = OMA$ in the triangles NMQ, OMQ, because the side MQ is common, and $MO = MN$, and it will make $NQ = QO$, and AM perpendicular to NO. Draw the line CG from the center C to the point of contact, and it will be parallel to AM because it is also perpendicular to NO; and set $AB = a$, $CE = b$, $CA = c$, and $AP = x$, $PM = y$, and thus $AM = \sqrt{xx + yy}$. In the similar triangles AOM, AQO it will be AM, $AO :: AO$, AQ, and substituting the analytical

Math. Mag. **95** (2022) 99–109. doi:10.1080/0025570X.2022.2023301 © Mathematical Association of America
MSC: 01-01A50

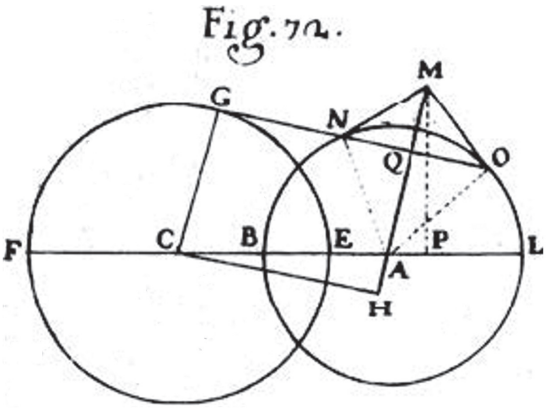


Figure 1 Fig. 72 as it appears in Agnesi's *Instituzioni Analitche* [1] in Table XIV.

values one will find

$$AQ = \frac{aa}{\sqrt{xx + yy}}.$$

Draw CH perpendicular to MA, lengthened if needed, to make $HQ = CG$, and then

$$HA = b - \frac{aa}{\sqrt{xx + yy}};$$

but the triangles CAH, AMP are similar, and thus PA, $AM :: AH$, AC, that is

$$x, \sqrt{xx + yy} :: b - \frac{aa}{\sqrt{xx + yy}}, c.$$

and by multiplying the outside and the middle terms,

$$cx = b\sqrt{xx + yy} - aa.$$

that is

$$cx + aa = b\sqrt{xx + yy},$$

and squaring

$$ccxx + 2aacx + a^4 = bbxx + bbyy,$$

or

$$yy + \frac{\overline{bb - cc} \times xx}{bb} - \frac{2aacx}{bb} - \frac{a^4}{bb} = 0.$$

Three cases of this equation have to be separated: when $b = c$, when b is larger than c, and when c is larger than b.

Agnesi then constructed the equations of the parabola when $b = c$, of the ellipse when b is larger than c, and the hyperbola when c is larger than b. The standard equations of the three conics in "modern" notation are

1. $y^2 = \dfrac{2a^2}{b}x + \dfrac{a^4}{b^2}$ when $b = c$,

2. $\dfrac{\left(x - \frac{a^2c}{b^2-c^2}\right)^2}{\left(\frac{a^2b}{b^2-c^2}\right)^2} + \dfrac{y^2}{\left(\frac{a^2}{\sqrt{b^2-c^2}}\right)^2} = 1$ when $b > c$,

3. $\dfrac{\left(x + \frac{a^2c}{c^2-b^2}\right)^2}{\left(\frac{a^2b}{c^2-b^2}\right)^2} - \dfrac{y^2}{\left(\frac{a^2}{\sqrt{c^2-b^2}}\right)^2} = 1$ when $b < c$.

After some more calculations to show how to graph the conics according to techniques that had been introduced earlier in the book, the presentation of the problem is concluded with the following remark:

> *In this problem it was always assumed that the circle EFG is larger than the circle BNO, that is b is larger than a, but even if it was $b = a$, and a larger than b, the locus of the required points would always be a parabola in the first case, an ellipse in the second, and two opposite hyperbolas in the third, and it is therefore useless to distinguish these cases, which will not change the loci.*

The formulae of the three conics depend only on the radii of the circles and the distance between their centers. Agnesi placed the center of the circle to which the secants are drawn at the origin of the axis, so these three numbers completely describe the geometry of the problem and the conclusion seems reasonable. To gain a better understanding of the problem an example is in order.

Let the equations of the circles be $(x + 3)^2 + y^2 = 25$ and $x^2 + y^2 = 16$, with $a = 4$, $b = 5$, and $c = 3$ (distance between the two centers). Thus, according to Agnesi's results, one expects the locus studied in the problem to be the ellipse described by the equation

$$\frac{\overline{bb - cc} \times xx}{bb} - \frac{2aacx}{bb} = \frac{a^4}{bb} - yy,$$

that is

$$\frac{\overline{5^2 - 3^2} \times xx}{5^2} - \frac{2(16)3x}{5^2} = \frac{4^4}{5^2} - yy.$$

The standard equation is

$$\frac{(x-3)^2}{25} + \frac{y^2}{16} = 1,$$

and the ellipse has center $(3, 0)$, vertices $(-2, 0)$ and (8.0), covertices $(3, 4)$ and $(3, -4)$. A graph of this problem seems to suggest a different result (see Figure 2,

produced using the software package GeoGebra. To see an animation of this geometric configuration, visit Agnesi Circle Problem - Ellipse*, take hold of the point G and move it on the circle to see the point M move on the ellipse. The animation can also be run automatically by pressing the play button in the lower left of the interface).

Something is peculiar. This graph shows as the wanted locus only part of the ellipse, starting and ending at the points of intersection of the circle on the right with the ellipse, points generated by the lines tangent to both circles. Once the point G moves further to the left, the tangent line to the left circle will not touch the right circle and there will be no more points satisfying the construction described in the problem. Could the textbook be so wrong? There are other mistakes in the *Instituzioni* [**4**], as in most textbooks, but since Agnesi was a well-respected mathematician [**3**] it is hard to believe that she could have obtained such an incorrect solution. At this point the problem requires more investigation! If Agnesi's solution is indeed wrong, where is the mistake? What is the correct solution?

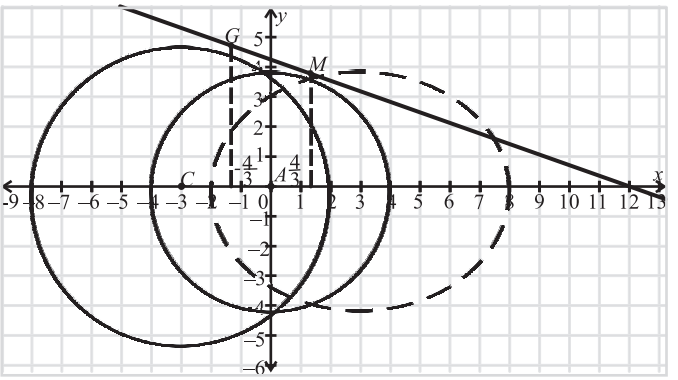


Figure 2 Illustration of Agnesi's problem using the circles $(x+3)^2+y^2=25$ and $x^2+y^2=16$.

Working on the problem in a different way

Agnesi used geometry and algebra to solve the problem since this is included in the precalculus part of her book. What happens if one uses analytic geometry and a more "modern" approach?

By symmetry, it is sufficient to construct the solution curve in the upper half-plane. Referring to Agnesi's illustration, using her choice of the variables, and placing the origin of the axes at the point A, let $AB=a$, $CE=b$, $CA=c$, and $AP=x$, $PM=y$, so that $AM=\sqrt{xx+yy}$ and the equations of the two circles are $(x+c)^2+y^2=b^2$ and $x^2+y^2=a^2$. See Figure 1.

Let the point G have coordinates (x_G, y_G) . The slope of the line tangent to the circle at G, which can also easily be found without using derivatives, is

$$y'_G=-\frac{(x_G+c)}{\sqrt{b^2-(x_G+c)^2}}=-\frac{x_G+c}{y_G}$$

for $y_G\neq 0$, and the equation of the tangent line is

$$y=y'_G x+\frac{y_G^2+(x_G+c)x_G}{y_G}.$$

*Online at http://ggbm.at/H6TXQdY6.

If $y_G = 0$, the tangent line is vertical, and the point M will be on the x-axis. Set

$$C_G = \frac{y_G^2 + (x_G + C)x_G}{y_G}$$

so that the equation can be written as $y = y'_G x + C_G$. To determine the points N and O, determine the intersection of this line and the circle with radius $AB = a$ by solving the quadratic equation

$$x^2 + (y'_G x + C_G)^2 = a^2,$$

which can be rewritten as

$$(1 + (y'_G)^2)x^2 + 2y'_G C_G x + (C_G^2 - a^2) = 0.$$

After some simplification, using the fact that $(x_G + c)^2 + y_G^2 = b^2$, the equation becomes

$$b^2x^2 - 2(b^2x_G + cy_G^2)x + y_G^2(C_G^2 - a^2) = 0.$$

The quadratic formula now yields

$$x = \frac{(b^2x_G + cy_G^2) \pm \sqrt{(b^2x_G + cy_G^2)^2 - b^2y_G^2(C_G^2 - a^2)}}{b^2}.$$

Let's notice that this equation can be solved only when

$$(b^2x_G + cy_G^2)^2 - b^2y_G^2(C_G^2 - a^2) \geq 0,$$

generating some bounds for the x-coordinate of the point G. Solving this second degree inequality provides a good algebra workout, but the result is represented by the rather simple conditions

$$\frac{(b^2 - c^2) - ab}{c} \leq x_G \leq \frac{(b^2 - c^2) + ab}{c}.$$

Since the point G is on the circle $(x + c)^2 + y^2 = b^2$, we also have the bounds

$$-c - b \leq x_G \leq -c + b.$$

When

$$(b^2x_G + cy_G^2)^2 - b^2y_G^2(C_G^2 - a^2) = 0,$$

the points N and O coincide (their x-coordinate is $(b^2x_G + cy_G^2)/b^2$), the tangent to G is also tangent to the other circle, and the point M is on the circle. Thus the problem can be solved when

$$x_G \in \left[\frac{(b^2 - c^2) - ab}{c}, \frac{(b^2 - c^2) + ab}{c}\right] \cap [-c - b, -c + b].$$

Agnesi does not mention any restrictions on the choice of the point G. This is unusual because in other cases she discusses in detail the conditions for the existence of real solutions, and she is generally very meticulous.

When

$$(b^2x_G + cy_G^2)^2 - b^2y_G^2(C_G^2 - a^2) > 0,$$

the two solutions are

$$x_N = \frac{(b^2x_G + cy_G^2) - \sqrt{(b^2x_G + cy_G^2)^2 - b^2y_G^2(C_G^2 - a^2)}}{b^2}$$

and

$$x_0 = \frac{(b^2x_G + cy_G^2) + \sqrt{(b^2x_G + cy_G^2)^2 - b^2y_G^2(C_G^2 - a^2)}}{b^2}.$$

The points N and O are on the circle $x^2 + y^2 = a^2$, so their y-coordinates are

$$y_N = \sqrt{a^2 - x_N^2} \quad \text{and} \quad y_O = \text{sign}(x_{Int} - a)\sqrt{a^2 - x_0^2}$$

respectively, when $x_G > -c$ (i.e. G is between the two centers). The formula for y_o is slightly more complicated because this coordinate will be negative when the intersection of the line GO with the x-axis, with x-coordinate

$$x_{Int} = \frac{y_G^2 + (x_G + c)x_G}{x_G + c},$$

is inside the circle with radius a. On the other hand, when $x_G < -c$, we will have instead

$$y_N = \text{sign}(-x_{Int} - a)\sqrt{a^2 - x_N^2} \quad \text{and} \quad y_o = \sqrt{a^2 - x_0^2}$$

for the same reason.

The equations of the lines tangent to the circle at these points are

$$y = -\frac{x_N}{y_N}x + \frac{a^2}{y_N} \quad \text{and} \quad y = -\frac{x_o}{y_o}x + \frac{a^2}{y_o}$$

when $y_N \neq 0$ and $y_o \neq 0$. To find the intersection of these lines, and thus the coordinates of the point M, we solve the equation

$$-\frac{x_N}{y_N}x + \frac{a^2}{y_N} = -\frac{x_o}{y_o}x + \frac{a^2}{y_o}$$

to obtain

$$x_M = a^2\frac{y_N - y_o}{x_o y_N - x_N y_o} \quad \text{and} \quad y_M = -\frac{x_N}{y_N}x_M + \frac{a^2}{y_N}.$$

The problem is solved!

Although the coordinates of M are functions of the coordinates of G, the dependence is hidden by the complexity of the formulas. The analytic solution, while gaining in detail, is much less elegant than Agnesi's geometry-based work.

As a side comment, a solution that shows more explicitly the connection between M and G can be found by restating Agnesi's problem as follows:

> *Given the two circles EGF, BNO, and given their centers C, A; (see Fig. 1) if from any point G on the periphery of the circle EGF one draws the tangent GNO that meets the other circle BNO in the points N, O, and from one of these two points draws the tangent NM to the circle, find the locus of all the points M in which said tangents meet the line through the center A parallel to the radius CG.*

Determine the coordinates of the point O (or N) as done above. The equation of the line tangent to the circle at O has equation

$$y = -\frac{x_0}{y_0}x + \frac{a^2}{y_0}.$$

The slope of the line CG is $y_G/(x_G + c)$, so the equation of the line through A and parallel to this is $y = y_G/(x_G + c)x$. The x-coordinate of the point M at the intersection of these two lines is

$$x_M = \frac{a^2(x_G + c)}{y_O y_G + x_O(x_G + c)}.$$

Thus, $y_M = y_G/(x_G + c)x_M$.

Let us go back and verify the analytic work in the case of the circles $(x + 3)^2 + y^2 = 25$ and $x^2 + y^2 = 16$, with $a = 4$, $b = 5$, and $c = 3$.

According to the restriction found for the x-coordinate of the point G, it will be possible to construct the point M only when

$$x_G \in \left[-\frac{4}{3}, 12\right] \cap [-8, 2] = \left[-\frac{4}{3}, 2\right].$$

When G is the point with coordinates $\left(-\frac{4}{3}, \frac{10\sqrt{2}}{3}\right)$, the points N and O coincide, and they have coordinates $\left(\frac{4}{3}, \frac{8\sqrt{2}}{3}\right)$. A spreadsheet calculation using the formulas above yields the following points M for

$$x_G \in \left[-\frac{4}{3}, 2\right]$$

and $y_G > 0$, building only a section of the top part of the ellipse.

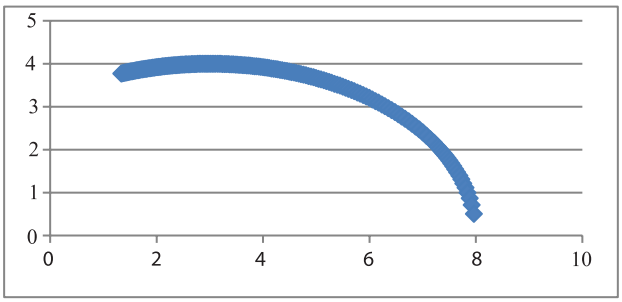


Figure 3 Graph of the solution set given the circles $(x + 3)^2 + y^2 = 25$ and $x^2 + y^2 = 16$.

Why did Agnesi state that one would obtain a full ellipse? Geometrically, the point M must be either on the circle on the right or outside of it, making it possible to construct tangent lines from M to the points N and O and vice versa. However, this construction is not possible once the points of the ellipse are inside the circle.

Asking more questions (just because . . .)

What about the other two conics? Do they fare any better than the ellipse? How many cases would have to be listed to have all possible relations between a, b, and c and thus also check Agnesi's closing claim that the relative size of the circles does not matter? Consider the following list of all possible cases.

$$a < b \begin{cases} b > c \begin{cases} a > c & (1) \\ a < c & (2) \\ a = c & (3) \end{cases} \\ b = c \quad a < c \quad (4) \\ b < c \quad a < c \quad (5) \end{cases} \quad a = b \begin{cases} b = c \quad a = c \quad (6) \\ b > c \quad a > c \quad (7) \\ b < c \quad a < c \quad (8) \end{cases} \quad a > b \begin{cases} b = c \quad a > c \quad (9) \\ b > c \quad a > c \quad (10) \\ b < c \begin{cases} a = c & (11) \\ a > c & (12) \\ a < c & (13) \end{cases} \end{cases}$$

Do all of these cases correspond to a geometrically solvable problem? Remember the restriction

$$x_G \in \left[\frac{(b^2 - c^2) - ab}{c}, \frac{(b^2 - c^2) + ab}{c} \right] \cap [-c - b, -c + b].$$

For a locus to exist at all, these two intervals cannot be disjoint. This will occur when $b - a < c$ because it will make

$$\frac{(b^2 - c^2) + ab}{c} < -c + b.$$

Geometrically, when $b \geq a + c$ the circle on the right is inside the circle on the left, so it is not possible for the tangent line to the larger circle to intersect the circle inside. Given that $c > 0$, the condition $b - a < c$ is always true when $a \geq b$ and $0 \geq b - a$ (cases (6) to (13) in the list above), and if $b \leq c$, since $a > 0$, then $b - a < c$ (cases (4) and (5)). The other three cases are problematic. For example, in case (2), if $a = 0.75$, $b = 3$, and $c = 2$, it is not possible to solve the problem because the circle on the right is nested inside the other one. But when $a = 2$, $b = 3$, and $c = 2.5$ it is possible to find a locus that satisfies the conditions in the problem and this locus is a part of an ellipse.

At this point it is also possible to see that Agnesi's claim that it is not important which circle is larger is also incorrect. While the problem cannot be solved when $a = 0.75$, $b = 3$, and $c = 2$, a solution exists when $a = 3$, $b = 0.75$, and $c = 2$.

The ellipse is not the only curve to be only partially drawn in some cases. When $b = c$, the required locus might only be part of the parabola with equation

$$y^2 = \frac{2a^2}{b}x + \frac{a^4}{b^2},$$

vertex $V\left(-\frac{a^2}{2b}, 0\right)$, as it happens when $a = b = c = 5$, where the solution exists only for $x \geq 0$, corresponding to $x_g \in [-5, 0]$.

But things get even more interesting! If $a = 6$, $b = c = 2$, the locus described by Agnesi is the whole parabola $y^2 = 36x + 324$, i.e., $x = \frac{1}{36}y^2 - 9$. (See Figure 4).[†]

A necessary condition for obtaining a complete conic is to be able to use the whole circle on the right side, having x_G span the whole interval $[-c - b, -c + b]$. Thus, it

[†]For an animation see Agnesi's Circle Problem, online at Agnesi's Circle Problem, online at http://gbm.at/VCQ2h35Yggbm.at/VCQ2h35Y moving the point M or pressing the play button as previously described.

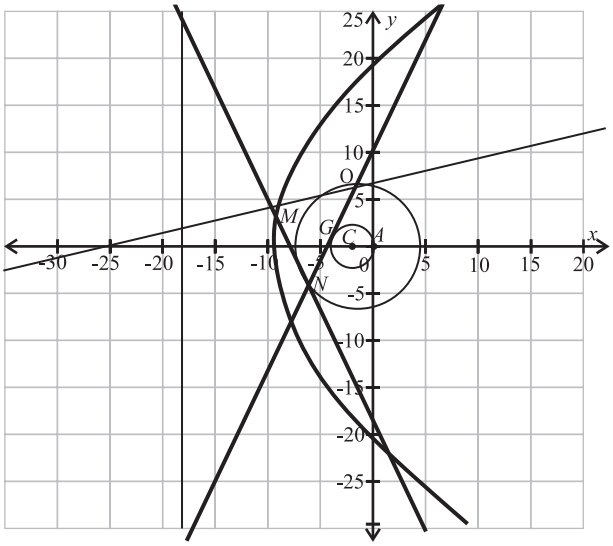


Figure 4 Illustration of Agnesi's problem using the circles $(x+2)^2+y^2=4$ and $x^2+y^2=36$.

is necessary to have

$$[-c-b,-c+b]\subseteq\left[\frac{(b^2-c^2)-ab}{c},\frac{(b^2-c^2)+ab}{c}\right].$$

This means that the desired locus is the entire conic when $-(a-b)<c<a-b$. Since $c>0$, this is a possible case only if $a>b$. In the other cases, the best result will be a part of a conic. For an additional animation that shows the incomplete construction of a hyperbola, visit http://ggbm.at/FGRmQEg4.

Can the problem be reversed?

Another issue to consider is whether the process presented in Problem III can be reversed. Since the procedure to construct M when G is given generates a one-to-one correspondence for appropriate values of x_G, this question seems reasonable. Given a point M (not inside the circle of radius a) in the upper half-plane, draw the tangent lines from it to the circle and find the line that passes through the points of tangency. Choose the point G determined by such a line when the line itself is tangent to the circle with equation $x^2+y^2=a^2$. Since the line through points on the circle (x,y) and $M(x_M,y_M)$ must be tangent to the circle, its slope can be found in two ways to produce the equation

$$\frac{y-y_M}{x-x_M}=-\frac{x}{y}.$$

Using the fact that the point is on the circle, this equation can be rewritten as

$$\frac{\sqrt{a^2-x^2}-y_M}{x-x_M}=-\frac{x}{\sqrt{a^2-x^2}}$$

implying that

$$x^2(x_M^2+y_M^2)-2a^2x_Mx+a^2(a^2-y_M^2)=0.$$

The solution(s) are

$$x = \frac{a^2 x_M \pm a\sqrt{y_M^2(x_M^2 + y_M^2 - a^2)}}{x_M^2 + y_M^2}.$$

The real solutions exist only when $x_M^2 + y_M^2 \geq a^2$, so the point M must be on or outside the circle Agnesi placed on the right. This highlights again the problem with the findings in the *Instituzioni*, when points inside the circle on the right are used. So, the two points O and N have coordinates

$$\left(x_O = \frac{a^2 x_M + a\sqrt{y_M^2(x_M^2 + y_M^2 - a^2)}}{x_M^2 + y_M^2}, \sqrt{a^2 - x_O^2}\right)$$

and

$$\left(x_N = \frac{a^2 x_M - a\sqrt{y_M^2(x_M^2 + y_M^2 - a^2)}}{x_M^2 + y_M^2}, \sqrt{a^2 - x_N^2}\right).$$

The slope of the line through O and N is

$$S = -\frac{2a^2 x_M}{(x_M^2 + y_M^2)\left(\sqrt{a^2 - x_O^2} + \sqrt{a^2 - x_N^2}\right)}.$$

The line is tangent to the other circle at the point G where $S = -(x_G + c)/y_G$. Solving this equation yields

$$x_G = -c \pm b\sqrt{\frac{S^2}{S^2 + 1}}.$$

Since the point G is on the circle, its y-coordinate is

$$y_G = \sqrt{b^2 - (x_G + c)^2} = \frac{b}{\sqrt{S^2 + 1}}.$$

When the point M is in the right half plane and $x_M > 0$, $x_G + c > 0$, so

$$x_G = -c + b\sqrt{\frac{S^2}{S^2 + 1}},$$

when M is in the left half plane, $x_G + c < 0$, so

$$x_G = -c - b\sqrt{\frac{S^2}{S^2 + 1}}.$$

Conclusion

Agnesi's construction is beautiful ... when it works. It is essential to be able to construct the triangle CAH for her procedure to work. And even under this restriction, there is no guarantee that the required locus is a whole conic. It is possible that Agnesi had specific examples in mind, and then over generalized her result. So, yes, a textbook can be wrong! Writers do make mistakes, but somebody else usually checks their

work. In Agnesi's case, she was the editor and the supervisor of the printing process [**3**] of her opus. On the other hand, Reverend Colson had worked on Newton's material, so he knew well the fundamentals of the mathematics of his time [**2**]. Did he limit himself to just being a translator of Agnesi's work? A mistake, however, does not mean that there is nothing to learn; sometimes mistakes can teach us a lot. Firstly, in order to find a mistake one must be an active reader who works along with the writer, checking the writer's work. Mistakes that are found must then be investigated in order to determine what went wrong. Lastly, readers can explore the problem independently, asking a lot of "what if?"-style questions and trying to answer them.

REFERENCES

[1] Agnesi, M. G. (1748), *Instituzioni Analitiche ad Uso della Gioventu' Italiana.* Milano.
[2] Colson, J. (1801). *Analytical Instituions in Four Books: Originally Written in Italian by Donna Maria Gaet Agnesi.* Published under the inspection of the Rev. John Hellins in London in 1801.
[3] Cupillari, A. (2007). *A Biography of Maria Gaetana Agnesi, An Eighteenth-Century Woman Mathematician: With Translations of Some of Her Work from Italian into English.* New York: Edwin Mellen Press.
[4] Cupillari, A. (2014). Maria Gaetana Agnesi's Other Curves (More than just the Witch). *Math. Mag.* 87(1): 3–13. doi.org/10.4169/math.mag/87.1.3
[5] Mazzotti, M. (2007). *The World of Maria Gaetana Agnesi, Mathematician of God.* Baltimore, MD: Johns Hopkins University Press.

Summary. In the pre-calculus portion of her book *Instituzioni Analitiche* (1748), Maria Gaetana Agnesi presents the following problem and its solution: "Given two circles EGF, BNO, and given their centers C, A; if from any point G on the periphery of the circle EGF one draws the tangent GNO that meets the other circle BNO in the points N, O, and from these two points draw two tangents NM, OM, find the locus of all the points M in which said tangents meet." The problem's solution is complicated by the need to consider several cases, and Agnesi's work contained errors. We discuss her work and present an illustrated and complete solution. In doing so, we hope not only to introduce a problem of geometric interest, but also to contribute to the historical understanding of Agnesi's work.

ANTONELLA CUPILLARI is an associate professor emerita of mathematics after retiring from Penn State Erie–The Behrend College in 2021. She is the author of the books *The Nuts and Bolts of Proofs* and *A Biography of Maria Gaetana Agnesi, an Eighteenth-Century Woman Mathematician: With Translations of Some of Her Work from Italian into English* and of several articles about classroom ideas and the history of mathematics, her field of research. She is still working on the 5th edition of her textbook and studying Agnesi's *Instituzioni*.

SAAMAN D. KHALILOLLAHI graduated from Penn State Erie–The Behrend College in 2021, earning concurrent degrees in computer science and mathematics. As an undergraduate he was involved in several research projects in robotics, industrial engineering, data analysis and mathematics, and was the programming mentor for the Harbor Creek Robotics Club from Harbor Creek Senior High, his alma mater. He is now enrolled in the MS Program in Computer Science and Engineering at Penn. State University—University Park.

Three Persons, Two Cuts: A New Cake-Cutting Algorithm

STEVEN J. BRAMS
New York University
New York, NY 10012
steven.brams@nyu.edu

PETER S. LANDWEBER
Rutgers University
Piscataway, NJ 08854
landwebe@math.rutgers.edu

Everyone knows the cake-cutting procedure, "I cut, you choose": Person A (assumed male) divides the cake, which may be any heterogeneous divisible good, into two equal parts—in terms of his preferences—and person B (assumed female) chooses the piece that she thinks is at least as valuable as the other piece. Thereby each person believes he or she received a piece at least as valuable as the piece obtained by the other person. This division is envy-free, because A and B do not envy the other person for having obtained a more valuable piece.

But can this procedure be extended to more than two persons? John Selfridge and John Conway, circa 1960, independently proposed an extension for dividing a cake into envy-free portions for three persons, whom we will refer to as players. It was never published but is described in Brams and Taylor [**8**, pp. 116–120]. Instead of using the minimal two cuts to divide a cake into three pieces, however, it requires up to five cuts. Moreover, the portions that players A, B, and C receive may not be connected (or contiguous)—one or more of them may get pieces from different parts of the cake (e.g., both ends) rather than a single connected piece.

Twenty years later Stromquist proposed an algorithm that gives players connected pieces, using two cuts for three persons, but it was not discrete. It used four simultaneously moving knives, requiring A, B, or C to call "stop" when their moving knives and the moving knife of a referee create a tie between different pieces of the cake [**19**].

Some 25 years after this discovery, Barbanel and Brams [**4**] found a much simpler three-person moving-knife algorithm that required a player, or a player and a referee, to move two parallel knives simultaneously. By "squeezing" a piece desired by two players until it creates a tie with another piece for one of the players, it enables all three players to obtain at least a tied-for-most-valuable piece.

They also showed how this algorithm could be extended to four players, but instead of using only the minimal three cuts, it required up to five. This is fewer than the 11 cuts that an earlier four-person moving-knife procedure of Brams, Taylor, and Zwicker required [**9**]. Besides three-person and four-person algorithms that use relatively few cuts, envy-free algorithms that work for any number n of players have been developed. These include an n-person moving-knife algorithm [**7**] which uses a finite number of cuts, but this number is unbounded—an upper bound cannot be specified independently of the preferences of the players—and an algorithm based on Sperner's lemma [**21**], which requires convergence to an exact division so also is unbounded.

Aziz and Mackenzie [**2**] provided an algorithm that has two advantages: It does not depend on continuously moving knives or convergence, and it is finite and bounded. Its

Math. Mag. **95** (2022) 110–122. doi:10.1080/0025570X.2022.2023300 © Mathematical Association of America
MSC: 91A05, 92A80

disadvantage is that its bounds are presently extremely large, making it, like Brams and Taylor's algorithm, computationally complex. In particular, it does not run in polynomial time. Recently, it was modified and extended to chore division [12], and a simplified four-person discrete cake-cutting algorithm was found [1]. In this paper, we return to the three-person cake-cutting problem but provide an envy-free algorithm that:

- does not require moving knives (unlike those of Brams and Barbanel [4], Brams, Taylor, and Zwicker [9], and Stromquist [19];
- requires only the minimal two cuts and so gives connected pieces (unlike that of Selfridge and Conway).

The algorithm is not computationally complex for three players, but it does require a number of steps, which we specify later, and certain restrictions on how players value a cake, in order to make it computationally feasible.

After describing this envy-free algorithm, we show how it can be extended to give an envy-free allocation that is

- *maximin envy-free*—it maximizes the minimum value that any player receives from two cuts.

We also show how to find an allocation, that is,

- *maximally equitable*—all players receive equally valued shares that are maximal.

Both kinds of allocations are efficient (Pareto-optimal): There is no other allocation in which all the players do at least as well and one or more does better.

In the case of envy-freeness, efficiency follows from the fact that any envy-free allocation that uses the minimal $n - 1$ cuts is efficient [8, pp. 149–151], [14], at least when the value functions are strictly positive; maximin envy-freeness additionally ensures that the worst-off player does as well as possible. In the case of maximal equitability, an allocation that gives all players equally valued shares that are as great as possible cannot give any player more without giving another player less, so it is also efficient, again when the value functions are strictly positive. Although it is well known that there exists an envy-free allocation that is also equitable [18], it need not be efficient. For more on the efficiency of envy-free allocations—but not algorithms to find them—see Barbanel [3].

Players submit value functions

We make the following assumptions:

1. The cake is defined by the interval [0, 1], and a division of the cake is a partition, indicated by points on the interval, into three connected subintervals that together comprise the entire cake.
2. Each player has a probability density function (pdf) over the cake, called a value function, whose measure is assumed to be nonatomic and absolutely continuous—that is, the pdf is continuous on [0, 1], takes only nonnegative values, is positive somewhere on each subinterval, and has (Riemann) integral equal to one.

Our algorithm differs from most of the aforementioned algorithms in that each player, independently of the others, does not place marks on the unit interval or move knives over it. Unlike the "squeezing procedure" of Barbanel and Brams [4] on which

it is modeled, it determines directly—without using moving knives—the two points on [0, 1] at which to cut the cake so that a tie occurs with one of the two enlarged pieces that player A, the squeezer, values equally. (The intermediate value theorem and continuity ensure that a tie must occur.)

It then assigns the tied enlarged piece to the player (say, B) for whom the squeezed piece ties; C gets the squeezed piece, which it still thinks is the most valuable; and A gets the remaining enlarged piece, which it thinks ties with the other enlarged piece and is preferable to the squeezed piece. In this way, all players receive at least tied-for-most valuable pieces, rendering the resulting allocation envy-free.

Instead of using moving knives, however, our algorithm uses a referee—to whom the players submit their value functions—who solves equations simultaneously to determine the cutpoint, when a player equally values two pieces, at which the moving-knife procedure would make a cut. Although such an algebraic solution obviates the need for the players to physically move knives, not all solutions to the integral equations we give can be expressed in closed form and may instead need to be approximated.

More specifically, the pdfs of Examples 1 and 2 in Section 4 are piecewise linear, which makes their integrals over subintervals quadratic and easy to evaluate, but for functions for which one cannot evaluate a definite integral in closed form, this will not be the case. Brânzei [**10**] and Kurokawa, Procaccia, and Wang [**17**] specify value functions that yield closed-form solutions, so they do not require convergence to an exact solution. Exact solutions for the examples in Section 4, expressed in terms of radicals, can be found using the Wolfram Equation Solver [**23**], but in Section 4 we give their decimal approximations.

An important advantage of an algebraic solution is that the referee can determine which envy-free solution is maximin. By contrast, one player's movement of knives cannot give this solution, because it requires the comparison of different players' valuations at different envy-free cutpoints.

Stromquist defined a "finite protocol" for cake-cutting to be one in which the players make finitely many marks in order to determine an envy-free allocation [**20**]. He showed that no finite protocol yields an envy-free allocation for three or more players. Given his definition, our algorithm is not finite, because it uses information on the continuous pdfs over infinitely many points to determine an envy-free allocation.

But our algorithm does not require infinitely many continuous choices by the players. Instead, after submitting their pdfs to a referee, he or she makes decisions—specifically, when there is a tie in the value of two pieces—that determines where cuts are to be made. The players, individually, cannot do this on their own.

Ianovski [**15**] argued that "a moving knife protocol is certainly less than an effective procedure in the algorithmic sense" because, as Dehghani et al. [**12**] pointed out, "the continuous movement of a knife cannot be captured by any finite protocol." This difficulty disappears, however, by introducing a referee, who uses the value functions of the players to find cutpoints (in a manner to be illustrated by the examples in the next section).

Thereby the players do not need to make decisions about the stoppage of continuously moving knives. The referee uses this information to do the rest, comparing the values that players attach to different pieces and indicating when they are equal. His or her decisions, using the algorithm described in the next section, are purely mechanical.

The envy-free algorithm specifies the steps—at most five—to process this information and determine an envy-free allocation. While the algorithm terminates after a finite number of steps. an additional step is needed to find a maximin envy-free allocation. A maximally equitable allocation may require that different sets of equations be solved simultaneously.

The algorithms that yield maximin envy-free and maximally equitable allocations provide exact solutions if the players' value functions are sufficiently simple (see Kurokawa, Lai, and Proaccia [**16**] and Kurokawa, Proaccia, and Wang [**17**]) without having physically to move knives and say at what point two pieces have the same value (as noted earlier, the referee does this by solving equations). Since the players do not need to make an infinite number of continuous choices over time, the algorithm offers an immediate calculable solution that precludes the need for moving knives.

Because we assume that the players submit their value functions independently to a referee, they cannot use information about the other players' functions to manipulate the algorithm to their advantage. But to ensure that the proper calculations are made, it would be advisable to specify a checking process for verifying their correctness.

Envy-free and equitable algorithms

Our envy-free algorithm is implemented in the five steps specified next, where the value functions (pdfs) of A, B, and C are, respectively, $v_A(x)$, $v_B(x)$, and $v_C(x)$. Later we will specify the additional steps needed to find a maximin envy-free and a maximally equitable allocation.

The latter allocations (maximin, envy-free, and maximally equitable) cannot be found by moving-knife procedures, which assume that players know only their own value functions, whereas the referee in our setup knows the value functions of all the players. Our algorithms, therefore, afford benefits that moving-knife procedures do not provide.

Envy-freeness Step 1. Players A, B, and C submit their pdfs over $[0, 1]$ to a referee, who determines the two cutpoints that each of A, B, and C would make to divide the cake into three equally valuable pieces of $1/3$ each: (a_1, a_2); (b_1, b_2); (c_1, c_2), where the first number in each ordered pair indicates the cutpoint closer to 0 and the second the cutpoint closer to 1. Assume that A's first cutpoint, a_1, is the one of the three closest to 0.

Step 2. A referee determines whether cutpoints at a_1 and a_2 yield an envy-free allocation, whereby A gets $[0, a_1]$ and B and C most value different pieces—either $[a_1, a_2]$ or $[a_2, 1]$. If yes, then B and C each get their more valued pieces, which concludes the algorithm; if no, then go to Step 3.

Step 3. If both B and C most value the middle piece, then go to Step 4; if both B and C most value the right piece, then go to Step 5. In each case, calculate the cutpoints a and b that yield an envy-free allocation, as given in Step 4 or Step 5.

To connect our algorithm with the squeezing procedure, we allude in Steps 4 and 5 to a piece (middle or right) that is being squeezed by moving knives. The squeezing stops when the value of this piece to a player ties with the player's value of another piece. By setting definite integrals equal to each other, we find the cutpoints, given by the limits of integration, that equalize the value of two pieces to players.

Step 4. *(middle piece squeezed by A's equally expanding pieces on the left and right).* Determine the upper and lower limits of integration, a and b, by solving simultaneously two pairs of integral equations, the first pair given by (1a) and the second pair given by (1b):

$$\int_0^a v_A(x)\,dx = \int_b^1 v_A(x)\,dx \quad \text{and} \quad \int_a^b v_B(x)\,dx = \int_0^a v_B(x)\,dx \tag{1a}$$

$$\int_0^a v_A(x)\,dx = \int_b^1 v_A(x)\,dx \quad \text{and} \quad \int_a^b v_B(x)\,dx = \int_b^1 v_B(x)\,dx. \tag{1b}$$

The equation to the left of "and" of (1a) and of (1b), which are the same, ensure that the value of the left and right pieces for A (the squeezer) are equal. The equations on the right of (1a) and (1b) ensure that the value of the middle piece (between a and b) for B equals

- its value of the left piece (between 0 and a), given by (1a); or
- its value of the right piece (between b and 1), given by (1b).

The referee chooses the piece—either on the left or on the right—for which the solution to the equations of (1a) or (1b) gives the larger value of $\int_a^b v_B(x)\,dx$. This is the piece that is the first to tie with the middle piece under the squeezing procedure, because it is reduced less than would be the other piece when it ties with the middle piece.

Next, determine the upper and lower limits of integration, a and b, by solving simultaneously the two sets of integral equations, analogous to (1a) and (1b), but in which v_C is substituted for v_B

$$\int_0^a v_A(x)\,dx = \int_b^1 v_A(x)\,dx \quad \text{and} \quad \int_a^b v_C(x)\,dx = \int_0^a v_C(x)\,dx \tag{2a}$$

$$\int_0^a v_A(x)\,dx = \int_b^1 v_A(x)\,dx \quad \text{and} \quad \int_a^b v_C(x)\,dx = \int_b^1 v_C(x)\,dx. \tag{2b}$$

The referee chooses the piece—either on the left or on the right—for which the solution to the equations of (2a) or (2b) for a and b gives the larger value of $\int_a^b v_C(x)\,dx$. This is the piece that is the first to tie with the middle piece under the squeezing procedure, because it is reduced less than would be the other piece when it ties with the middle piece.

Compare the larger values given by the solutions of (1a) and (1b), and by the solutions of (2a) and (2b), and choose that which is greater (in the case of a tie, either can be chosen). This determines whether B or C is the first to receive an enlarged piece—either on the left or the right—that ties with the diminished middle piece. Assume it is B (C) that receives the enlarged piece when there is a tie; then C (B) will receive the diminished middle piece, which it still thinks is the most valuable; and A will receive the remaining enlarged piece on the left or the right. This concludes the algorithm.

Step 5. *(right piece squeezed by A's equally expanding pieces on the left).* The setup of pairs of integral equations to find cutpoints when the squeezed right piece first equals one of the expanded left pieces is completely analogous to the setup of equations in Step 4, so we will not provide details here. (In the next section, however, we give all pairs of equations that were solved to find envy-free cutpoints in the two examples in this section.)

This concludes the envy-free portion of the algorithm.

Maximin envy freeness Steps 4 and 5 can together be described informally as follows: Players B or C watch A squeeze either the middle piece or the right piece, keeping the other two expanding pieces equal, until the squeezed piece ties with one of these expanding pieces. This squeezing process, however, does not ensure a maximin envy-free allocation.

To find such an allocation, A further squeezes the squeezed piece (middle or right piece), which created the tie and an envy-free allocation, creating a second tie for the player who was not squeezed out initially. In this case, the maximin allocation will generally be one between the first tie and the second tie, giving each player a different piece that maximizes the worst-off player's value (see equations (4a) and (4b) later for an example).

That there always exists such a maximin follows from the fact that the integrals giving the players' valuations depend continuously on the endpoints of a closed interval, so there must be a point at which the minimum that the players receive is maximal. This is found by increasing the share of the player who receives the least-valuable piece when there is a tie until it itself ties with the piece of one of the other two players.

When the players' value functions are not the simple piecewise linear functions as in the examples in the next section, there may be more than two ties, which is a complication we will not consider here. As we noted earlier, this maximin allocation cannot be found with moving knives alone, because the player being squeezed can only indicate when it finds two pieces equal—but not if there is another envy-free allocation that maximizes the minimum value that the players receive.

Maximal equitability Solve simultaneously the equations that equalize the players' shares, as they value them, by setting them equal to each other. If there is more than one solution, the referee chooses the one that maximizes the equal value that all players receive, which is efficient (Pareto-optimal).*

In Section 4, we show in both Examples 1 and 2 that the maximal equitable allocation need not be envy-free, much less maximally envy-free. These examples illustrate the possible conflict between maximal equitability and maximin envy-freeness, whose allocations give shares to each player that differ by several percentage points.

Two examples

Example 1 Assume that A and B have the following piecewise linear value functions that are symmetric and V-shaped:

$$v_A(x) = \begin{cases} -4x + 2 & \text{for } x \in \left[0, \frac{1}{2}\right] \\ 4x - 2 & \text{for } x \in \left[\frac{1}{2}, 1\right] \end{cases} \qquad v_B(x) = \begin{cases} \frac{-4x+3}{2} & \text{for } x \in \left[0, \frac{1}{2}\right] \\ \frac{4x-1}{2} & \text{for } x \in \left[\frac{1}{2}, 1\right] \end{cases}$$

Whereas both functions have maxima at $x = 0$ and $x = 1$ and a minimum at $x = 1/2$, A's function is steeper (higher maximum, lower minimum) than B's, as illustrated in Figure 1. In addition, suppose that a third player, C, has a uniform value function, $v_C(x) = 1$, for $x \in [0, 1]$.

Before applying the algorithm, it is apparent that an envy-free allocation of the cake will be one in which A gets a portion to the left of $1/2$, B a portion to the right of $1/2$ (A and B could be interchanged), and C the portion in the middle. In Figure 1, we give the maximin envy-free cutpoints, represented by the "knives" below the graph, that show that A and B receive equal-length pieces between 0 and 0.314, and between 0.686 and 1. If they were not equal, then the player whose portion is shorter would envy the player whose portion is longer. Moreover, the distance separating the cutpoints must

*As with maximin envy-free allocations, equitable allocation may not be calculable in closed form using only algebraic operations for some value functions [**11**]. The vulnerability of maximin equitability to strategizing by the players is discussed by Brams, Jones, and Klamler [**5**].

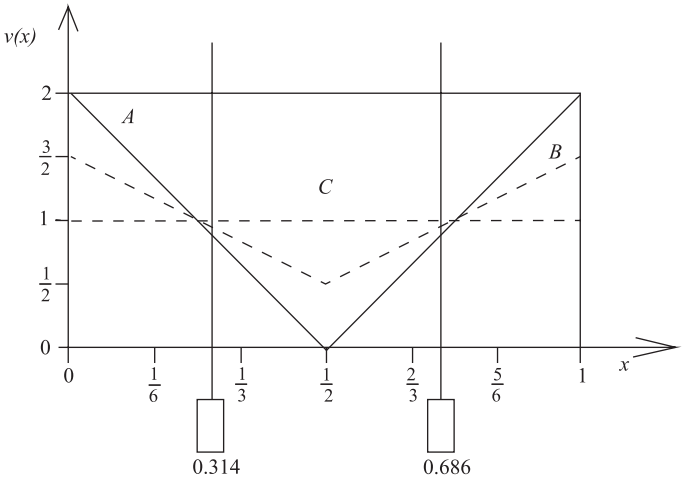


Figure 1 Value functions $v(x)$ of A, B, and C in Example 1 and maximin envy free cutpoints.

be at least $1/3$ (it is 0.372); otherwise, C will envy A and B for getting pieces that it considers worth more than $1/3$.

Both equations (1a) and (1b) are applicable to finding a and b when a tie occurs—for the first time with B—between the middle piece and the left and right pieces. Equation (3a) ensures that the cutpoints are equidistant from 0 and 1 as A squeezes equally from the left and right, and equation (3b) ensures that B's middle piece, whose value is given by the left side of (3b), ties with its right piece, whose value is given by the right side of (3b). Thereby, B can obtain this piece (it could also be the left piece because of the symmetry of the value functions about $x = 1/2$):

$$\int_0^a (-4x + 2)\, dx = \int_b^1 (4x - 2)\, dx \tag{3a}$$

$$\int_a^{1/2} \frac{-4x + 3}{2}\, dx + \int_{1/2}^b \frac{4x - 1}{2}\, dx = \int_0^a \frac{-4x + 3}{2}\, dx \tag{3b}$$

After integration and evaluation, the preceding equations become

$$-2a^2 + 2a = -2b^2 + 2b \quad \text{and} \quad 2a^2 - 3a = -b^2 + \frac{b}{2} - \frac{1}{2}$$

When solved simultaneously, they yield a feasible solution of $a = 0.271$ and $b = 0.729$. This gives A a share of 39.6% for the subinterval $[0, 0.271]$, B a share of 36.2% for the subinterval $[0.729, 1]$, and C a share of 45.8% for the subinterval $[0.271, 0.729]$.

In fact, however, we can give the worst-off player (B) more without creating envy by squeezing the middle piece further so that C's share (45.8%) decreases as B's share (36.2%) increases (A's share of 39.6% will also increase). Indeed, the middle piece can be reduced to $[1/3, 2/3]$ while preserving envy-freeness, which is when there is a tie for C between the diminished middle piece and the enlarged left and right pieces.

At these cutpoints, C's share is 33.3%, B's share 38.9%, and A's share 44.4%, which is the envy-free allocation farthest from being maximin. In effect, squeezing the middle piece until it ties for C with the left and right pieces carries the squeezing process too far to produce a maximin envy-free allocation, just as squeezing this piece until it ties for B is not far enough; the maximin envy-free allocation lies in between.

In Example 1, because it is B who receives the smallest share (36.2%) from the envy-free allocation, the referee needs to increase B's share. He or she does so by further squeezing the squeezed player's (C's) allocation until it ties with B's while preserving envy-freeness. More specifically, the referee ensures that the squeezer (A) equally values the enlarged pieces on the left and right, using Equation (4a), which is the same as Equation (3a), and ensuring that B's left share and C's middle share (in their respective measures) are equal, using Equation (4b)

$$\int_0^a (-4x+2)\,dx = \int_b^1 (4x-2)\,dx \tag{4a}$$

$$\int_0^a \frac{-4x+3}{2}\,dx = \int_a^{1/2} dx + \int_{1/2}^b dx \tag{4b}$$

After integration and evaluation, the preceding equations become

$$-2a^2+2a = -2b^2+2b \quad \text{and} \quad -a^2+\frac{5}{2}a = b$$

When solved simultaneously, they yield a feasible solution of $a = 0.314$ and $b = 0.686$. This gives A a share of 43.1% for the left subinterval [0, 0.314], B a share of 37.2% for the same-length right subinterval [0.686, 1], and C a share of 37.2% for the middle subinterval [0.314, 0.686].

This allocation is envy-free, because C's reduced middle share is still greater than what it perceives A and B receive on the two sides (31.4%). Moreover, A and B, which continue to receive the same-length subintervals, will not be envious of each other or value the middle piece more.

Because this allocation maximizes the minimum share that any player receives, it is the maximin envy-free allocation. Squeezing C's middle piece further would reduce C's share below 37.2% while increasing A's and B's shares, preventing the envy-free allocation from being maximin.

Recall that an allocation is maximally equitable when all players receive equally valued shares, as each player values its share, that are maximal. As shown in Brams, Jones, and Klamler [**5**], when A's share is set equal to C's, and C's share is set equal to B's (in each player's measure), this share in Example 1 is 39.3%. But this allocation is not envy-free because it requires that a be closer to 0 (at $x = 0.269$) than b is to 1 (at $x = 0.662$). The horizontal length of C's piece ($0.662 - 0.269 = 0.393$) is exactly the common share that all players receive.

Table 1 summarizes the subintervals and the percent shares that each player receives according to the different algorithms:

Allocations	A	B	C
Envy-free allocation	[0, 0.271] 39.6%	[0.729, 1] 36.2%	[0.271, 0.729] 45.8%
Envy-free allocation (maximin)	[0, 0.314] 43.1%	[0.686, 1] 37.2%	[0.314, 0.686] 37.2%
Equitable allocation	[0, 0.269] 39.3%	[0.662, 1] 39.3%	[0.269, 0.662] 39.3%

TABLE 1: Subintervals and percent shares to each player in Example 1.

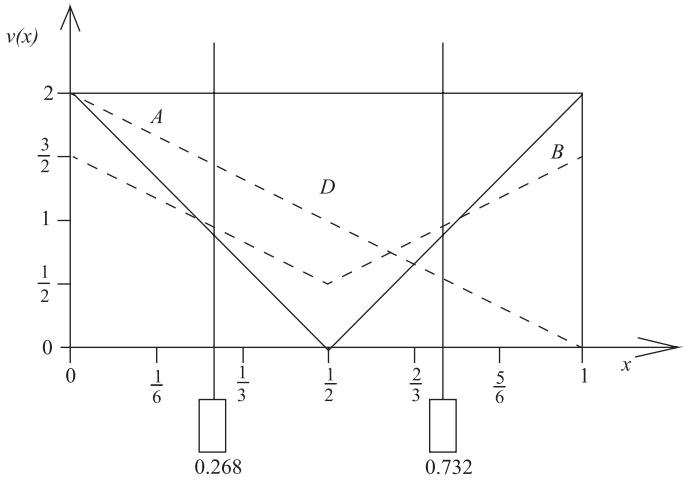


Figure 2 Value functions $v(x)$ of A, B, and D in Example 2 and maximin envy free cutpoints.

The maximin envy-free allocation only slightly raises the minimum value that a player receives from 36.2% to 37.2% (for B) while lowering C's share by 8.6 percentage points and raising A's share by 3.5 percentage points. Thus, the biggest effect of ensuring that the envy-free allocation is maximin is not on the minimum share that the worst-off player receives but on the shares of the other players, benefiting one player (A) and hurting the other (C).

Example 2 Modify Example 1 by replacing player C with player D, whose value function is $v_D(x) = -2x + 2$ over $[0, 1]$. Thus, D has a downward sloping value function, most valuing cake near $x = 0$ and least valuing it near $x = 1$, as shown in Figure 2.

To determine whether A, B, or both players prefer (before squeezing) the middle piece or the right when D trisects the cake, we calculate the $1/3$ cutpoints d_1 and d_2 that define this piece for D

$$\int_0^{d_1} (-2x + 2)\, dx = \frac{1}{3} \qquad \text{and} \qquad \int_{d_2}^{1} (-2x + 2)\, dx = \frac{1}{3}$$

The left equation gives $d_1 = 0.184$, and the right equation gives $d_2 = 0.423$. Manifestly, both A and B prefer the right subinterval $[0.423, 1]$, which gives each player more than 50% of the total value of the cake, to the middle subinterval $[0.184, 0.423]$. Hence, Step 5 of the algorithm is applicable, in which D squeezes the right piece from the left and the middle so that it ties for A or B with the enlarged middle or right pieces.

Enlarging the left piece $[0, 0.184]$ and the middle piece $[0.184, 0.423]$ equally requires that these two pieces be equal for D:

$$\int_0^a (-2x + 2)\, dx = \int_a^b (-2x + 2)\, dx$$

$$-2a^2 + 4a = -b^2 + 2b \tag{5a}$$

D will stop squeezing whichever of the following two events occurs first:

- The enlarged left piece, which A values more than B, equals the diminished right piece for A:

$$\int_0^a (-4x+2)\,dx = \int_b^1 (4x-2)\,dx$$

$$-2a^2 + 2a = -2b^2 + 2b \tag{5b}$$

Solving (5a) and (5b) simultaneously yields a feasible solution of $a = 0.268$ and $b = 0.732$. This gives B a share of 34.0% for the subinterval [0, 0.236]—when a tie for A occurs with the enlarged left piece—D a share of 46.4% for the subinterval [0.268, 0.732], and A a share of 39.2% for the subinterval [0.732, 1]. Hence, when a tie occurs for A with the diminished right piece, the players receive pieces in the order A-D-B, going from left to right.

- The enlarged middle piece, which B values more than A, equals the diminished right piece for B:

$$\int_a^{1/2} \frac{-4x+3}{2}\,dx + \int_{1/2}^b \frac{4x-1}{2}\,dx = \int_b^1 \frac{4x-1}{2}\,dx$$

$$a^2 - \frac{3}{2}a = -2b^2 + b \tag{5c}$$

Solving (5a) and (5c) simultaneously yields a feasible solution of $a = 0.267$ and $b = 0.727$. This gives D a share of 46.3% for the subinterval [0, 0.267]—when a tie for B occurs with the enlarged middle piece—B a share of 33.4% for the subinterval [0.267, 0.727], and A a share of 39.7% for the subinterval [0.727, 1]. Hence, when a tie occurs for B, the players receive pieces in the order D-B-A.

Because the D-B-A solution occurs slightly before (at $a = 0.267$) the A-D-B solution (at $a = 0.268$), the envy-free algorithm gives the D-B-A solution, whereby D gets the left piece, where its value is most concentrated. If the squeezing of the right piece continues to $a = 0.268$, the right piece will be slightly reduced—giving D 46.3% rather than 46.4%—so that it ties with the middle piece for B, giving the A-D-B solution, in which D gets the middle piece.

We know that the latter solution is envy-free for A and B because these players get left and right pieces, respectively, that are equidistant from the endpoints of 0 and 1. Moreover, D will not be envious of either A or B because, as squeezer, D maintains a tie between the middle piece and left piece, both of which D values more than the diminished right piece.

Between the two envy-free solutions of D-B-A and A-D-B, A-D-B gives the worst-off player (A) a bit more (34.0%) than the D-B-A solution gives the worst-off player (33.4% for B), so A-D-B is the maximal envy-free solution. Unlike Example 1, the D-B-A solution does not give A and B equal-length pieces, whereas the A-D-B solution does.

If the order of players from left to right is A-D-B, then there is an equitable solution, which can be found by setting A's valuation of its piece equal to B's and requiring that D value A's piece the same as its piece:

$$\int_0^a (-4x+2)\,dx = \int_0^a \frac{4x-1}{2}$$

$$\int_0^a (-2x+2)\,dx = (-2x+2)\,dx$$

After integration and evaluation, the preceding equations become

$$-a^2 + 2a = -2b^2 + 2b \qquad \text{and} \qquad -2a^2 + 4a = -b^2 + 2b$$

When solved simultaneously, they yield a feasible solution of $a = 0.256$ and $b = 0.675$, with A getting the left piece, D the middle piece, and B the right piece. This gives each player a share of 38.2%, but it is not an envy-free allocation, because A and B receive different-length pieces.

By comparison, if the left-to-right order of the players is D-B-A, there is a different equitable solution, which can be found by setting D's valuation of its piece equal to A's and requiring that B value A's piece the same as its piece:

$$\int_0^a (-2x + 2)\, dx = \int_b^1 (4x - 2)\, dx$$

$$\int_a^{1/2} \frac{-4x + 3}{2}\, dx + \int_{1/2}^b \frac{4x - 1}{2}\, dx = \int_b^1 (4x - 2)\, dx$$

After integration and evaluation, the preceding equations become

$$-a^2 + 2a = -2b^2 + 2b \qquad \text{and} \qquad a^2 - \frac{3}{2}a + \frac{1}{2} = -3b^2 + \frac{5}{2}b.$$

When solved simultaneously, they yield a feasible solution of $a = 0.219$ and $b = 0.734$, with D getting the left piece, B the middle piece, and A the right piece. This gives each player a share of 39.0%, but it is not an envy-free allocation, because D envies B for getting a piece that D values at 53.9%.

Table 2 shows the subintervals, and the percent shares, that each player receives according to the algorithm and the other properties that we have discussed: Unlike

ALLOCATIONS	A	B	C
Envy-free allocation Left-to-right order: D-B-A	[0.727, 1] 39.7%	[0.267, 0.727] 33.4%	[0, 0.267] 46.3%
Envy-free allocation (maximin) Left-to-right order: A-D-B	[0, 0.268] 39.2%	[0.732, 1] 34.0%	[0.268, 0.732] 46.4%
Equitable allocation Left-to-right order: A-D-B	[0, 0.256] 38.2%	[0.675, 1] 38.2%	[0.256, 0.675] 38.2%
Equitable allocation (maximal) Left-to-right order: D-B-A	[0.734, 1] 39.0%	[0.219, 0.734] 39.0%	[0, 0.219] 39.0%

TABLE 2: Subintervals and percent shares to each player in Example 2.

Example 1, the order from left to right in which the players receive pieces makes a difference for both the envy-free and the equitable allocations.[†]

Whereas both orders of the two envy-free allocations are efficient, this is not so for the equitable allocations: The order D-B-A Pareto-dominates the order A-D-B, making D-B-A the maximally equitable allocation. On the other hand, the order A-D-B is the maximin envy-free allocation because the minimum it gives to a player (34.0% to A) is greater than the minimum that D-B-A gives to a player (33.4% to B).

[†]Another difference is that the maximin envy-free allocation in Example 1 occurs when the value to the worst-off player (B) of its piece (37.2%) equals the value to the recipient (C) of the squeezed piece. In Example 2, the maximin envy-free allocation does not equalize these values because it is based on a different order (A-D-B) from the Barbanel-Brams [4] envy-free allocation (D-B-A), which creates a tie earlier (when $a = 0.267$).

Conclusions

We have described a three-person, two-cut algorithm for finding an envy-free allocation of a cake—inspired by the continuous moving-knife procedure of Barbanel and Brams [**4**]—when the players submit their value functions to a referee instead of moving knives. This allocation is efficient and gives connected pieces, as we illustrated with two examples in which the players' value functions are piecewise linear.

But more than duplicating the results of a moving-knife procedure, the players' submissions of their value functions enable the referee to make calculations, which are purely mechanical, that yield a maximin envy-free allocation. In general, this raises the value that the worst-off player receives, which makes it arguably fairer than the Barbanel-Brams envy-free allocation. (This is also true of the maximin envy-free division of indivisible items [**6**, **17**]; for empirical evidence in support of maximin, see Gal et al. [**13**]).

Similarly, a referee can use the value functions to find a maximally equitable allocation, which gives all players equal shares—as they value them—that are maximal. While the latter allocation may raise the value of the piece that the worst-off player receives from the maximin envy-free allocation, it may do so at the price of creating envy.

It is not evident that the envy-free algorithms can be extended to $n > 3$ players and still use the minimal $n - 1$ cuts. In particular, the extensions present two hurdles: finding the order of players along a line to which to assign a portion of the cake, and determining the cutpoints that give them envy-free pieces. Although one can try the $n!$ different orders and solve equations that guarantee envy-freeness, as in Willson [**22**], the advantage of our algorithm is the guarantee that it provides of a solution, which will be exact for certain classes of value functions.

Acknowledgments We thank the previous editor Michael A. Jones, an associate editor, and two anonymous referees for their valuable comments on earlier versions of this article, and the current editor for his efforts to put the article in final form.

REFERENCES

[1] Amanatidis, G., Christodoulou, G., Fearnley, E., Psomas, C.-A., Vakaliou, E. (2018). An improved envy-free cake cutting protocol for four agents. In: *Algorithmic Game Theory*. Lecture Notes in Computer Science, Vol. 11059. Springer, pp. 87–99.

[2] Aziz, H., Mackenzie, S. (2016). A discrete and bounded envy-free cake cutting protocol for any number of agents. *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016*. Los Alamitos, CA: IEEE Computer Society, pp. 416–427.

[3] Barbanel, J. B. (2005). *The Geometry of Efficient Fair Division*. New York: Cambridge University Press.

[4] Barbanel, J. B., Brams, S. J. (2004). Cake division with minimal cuts: Envy-free procedures for 3 persons, 4 persons, and beyond. *Math. Social Sci.* 48(3): 251–259. doi.org/10.1016/j.mathsocsci.2004.03.006

[5] Brams, S. J., Jones, M. A., Klamler, C. (2006). Better ways to cut a cake. *Not. AMS* 53(11): 1314–1321.

[6] Brams, S. J., Kilgour, D. M., Klamler, C. (2017). Maximin envy-free division of indivisible items. *Group Decis. Negot.* 26(1): 115–131. doi.org/10.1007/s10726-016-9502-x

[7] Brams, S. J., Taylor, A. D. (1995). An envy-free cake division protocol. *Amer. Math. Monthly.* 102(1): 9–18. doi.org/10.2307/2974850

[8] Brams, S. J., Taylor, A. D. (1996). *Fair Division: From Cake-Cutting to Dispute Resolution*. New York: Cambridge University Press.

[9] Brams, S. J., Taylor, A. D., Zwicker, W. S. (1997). A moving-knife solution to the four-person envy-free cake division problem. *Proc. Amer. Math. Soc*. 125(2): 547–554. doi.org/10.1090/S0002-9939-97-03614-9

[10] Brânzei, S. (2015). A note on envy-free cake cutting with polynomial valuations. *Inf. Process. Lett.* 15: 93–95. doi.org/j.ipl.2014.07.005

[11] Chèze, G. Cake cutting: Explicit examples for impossibility results. (2019). *Math. Social Sci.* 102: 68–72. doi.org/10.1016/j.mathsocsci.2019.09.005

[12] Dehgani, S., Farhadi, A., Taghi, M., Aghayi, H., Yami, H. (2018). Envy-free chore division for an arbitrary number of agents. *Proc. of the Twenty-Ninth Annual ACM-SIAM Symp. on Disc. Algorithms*, pp. 2564–2583.
[13] Gal, Y., Mash, M., Procaccia, A. D., Zick, Y. (2017). Which is the fairest (rent division) of them all? *J. ACM*. 64(6). Article 39. doi.org/10.1145/3131361
[14] Gale, D. (1993). Mathematical entertainments. *Math. Intelligencer*. 15(1): 48–52. doi.org/10.1007/BF03025257
[15] Ianovski, E. (2012). Cake cutting mechanisms. https://arxiv.org/pdf/1203.0100.pdf
[16] Kurokawa, D., Lai, J. K., Proaccia, A. D. (2017). How to cut a cake before the party ends. *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, pp. 555–561.
[17] Kurokawa, D., Procaccia, A. D., Wang, J. (2018). Fair enough: Guaranteeing approximate maximin shares. *J. ACM*. 65(2): Article 8. doi.org/10.1145/3140756
[18] Neyman, J. (1946). Un théorème d'existence. *C. R. Acad. Sci. Paris*. 222, 843–845.
[19] Stromquist, W. (1980). How to cut a cake fairly. *Amer. Math. Monthly*. 18(8): 640–644. doi.org/10.1080/00029890.1980.11995109
[20] Stromquist, W. (2008). Envy-free cake divisions cannot be found by finite protocols. *Elec. J. Comb*. 15: #R11. doi.org/10.37236/735
[21] Su, F. E. (1999). Rental harmony: Sperner's lemma in fair division. *Amer. Math. Monthly*. 106(10): 430–442. doi.org/10.2307/2589747
[22] Willson, S. J. (1995). Fair division using linear programming. http://swillsoon.public.iastate.edu/FirDivisionUsingLPUnpublished6.pdf.
[23] Wolfram Equation Solver. (2020). https://www.wolframalpha.com/examples/EquationSolving.html.

Summary. We describe a 3-person, 2-cut envy-free cake-cutting algorithm, inspired by a continuous moving-knife procedure, that does not require that the players continuously move knifes across the cake. By having the players submit their value functions over the cake to a referee—rather than move knives according to these functions—the referee can ensure that the division is not only envy-free but also maximin. In addition, the referee can use the value functions to find a maximally equitable division, whereby the players receive equally valued shares that are maximal, but this allocation may not be envy-free.

STEVEN J. BRAMS has long been interested in fair division and is the author or coauthor of three books on the subject, the latest being *Mathematics and Democracy: Designing Better Voting and Fair-Division Procedures*. He has applied game theory and social-choice theory to voting and elections, bargaining and fairness, international relations, and the Bible, theology, and literature. He is Professor of Politics at New York University and works regularly with mathematicians, computer scientists, and economists on a variety of interdisciplinary topics.

PETER S. LANDWEBER is a retired Professor of Mathematics at Rutgers University, New Brunswick, NJ. Having followed the work of Steve Brams on fair division for many years, he is delighted to join in as a coauthor.

The Lucky Tickets

ARTEM LOGACHOV
Sobolev Institute of Mathematics
Novosibirsk State University
of Economics and Management
Novosibirsk, 630099, Russia
avlogachov@mail.ru

OLGA LOGACHOVA
Siberian State Univ. of
Geosystems and Technologies
Novosibirsk State University
of Economics and Management
Novosibirsk, Russia
omboldovskaya@mail.ru

STANISLAV RUSSIJAN
Donetsk National Technical University
Donetsk, 83001, Ukraine
st_russ@mail.ru

What is a lucky ticket? For some, checking if their ticket number is lucky is a type of mathematical amusement. Sometimes a six-digit number is used to enumerate tickets sold, for example, inside a city bus or at a train station. A ticket with a six-digit number is called a *lucky ticket* if the sum of the first three digits equals the sum of the last three digits. In Russia such tickets are usually called the *Moscow lucky tickets*. See Figure 1* with two Moscow lucky tickets, 713128 and 649784. We have $7+1+3=11=1+2+8$ and $6+4+9=19=7+8+4$.

Figure 1 Two Moscow lucky tickets.

Interestingly, there is a competing definition of a lucky ticket. One calls a ticket with a six-digit number lucky if the sum of the digits located in the odd places is equal to the sum of the digits occupying the even places. In Russia, such tickets are called the *St. Petersburg lucky tickets*. For example, a ticket with a number 896665 is the St. Petersburg lucky ticket since $8+6+6=20=9+6+5$. Evidently, the cardinality of Moscow lucky ticket numbers matches the cardinality of St. Petersburg lucky ticket numbers, as there is a bijection between the two sets. Thus, without loss of generality we can consider only the lucky tickets according to the Moscow version.

*The online version of this article has color diagrams.

Math. Mag. **95** (2022) 123–129. doi:10.1080/0025570X.2022.2023302 © Mathematical Association of America
MSC: 05A10, 11B65

In playing this game, one can go beyond the tickets with six-digit numbers—any number with an even number of digits will work. Take, for example, U. S. dollar bills, which have eight-digit serial numbers on them, and define a *lucky dollar* as shown in Figure 2.

Figure 2 *A lucky dollar.*

Several papers about lucky tickets were published in *Kvant* magazine in the Soviet Union and in present day Russia. *Kvant* is a popular science magazine with an emphasis on physics and mathematics for high school students and teachers. Its first issue was published in 1970, in the Soviet Union. The magazine continues its mission in Russia today. Some selected articles from *Kvant* were translated into English and published in the *Quantum Magazine* between 1990 and 2001. One of them was about the lucky tickets [2]. A collection of articles from *Kvant* (in Russian) devoted entirely to lucky tickets is available online [3].

Formulas for calculating the total number of lucky tickets with n-digit numbers, where n is even, can be found in [3], as well as in Savin and Fink [2], and [4, pp. 96-98]. It may seem surprising, but the following equality can be obtained by using generating functions and the Cauchy integral formula:

$$\begin{aligned} \#\{n \text{ digit lucky tickets}\} &= \frac{1}{(\frac{9n}{2})!}\left(\left(\frac{1-z^{10}}{1-z}\right)^n\right)^{(\frac{9n}{2})}\Bigg|_{z=0} \\ &= \frac{1}{\pi}\int_0^{\pi}\left(\frac{\sin 10x}{\sin x}\right)^n dx, \end{aligned} \tag{1}$$

where $\#A$ refers to the cardinality of the set A, and $f^{(r)}(z)$ denotes the r-th derivative of a function $f(z)$. The long vertical line with $z = 0$ means that the function value is calculated at the point $z = 0$.

The main problem

Let us consider a more general problem. An n-digit ticket is called the k-lucky ticket ($k < n$) if the sum of the first k digits is equal to the sum of the last $n - k$ digits. Specifically, an n-digit ticket number $a_1 \dots a_k a_{k+1} \dots a_n$ is k-lucky if the following equality holds

$$\sum_{r=1}^{k} a_r = \sum_{r=k+1}^{n} a_r.$$

Let us calculate the total number of k-lucky tickets for $n = 2$ and $n = 3$. Obviously, the two-digit numbers ($n = 2$) can have only 1-lucky tickets. They consist of the pairs of identical digits. Thus, there are exactly 10 of them.

For the three-digit numbers ($n = 3$), there are the 1-lucky and the 2-lucky tickets. Evidently, the total number of 1-lucky tickets is the same as the total number of 2-lucky tickets. This is why we only need to find the total number of 1-lucky tickets. If the first digit of the ticket is r, where $0 \le r \le 9$, then the second digit can be chosen in $r + 1$ ways, taking values from 0 to r. The third digit is then uniquely determined by the first two. Thus, there is only one 1-lucky ticket in which the first digit is zero (the ticket number is 000), two 1-lucky tickets with the first digit 1 (numbers 101 and 110), and so on, ending with ten 1-lucky ticket whose first digit is 9, specifically

$$909, 918, 927, 936, 945, 954, 963, 972, 981, 990.$$

Summing the above cases, we obtain that the total number of 1-lucky tickets (and, respectively, 2-lucky tickets) is equal to $\sum_{r=0}^{9}(r + 1) = 55$.

It is easy to check that as we allow more and more digits in a number, it becomes harder to calculate the number of k-lucky tickets. This is why we are interested in finding the general formula for calculating the number of k-lucky tickets. We will derive this formula using differential calculus and contour integration from complex analysis.

First, for a given n, we show that the number of k-lucky (n-digit) ticket numbers is equal to the number of n digit numbers whose digits sum to $9k$. For this, we replace the first k digits of a k-lucky ticket number by their complements to nine. That is, we consider the following one-to-one correspondence

$$a_1 \dots a_k a_{k+1} \dots a_n \Longleftrightarrow (9 - a_1) \dots (9 - a_k) a_{k+1} \dots a_n.$$

Next, we observe that the sum of the digits on the right-hand side is equal to $9k$. For example, the 2-lucky ticket number 382243 ($3 + 8 = 2 + 2 + 4 + 3$) is mapped into the number $(9 - 3)(9 - 8)2243$, i.e. 612243 ($6 + 1 + 2 + 2 + 4 + 3 = 18 = 9 \cdot 2$). Thus, we have reduced the problem to calculating the number of n digit numbers, whose digits sum up to $9k$.

Now, in order to solve this new problem, we consider the function

$$F(z) = z^0 + z^1 + z^2 + z^3 + z^4 + z^5 + z^6 + z^7 + z^8 + z^9.$$

Functions of this type are called *generating functions*. Taking the nth power of this function, we obtain

$$F^n(z) = (z^0 + z^1 + z^2 + z^3 + z^4 + z^5 + z^6 + z^7 + z^8 + z^9)^n.$$

When we multiply this out, we obtain a sum consisting of 10^n terms. Every term is a product of the form $z^{k_1} z^{k_2} \dots z^{k_n}$, where $k_r = 0, \dots, 9$ and $r = 1, \dots, n$. There is a one-to-one correspondence between every such term and a ticket number with digits $k_1 k_2 \dots k_n$. Hence, the terms $z^{k_1} z^{k_2} \dots z^{k_n} = z^{9k}$ correspond to the ticket numbers whose digits sum up to $9k$. For example, the product

$$\underbrace{z^9 \cdot \ldots \cdot z^9}_{k} \cdot \underbrace{z^0 \cdot \ldots \cdot z^0}_{n-k}$$

corresponds to the ticket number

$$\underbrace{9 \dots 9}_{k} \underbrace{0 \dots 0}_{n-k}.$$

We conclude that the number of tickets whose digits sum to $9k$ is equal to the coefficient in front of z^{9k} in the polynomial $F^n(z)$.

Our problem has been reduced to the finding the coefficientof z^{9k} in $F^n(z)$. Now, since $F(z)$ is a geometric sum of ten terms, we can represent it as follows:

$$F(z) = \begin{cases} \dfrac{1-z^{10}}{1-z} & \text{if } z \neq 1, \\ \\ 10 & \text{if } z = 1. \end{cases}$$

In order to find the coefficient in front of z^{9k}, we need to calculate the $(9k)$th derivative of $F^n(z)$ at 0, and divide it by $(9k)!$ (that is, compute the coefficient of z^{9k} in the Maclaurin series of $F^n(z)$). Since $F^n(z)$ is an entire function, its derivatives at zero over the set of real numbers coincide with their derivatives over the set of complex numbers. Therefore, we can apply the Cauchy integral formula from complex analysis:

$$f^{(m)}(z_0) = \frac{m!}{2\pi i} \oint\limits_{|z-z_0|=1} \frac{f(z)}{(z-z_0)^{m+1}}\, dz.$$

In our case $f(z) = \left(\dfrac{1-z^{10}}{1-z}\right)^n$, $m = 9k$, and $z_0 = 0$. Consequently, we have

$$\frac{1}{(9k)!}\left(\left(\frac{1-z^{10}}{1-z}\right)^n\right)^{(9k)}\Bigg|_{z=0} = \frac{1}{2\pi i}\oint\limits_{|z|=1}\left(\frac{1-z^{10}}{1-z}\right)^n \frac{1}{z^{9k+1}}\, dz.$$

We make the substitution $z = e^{i\varphi}$ on the right-hand side to obtain

$$\frac{1}{2\pi}\int\limits_0^{2\pi}\left(\frac{1-e^{10i\varphi}}{1-e^{i\varphi}}\right)^n \frac{e^{i\varphi}}{e^{(9k+1)i\varphi}}\, d\varphi.$$

We now make the substitution $x = \frac{\varphi}{2}$ and carry out the following computation:

$$\begin{aligned}
&= \frac{1}{\pi}\int\limits_0^{\pi}\left(\frac{1-e^{20ix}}{1-e^{2ix}}\right)^n \frac{1}{e^{18kix}}\, dx \\
&= \frac{1}{\pi}\int\limits_0^{\pi}\left(\frac{(e^{-10ix}-e^{10ix})/2}{(e^{-ix}-e^{ix})/2}\right)^n \frac{e^{10nix}}{e^{nix}}\,\frac{1}{e^{18kix}}\, dx \\
&= \frac{1}{\pi}\int\limits_0^{\pi}\left(\frac{\sin 10x}{\sin x}\right)^n e^{9(n-2k)ix}\, dx \\
&= \frac{1}{\pi}\int\limits_0^{\pi}\left(\frac{\sin 10x}{\sin x}\right)^n \Big(\cos(9(n-2k)x) + i\sin(9(n-2k)x)\Big)\, dx. \qquad (2)
\end{aligned}$$

Since the left-hand side of (2) is known to be a real number, the imaginary part of the right-hand side is zero. Consequently, the number of k-lucky tickets among all n

digit numbers is equal to

$$\frac{1}{(9k)!}\left(\left(\frac{1-z^{10}}{1-z}\right)^n\right)^{(9k)}\Bigg|_{z=0} = \frac{1}{\pi}\int_0^\pi \left(\frac{\sin 10x}{\sin x}\right)^n \cos(9(n-2k)x)\,dx. \qquad (3)$$

It is easy to check that, if n is even and $k = \frac{n}{2}$, then the formulas (3) and (1) coincide. We solved our problem!

We used *Maple* to implement formula (3). Curiously, in our calculations, the running time for the left-hand side of (3) was shorter than for the right-hand side of the same formula.

Notice that we solved our problem for the ticket numbers written in the decimal numeral system. This numeral system employs 10 as the base, and utilizes 10 different numerals—the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Now, we ask the same question when the ticket numbers are written in an arbitrary positional base d numeral system (where d is a natural number greater than 1). How do we find the number of k-lucky tickets in base d?

By analogy with the decimal system, it can be proved that the number of k-lucky tickets is equal to

$$\frac{1}{((d-1)k)!}\left(\left(\frac{1-z^d}{1-z}\right)^n\right)^{((d-1)k)}\Bigg|_{z=0} = \frac{1}{\pi}\int_0^\pi \left(\frac{\sin dx}{\sin x}\right)^n \cos((d-1)(n-2k)x)\,dx. \qquad (4)$$

It is easy to see, that the number of k-lucky tickets tends to infinity when the base $d \to \infty$.

As we see in Figure 1, the ticket numbers for the American Freedom Train and for the Moscow trolleybus consist of six digits. Now we have a formula that, for a given n, computes the total number of k-lucky tickets (Σ). Table 1 counts the number of k-lucky tickets for six-digit numbers

k	1	2	3	4	5
Σ	2002	25927	55252	25927	2002

TABLE 1: The number of k-lucky tickets for the six-digit ticket numbers

The numbers on the US dollar bills (see Figure 2) consist of eight digits. The counts of k-lucky tickets, $k = 1, \ldots, 7$, among the eight-digit ticket numbers are given in Table 2.

k	1	2	3	4	5	6	7
Σ	11440	429220	2706880	4816030	2706880	429220	11440

TABLE 2: The number of k-lucky tickets for the eight-digit ticket numbers

Now we ask the following question: How do we find the number of tickets that are simultaneously k_1-lucky, k_2-lucky, ..., k_r-lucky, where $0 < k_1 < k_2 < \ldots < k_r < n$? It is easy to notice that the numbers on such tickets are of the form

$a_1 \dots a_{k_1} 0 \dots 0 a_{k_r+1} \dots a_n$, where $a_1 + \dots + a_{k_1} = a_{k_r+1} + \dots + a_n$. It follows that the total number of such tickets is equal to the number of k_1-lucky tickets among $n - k_r + k_1$ digit ticket numbers. This quantity is expressed by the following formula:

$$\frac{1}{(9k_1)!}\left(\left(\frac{1-z^{10}}{1-z}\right)^{n-k_r+k_1}\right)^{(9k_1)}\Bigg|_{z=0}$$
$$= \frac{1}{\pi}\int_0^{\pi}\left(\frac{\sin 10x}{\sin x}\right)^{n-k_r+k_1}\cos(9(n-k_r-k_1)x)\,dx.$$

Calculation of combinatorial sums using integrals

In this section, we only consider numbers written in the binary system. Here, the total number of k-lucky tickets can be found directly via combinatorial considerations. As you know, in the binary system, any number can be represented as a sequence of 0s and 1s. This allows for the use of simple combinatorics.

First, we compute the number of k-lucky tickets such that the sum of the first k digits adds up to r. In such numbers, the first k digits consist of r ones and $k - r$ zeros. The remaining $n - k$ digits will also consist of r ones and $n - k - r$ zeros. Among the first k digits, we place r ones in $\binom{k}{r}$ ways, and the same is true for the remaining $n - k$ digits. (Here, $\binom{k}{r} = \frac{k!}{r!(k-r)!}$ is the binomial coefficient.) Therefore, the total number of k-lucky tickets for which the sum of the first k digits is r equals $\binom{k}{r}\binom{n-k}{r}$. Since each digit is either 0 or 1, we require $0 \le r \le \min(k, n-k)$. For simplicity we assume that $\min(k, nk) = k$. Thus, the total number of k-lucky tickets among all n digit binary ticket numbers is

$$\sum_{r=0}^{k}\binom{k}{r}\binom{n-k}{r}. \tag{5}$$

Let us calculate the same quantity in an alternative way. We find a one-to-one correspondence between the k-lucky tickets and the tickets whose digits sum to k. Specifically, we replace the first k digits with their opposites. Thus, a k-lucky number of the form $a_1 \dots a_k a_{k+1} \dots a_n$ is mapped into $(1 - a_1) \dots (1 - a_k) a_{k+1} \dots a_n$. We will arrive with the numbers whose digits sum up to k. Their total number is $\binom{n}{k}$ (we choose k places for the ones from n).

It follows from (4) and (5) that

$$\binom{n}{k} = \sum_{r=0}^{k}\binom{k}{r}\binom{n-k}{r} = \sum_{r=0}^{k}\binom{k}{r}\binom{n-k}{n-k-r}$$
$$= \frac{1}{\pi}\int_0^{\pi}\left(\frac{\sin 2x}{\sin x}\right)^{n}\cos((n-2k)x)\,dx$$
$$= \frac{2^n}{\pi}\int_0^{\pi}\cos^n(x)\cos((n-2k)x)\,dx.$$

Note that the equality $\binom{n}{k} = \sum_{r=0}^{k}\binom{k}{r}\binom{n-k}{n-k-r}$ is known as the particular case of Vandermonde's identity (see Askey [1, pp. 59 and 60]).

Acknowledgments The authors would like to thank Yevgeniy Kovchegov of Oregon State University, and Anatoly Yambartsev of the Institute of Mathematics and Statistics at the University of São Paulo, for several helpful suggestions that greatly improved the quality of this paper.

REFERENCES

[1] Askey, R. (1975). *Orthogonal Polynomials and Special Functions.* Philadelphia, PA: SIAM.
[2] Savin, A., Fink, L. (1992). A conversation in a streetcar. What are the chances of getting a "lucky ticket"? *Quantum.* 2(4): 23–25.
[3] Various Authors. (2017). *Let's count the lucky tickets.* Articles from "Kvant". http://ega-math.narod.ru/Quant/Tickets.htm.
[4] Vilenkin, N. (1972). *Combinatorial Mathematics for Recreation.* Moscow: Mir Publishers.

Summary. "Lucky tickets" are n-digit numbers whose digits satisfy various arithmetical equations. For example, a Moscow lucky ticket is a six-digit number in which the sum of the first three digits is the same as the sum of the last three digits. We summarize the story of lucky tickets, and introduce the notion of k-lucky tickets. Then we find formulas for calculating of the total numbers of such tickets.

ARTEM LOGACHOV obtained a Ph.D. in mathematics from the Sobolev Institute of Mathematics, Siberian Branch of the RAS, Russia, in 2015. He is a senior researcher at the Laboratory of Probability and Mathematical Statistics, Sobolev Institute of Mathematics. He also lectures on various mathematical disciplines for students of several universities in Novosibirsk, Russia. He is fond of interesting mathematical problems not only in higher mathematics, but also in the school course of mathematics, and he is always looking for original solutions. He likes to play chess.

OLGA LOGACHOVA obtained a Ph.D. in mathematics from the Institute of Applied Mathematics and Mechanics of NASU, Ukraine, in 2010. She is an associate professor in two universities: Siberian State University of Geosystems and Technologies and Novosibirsk State University of Economics and Management, Russia. Her research interests are probability theory, economic modeling, and partial differential equations. She enjoys playing guitar.

STANISLAV RUSSIJAN obtained a Ph.D. in Technical Sciences from Donetsk National Technical University, Ukraine, in 2012. He is an associate professor in the Department of Higher Mathematics in Donetsk National Technical University, Ukraine. His research interests include higher mathematics, economic modeling, and electrical engineering. He enjoys reading, music, and tourism.

Collatz Meets Fibonacci

MICHAEL ALBERT
Otago University, Otago, New Zealand
michael.albert@otago.ac.nz

BJARKI GUDMUNDSSON
Google Inc., Switzerland
suprdewd@gmail.com

HENNING ULFARSSON
Reykjavik University, Reykjavk, Iceland
henningu@ru.is

A well-known conjecture due to Lothar Collatz from 1937 states that when the function

$$f(x) = \begin{cases} x/2 & \text{if } x \equiv 0 \pmod 2 \\ 3x+1 & \text{if } x \equiv 1 \pmod 2 \end{cases}$$

is iterated from an initial positive integer x we eventually reach the cycle $(1, 4, 2)$. Although many have tried either to prove it or to find a counterexample, it remains open to this day. In what follows, we will assume that the conjecture is true. A collection edited by Lagarias [1] contains many articles related to the conjecture and its generalizations, as well as an annotated bibliography. One of the intriguing features of this function is that, starting from any initial value x, the sequence of iterates: x, $f(x)$, $f(f(x)), \ldots$, behaves irregularly at first before its seemingly inevitable decline from some power of two down to the three-element cycle. In this paper, we consider the initial interesting phase of these iterations as seen from a distance. That is, only with regard to their relative values, and not the actual numbers produced.

For example, starting the iteration at 12 leads to the sequence

$$\underbrace{12,\ 6,\ 3,\ 10,\ 5}_{\text{trajectory}},\ 16,\ 8,\ 4,\ 2,\ 1,\ 4,\ 2,\ 1,\ \ldots\ .$$

The sequence of numbers up until the first power of two is the interesting phase of the iteration, which we will call its *trajectory*. The elements of a trajectory are all distinct, and viewed from a distance we might replace each element of a trajectory by its rank (i.e., the i^{th} smallest number of the trajectory is replaced by i). The resulting sequence is a permutation, and we shall call permutations produced in this manner *Collatz permutations*. We denote the Collatz permutation obtained from initial value x by $C(x)$. So we have: $C(12) = 5\,3\,1\,4\,2$. Clearly, the map C is not one to one. For instance:

$$1 = C(5) = C(21) = C(85) = \cdots = C((2^{2k}-1)/3) = \cdots \quad k \in \mathbb{Z}_{\geq 2}$$

However, as the length of $C(x)$ increases, coincidences become more rare. For example, the only other $x \leq 1000$ with $C(x) = C(12)$ is $x = 908$. It seems natural to ask: *Among the permutations of length n, how many are Collatz permutations?*

Considering only those $x \leq 10^8$ for which the length of $C(x)$ is at most seven produces Table 1. As the reader will have noticed, the values in the table are the Fibonacci

Math. Mag. **95** (2022) 130–136. doi:10.1080/0025570X.2022.2023307 © Mathematical Association of America
MSC: 05A05, 11B83

Length	Collatz Permutations
1	1
2	1
3	2
4	3
5	5
6	8
7	13

TABLE 1: Number of Collatz permutations of length less than eight (experimentally).

numbers. In what follows we will explain this phenomenon and show that it persists through length 14. Beyond that point, excess permutations appear—and we will explain how and why this occurs as well.

Types of trajectories

The appearance of the Fibonacci numbers in the enumeration of Collatz permutations is easy to explain. The steps that occur in a trajectory are of two types depending on the parity of the argument. We can call them: *up steps* ($x \mapsto 3x+1$ when x is odd), denoted by u; and *down steps* ($x \mapsto x/2$ when x is even), denoted by d. Two up steps can never occur consecutively since $3x+1$ is even when x is odd.* The step types in a trajectory can be recovered from the resulting Collatz permutation according to the pattern of rises and descents. We call the resulting sequence of us and ds the *type* of the trajectory. As well as not containing consecutive us, the last symbol in such a sequence must be a d (since there is a "hidden" u occurring next to take us to a power of two). Recall that the trajectory produced by the starting number $x = 12$ was $12, 6, 3, 10, 5$ with the type $ddud$. The "hidden" up step, which we do not record, is from the 5 to the first power of two, 16.

As is well known, the number of such sequences of length n is given by F_n, the n^{th} Fibonacci number (with $F_1 = F_2 = 1$ and $F_n = F_{n-1} + F_{n-2}$ for $n > 1$).† So, in order to show that there are at least F_n Collatz permutations of length n it will be enough to show that any sequence of us and ds satisfying the necessary conditions above actually occurs as the type of some trajectory. To prove this, it is helpful to turn our focus to the end of a trajectory, rather than its beginning.

The final element of any trajectory is always a number of the form $(A-1)/3$ where $A = 2^a$ (it is not hard to see a must be even, but we will be making further requirements shortly, so ignore this for the moment). Think of u and d as functions and let their inverses be denoted U and D, that is:

$$U(x) = u^{-1}(x) = (x-1)/3 \qquad \text{and} \qquad D(x) = d^{-1}(x) = 2x.$$

A *witness* for a type, σ, is an $A = 2^a$ such that there is a trajectory ending at $(A-1)/3$ with type σ. For example, to find a witness for the type $uddud$ requires at least that $UDDUDU(A)$ be an integer. Note here that the "hidden u" has become explicit, and

*This is why some authors consider the "fundamental" steps of the Collatz iteration to be $x \mapsto (3x+1)/2$ for odd x, and $x \mapsto x/2$ for even x.

†The simplest way to see this is to note that any sequence of length n of this type, for $n \geq 3$ is obtained in one of two mutually exclusive ways: by prepending a d to a sequence of length $n-1$, or ud to a sequence of length $n-2$. Variations of this idea occur in many combinatorics texts, e.g., Exercise 10.3.1 in Matousek and Nesetril [3] and the hint for the solution at the back.

that the order of U's and D's in the functional application is the same as that in the type—because the type is written in the reverse of "normal" function application order! Now

$$UDDUDU(A) = \frac{8A - 29}{27}.$$

For this to be an integer, we must have that $A \equiv 7 \pmod{27}$. Remembering that $A = 2^a$, the least solution to this is $A = 2^{16} = 65536$. Unraveling the applications of U's and D's leads to $x = 19417$. The trajectory of x is

$$19417,\ 58252,\ 29126,\ 14563,\ 43690,\ 21845,$$

and $C(x) = 2\,6\,4\,1\,5\,3$.

We now check that, given a type σ, it is enough to find a witness A in this way (i.e., that the initial necessary integrality condition is also sufficient) and that infinitely many witnesses always exist.

Proposition 1. *If a type σ contains k u's, then there is a single congruence of the form $A = c \pmod{3^{k+1}}$ which must be satisfied in order for a trajectory of type σ to end with witness A. Consequently, there is a least witness $A = 2^a$ with $a \leq 2 \cdot 3^k$, and a general witness is of the form 2^{a+jd} where j is a nonnegative integer and $d = 2 \cdot 3^k$.*

Proof. This is a consequence of Theorem 1 of Everett [2], but we give a proof for completeness. From σ construct the sequence, $\Sigma = \Sigma_1 \Sigma_2 \cdots \Sigma_n$ of U's and D's obtained by changing each u to U and each d to D and then appending a U. Now compute $\Sigma(A)$ formally obtaining an expression of the form:

$$\frac{2^m A - b}{3^{k+1}}$$

where m is the number of ds in σ, k is the number of us, and b is not a multiple of three. That these three conditions are satisfied is easily verified by formal application of either U or D to an expression of the same type. In order that A be a witness to a trajectory of type σ, it is necessary that this be an integer which, since 2 and 3^{k+1} are relatively prime, leads to a necessary condition of the form $A = c \mod 3^{k+1}$. To see that this condition is also sufficient suppose inductively, that $\Gamma(A) \in \mathbb{Z}$ for some suffix $\Gamma = \Sigma_\ell \Sigma_{\ell+1} \cdots \Sigma_n$ of Σ. If Γ begins with D, then formally $\Gamma(A) = (2^k A - 2b)/3^m$ for some $k > 0$ and integer b. If this is an integer, then so is $d\Gamma(A) = (2^{k-1} A - b)/3^m$, which is the next element of the trajectory we are trying to construct. The case when Γ begins with U is easier since $u\Gamma(A) = 3\Gamma(A) + 1$ and so is still an integer as well. To finish, note that the size of the multiplicative group $(\mathbb{Z}/3^{k+1}\mathbb{Z})^\times$ is given by the Euler ϕ-function $\phi(3^{k+1}) = 2 \cdot 3^k$. Furthermore, it is well-known that two is a primitive root in this group, meaning that we can find a solution $a \in \{1, 2, \ldots, 2 \cdot 3^k\}$ to the equation $2^a \equiv c \pmod{3^{k+1}}$. ■

This proposition shows that every potential type has a witness and thereby proves that there are at least F_n Collatz permutations of length n.

Table 2 shows that there are exactly F_n Collatz permutations of length n for $n = 1, 2, \ldots, 14$ but for greater n there are more.

The shortest type that is associated with more than one Collatz permutation is $\sigma = uddudududdudd$, which has the integrality condition $2^a \equiv 16 \bmod 729$. The smallest solution to this equation is $a = 4$, corresponding to the trajectory 9, 28, 14, 7, 22, 11, 34, 17, 52, 26, 13, 40, 20, 10, 5 and the first permutation in Figure 1. However, the next solution to the integrality condition is $a = 490$. The initial

Length	#Perms	Length	#Perms	Excess
1	1	15	611	1
2	1	16	989	2
3	2	17	1600	3
4	3	18	2587	3
5	5	19	4185	4
6	8	20	6771	6
7	13	21	10953	7
8	21	22	17720	9
9	34	23	28669	12
10	55	24	46383	15
11	89	25	75044	19
12	144	26	121417	24
13	233	27	196448	30
14	377	28	317850	39
		29	514278	49
		30	832101	61
		31	1346346	77
		32	2178405	96

TABLE 2: Number of Collatz permutations of length less than 32.

$$\begin{array}{cccccccccccccc} \textcircled{3} & 12 & 7 & 2 & 10 & 5 & 13 & 8 & 15 & 11 & 6 & 14 & 9 & \textcircled{4}\ 1 \\ \textcircled{4} & 12 & 7 & 2 & 10 & 5 & 13 & 8 & 15 & 11 & 6 & 14 & 9 & \textcircled{3}\ 1 \end{array}$$

Figure 1 Two different permutations associated with the type $ududududdddd$.

number corresponding to this solution produces a different permutation than the smaller initial number, and it is shown in the second line in Figure 1.

In the next section we will explain why this type gives us two different permutations and when this should be expected.

Excess permutations

How can one type correspond to different permutations? Consider the type $\sigma = dududd$. If we are considering a potential witness t for it, then the trajectory that would be generated is:

$$DUDUDDU(t),\ UDUDDU(t),\ DUDDU(t),$$

$$UDDU(t),\ DDU(t),\ DU(t),\ U(t).$$

We can think of these as the following linear functions of t

$$\frac{16t-46}{27},\quad \frac{8t-23}{27},\quad \frac{8t-14}{9},\quad \frac{4t-7}{9},\quad \frac{4t-4}{3},\quad \frac{2t-2}{3},\quad \frac{t-1}{3}.$$

These lines are shown in Figure 2. We find a witness for the type wherever we find a vertical line $t = A$, where $A = 2^a$ and all the intersection points of $t = A$ with these lines are at integer heights. We can then read off the relative order of the lines at this

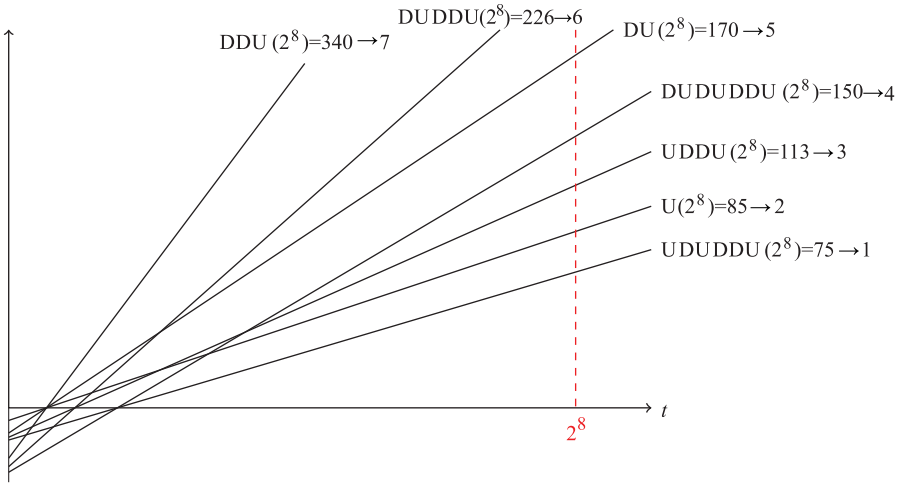


Figure 2 Linear functions determining permutations associated with the type *dududd*. Each function is evaluated at the witness 2^8 producing the permutation 4 1 6 3 7 5 2.

point according to the order they are crossed as we move up the line $t = A$ from the t-axis:

$DUDUDDU$	$UDUDDU$	$DUDDU$	$UDDU$	DDU	DU	U
↓	↓	↓	↓	↓	↓	↓
4	1	6	3	7	5	2

Consequently we see that if we have two potential witnesses such that there is no intersection point between two lines in the corresponding family lying between them, then they will determine the same permutation. On the other hand, if there were a witness between every pair of intersections of the lines for a type, and if these intersections all occurred at distinct abscissae, we might have up to $\binom{n}{2}$ witnesses for any given type producing distinct permutations. However, because the gap between successive witnesses is so large, the second potential witness always lies to the right of the rightmost intersection point among the lines. That is:

Proposition 2. *For any type σ, there are at most two distinct permutations $C(x)$ arising from x of type σ.*

Proof. Suppose that σ contains k occurrences of u. The values of a for which $A = 2^a$ are witnesses form an arithmetic progression with common difference $2 \cdot 3^k$, and in particular, the second witness is greater than $2^{2 \cdot 3^k}$. On the other hand, we can crudely bound the greatest abscissa of a point of intersection between the lines arising from σ as follows. Suppose that we have two such lines, one resulting from a suffix of Σ (notation as above) containing p occurrences of D and q occurrences of U (call this the pq-line), and the other from a longer suffix containing r occurrences of D and s occurrences of U (call this the rs-line). These lines have equations

$$y(x) = \frac{2^p x - b}{3^q}, \qquad \text{and} \qquad y(x) = \frac{2^r x - c}{3^s},$$

respectively. Suppose that the y intercept of the pq-line is less than that of the rs-line. The y-intercept of the pq-line is certainly greater than -2^{p-q}, while that of the rs-line is less than or equal to 0. So their intersection occurs to the left of the x-value

determined by:

$$\frac{2^p}{3^q}x - 2^{p-q} = \frac{2^r}{3^s}x,$$

which is

$$x = \frac{2^{p-q}}{2^p - 2^r 3^{s-q}} 3^s.$$

In the other case, the rs-line has lesser y-intercept than the pq-line. Using similar bounds gives an intersection point to the left of

$$x = \frac{2^{r-s}}{2^r - 2^p 3^{s-q}} 3^s.$$

In both estimates the denominator is a non-zero integer multiple of 2^p (since $p \leq r$ and $q \leq s$ and at least one of these inequalities is strict) and hence the first factors are bounded by $2^{-q}3^s$ and $2^{r-s-p}3^s$ respectively. Therefore, $x \leq 3^s \leq 3^k < 2^{2\cdot 3^k}$. ■

Recall that the type $\sigma = uddudududduddd$ had witnesses $2^4, 2^{490}, \ldots$. The greatest abscissa of an intersection point for the lines arising from this sequence is approximately 44.04. The first two witnesses are clearly on different sides of this intersection, explaining why we get two different Collatz permutations.

Propositions 1 and 2 show that the number of types of length n lies between F_n and $2F_n$. Moreover, the arguments leading up to Proposition 2 allow one to compute, given a type whether it corresponds to one or two Collatz permutations, and it is by these means that the data of Table 2 were derived.

To get an exact enumeration of the Collatz permutations one would need to understand which types are associated with two permutations. We call these *excess creating types* (*ET's*). Given an ET we can always create a new ET by prepending a d. This is because the extra d does not alter the integrality condition and can only increase the maximum intersection point. This at least shows that the number of ET's is non-decreasing.

Future work

As with so many aspects of the whole Collatz enigma, a few answers just seem to lead to more questions.

- How exactly does the number e_n of ET's of length n behave? The data above suggests that it might be something like "half Fibonacci rate" i.e., roughly proportional to the square root of the n^{th} Fibonacci number, or, equivalently, satisfying $e_n \sim e_{n-2} + e_{n-4}$.
- In our current dataset we always get $c = 16$ in the integrality conditions for ET's. This is probably just a result of our search procedure having explored only "small" values—but perhaps it might be a necessary condition.
- We have an extrinsic way of creating the Collatz permutations: run the Collatz process and see what comes out. Is there an intrinsic way to recognize these permutations, beyond the obvious condition that they cannot contain consecutive rises?
- There are several other maps similar to the Collatz map and there is also the modified Collatz sequence (where u is replaced with ud), as well as the Syracuse function where long down-steps are collapsed into one down-step. How do these analyses transfer to those contexts?

For the reader who is eager to start exploring Collatz permutations, we have a small code library supporting some basic features for working with their types. The code is written in the Sage open-source mathematics software system, but should run in Python with minor modifications. The code can be found at https://github.com/PermutaTriangle/CollatzPermutations.

Acknowledgments The authors are grateful to William Stein for access to the Sage Combinat Cluster, supported by NSF Grant No. DMS-0821725. We also thank the reviewers for comments that improved the paper.

REFERENCES

[1] Lagarias, J. C., (2010). *The Ultimate Challenge: The* $3x + 1$ *Problem*. Providence, RI: American Mathematical Society.

[2] Everett, C. J. (1977). Iteration of the number-theoretic function $f(2n) = n$, $f(2n + 1) = 3n + 2$. *Advances in Mathematics*. 25(1): 42–45. doi.org/10.1016/0001-8708(77)90087-1

[3] Matousek, J., Nesetril, J. (1998). *Invitation to Discrete Mathematics*. New York: Oxford University Press.

Summary. The Collatz map, a function from the set of positive integers to itself, maps an integer x to $x/2$ if x is even, and to $3x + 1$ if x is odd. The Collatz conjecture states that when the map is iterated the number one is eventually reached. We study permutations that arise as sequences from this iteration. We show that permutations of this type of length up to 14 are enumerated by the Fibonacci numbers. Beyond that, excess permutations appear. We will explain the appearance of these excess permutations and give an upper bound on their exact number.

MICHAEL ALBERT (MR Author ID: 24400) received his DPhil in mathematics as a Rhodes Scholar from Ontario at Oxford University. He held positions at the University of Waterloo and Carnegie Mellon University before moving to the University of Otago, where he is presently Professor and head of the Department of Computer Science. He is secretary of the New Zealand Mathematical Olympiad Committee, and has led New Zealand teams at the International Mathematical Olympiad on a number of occasions.

BJARKI GUDMUNDSSON (MR Author ID: 1168490) received his M.Sc. degree in Computer Science at Reykjavik University. He enjoys solving mathematical puzzles with computers and occasionally competes in programming contests. He currently works at Google Zurich. His personal website can be found at http://algo.is.

HENNING ULFARSSON (MR Author ID: 848375) received his PhD in mathematics from Brown University. He is an assistant professor at Reykjavik University. He has been a board member as well as the chair of the Icelandic Mathematical Society.

Using Rotation Matrices to Get ∇f and $\nabla \cdot \vec{F}$ in Cylindrical and Spherical Coordinates

FÁBIO M. S. LIMA
Institute of Physics
University of Brasília
70919-970, Brasília, DF, Brazil
fabio@fis.unb.br

PEDRO G. F. JORDÃO
Mathematics Department
University of Brasília
70910-900, Brasília, DF, Brazil
pedro103ssul@gmail.com

In nearly all calculus textbooks and classes, the gradient and divergence operators in cylindrical coordinates are derived by rewriting all rectangular coordinates (x, y, z) and unit vectors $(\hat{i}, \hat{j}, \hat{k})$ in terms of cylindrical ones, i.e., (r, θ, z) and $(\hat{r}, \hat{\theta}, \hat{k})$. Then, from the partial derivatives of r and θ with respect to x and y, one gets the 'inverse' partial derivatives of x and y with respect to r and θ, and then, after some rather tedious algebra, one arrives at the final expressions (see, e.g., Colley[**2**, Sec. 3.4]). In this note, we use rotation matrices to make these derivations less strenuous.

In introductory linear algebra, a rotation matrix is any matrix used to perform a rotation of the coordinate axes. In particular, the polar unit vectors $(\hat{r}, \hat{\theta})$ are easily determined by interpreting them as the unit vectors corresponding to the new coordinates axes x' and y' found with a *planar rotation* of θ around the z-axis, as illustrated in Figure 1. As shown in Anton [**1**, Sec. 4.9], these new unit vectors are related to the previous Cartesian ones by

$$\begin{pmatrix} \hat{i}' \\ \hat{j}' \end{pmatrix} = \begin{pmatrix} \hat{r} \\ \hat{\theta} \end{pmatrix} = R_\theta \begin{pmatrix} \hat{i} \\ \hat{j} \end{pmatrix}, \tag{1}$$

where

$$\hat{r} := \frac{\vec{r}}{|\vec{r}|} = \frac{(x\,\hat{i} + y\,\hat{j})}{r} = \cos\theta \hat{i} + \sin\theta\,\hat{j},$$
$$r := |\vec{r}| = \sqrt{x^2 + y^2},$$

and

$$R_\theta := \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \quad \theta \in \mathbb{R}, \tag{2}$$

is a *rotation matrix, counterclockwise* by θ. The unit vector $\hat{\theta} = a\,\hat{i} + b\,\hat{j}$ can be determined from the *orthonormality conditions* $\hat{\theta} \cdot \hat{r} = 0$ and $\hat{\theta} \cdot \hat{\theta} = 1$. This yields $a = \pm \sin\theta$ and $b = \mp \cos\theta$. We choose $\hat{\theta} = -\sin\theta\,\hat{i} + \cos\theta\,\hat{j}$ to make $\hat{\theta}$ a 90° *counterclockwise* rotation of $\hat{r}$, so that the ordered basis $\{\hat{r}, \hat{\theta}\}$ of $\mathbb{R}^2$ is *right-handed*.

Math. Mag. **95** (2022) 137–142. doi:10.1080/0025570X.2022.2025708 © Mathematical Association of America

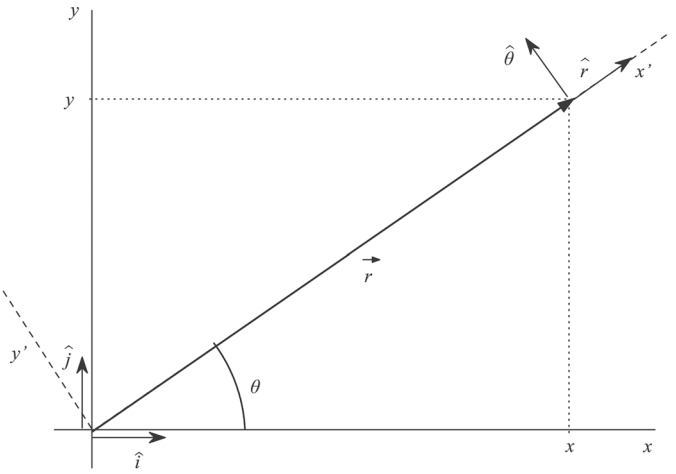


Figure 1 A counterclockwise rotation of θ of the coordinates axes and corresponding unit vectors. The new unit vectors $\hat{i}'$ and $\hat{j}'$ are just the polar unit vectors $\hat{r}$ and $\hat{\theta}$.

Since $\sin(-\theta) = -\sin\theta$ and $\cos(-\theta) = \cos\theta$, the matrix corresponding to a clockwise rotation by θ has the form

$$R_{(-\theta)} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{3}$$

from which it is clear that $R_\theta\, R_{(-\theta)} = R_0 = I_2$,* the identity matrix, so $R_{(-\theta)} = R_\theta^{-1}$, the inverse of R_θ.† By applying R_θ^{-1} to both sides of (1), one finds that

$$\begin{pmatrix} \hat{i} \\ \hat{j} \end{pmatrix} = R_\theta^{-1} \begin{pmatrix} \hat{r} \\ \hat{\theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \hat{r} \\ \hat{\theta} \end{pmatrix}. \tag{4}$$

With these algebraic ingredients in hand, let us derive our main results.

The *gradient* of a real scalar field $f(x, y, z)$ is defined by the formula

$$\nabla f := \frac{\partial f}{\partial x}\,\hat{i} + \frac{\partial f}{\partial y}\,\hat{j} + \frac{\partial f}{\partial z}\,\hat{k}\,. \tag{5}$$

In order to get the corresponding expression in cylindrical coordinates, one keeps z unaltered, whereas the in-plane coordinates are changed according to $x = r\cos\theta$ and $y = r\sin\theta$. The partial derivatives of $f(r, \theta, z)$ are then related to those of $f(x, y, z)$ with the aid of the chain rule. The result is

$$\begin{pmatrix} \dfrac{\partial f}{\partial r} \\[2ex] \dfrac{1}{r}\dfrac{\partial f}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \dfrac{\partial f}{\partial x} \\[2ex] \dfrac{\partial f}{\partial y} \end{pmatrix} = R_\theta \begin{pmatrix} \dfrac{\partial f}{\partial x} \\[2ex] \dfrac{\partial f}{\partial y} \end{pmatrix}. \tag{6}$$

*One might also get this result as a corollary of the geometrically obvious fact that $R_\alpha\, R_\beta = R_{\alpha+\beta}$. Incidentally, this identity provides a simple proof of the formulas for $\sin(\alpha+\beta)$ and $\cos(\alpha+\beta)$.

†Since R_θ^T also reduces to $R_{(-\theta)}$, it follows that $R_\theta^T = R_\theta^{-1}$, which shows that R_θ is an *orthogonal* matrix.

On applying the inverse R_θ^{-1} to both sides, one obtains

$$\begin{pmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{pmatrix} = R_\theta^{-1} \begin{pmatrix} \dfrac{\partial f}{\partial r} \\ \dfrac{1}{r}\dfrac{\partial f}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \dfrac{\partial f}{\partial r} \\ \dfrac{1}{r}\dfrac{\partial f}{\partial \theta} \end{pmatrix}. \tag{7}$$

On taking into account the unit vectors in (4), one finds

$$\begin{aligned} \nabla f &= \left(\cos\theta\,\frac{\partial f}{\partial r} - \frac{\sin\theta}{r}\,\frac{\partial f}{\partial \theta}\right)\left(\cos\theta\,\hat{r} - \sin\theta\,\hat{\theta}\right) \\ &\quad + \left(\sin\theta\,\frac{\partial f}{\partial r} + \frac{\cos\theta}{r}\,\frac{\partial f}{\partial \theta}\right)\left(\sin\theta\,\hat{r} + \cos\theta\,\hat{\theta}\right) + \frac{\partial f}{\partial z}\,\hat{k} \\ &= \frac{\partial f}{\partial r}\,\hat{r} + \frac{1}{r}\,\frac{\partial f}{\partial \theta}\,\hat{\theta} + \frac{\partial f}{\partial z}\,\hat{k}\,. \end{aligned} \tag{8}$$

The *divergence* of a real vector field

$$\vec{F} = F_x\,\hat{\imath} + F_y\,\hat{\jmath} + F_z\,\hat{k}$$

is defined by the formula

$$\nabla \cdot \vec{F} := \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}\,. \tag{9}$$

For simplicity, let $\vec{F} = \vec{F}_p + F_z\,\hat{k}$, where $\vec{F}_p = F_x\,\hat{\imath} + F_y\,\hat{\jmath}$. In cylindrical coordinates, taking into account the unit vectors in (4), one has*

$$\begin{aligned} \vec{F}_p &= F_x\left(\cos\theta\,\hat{r} - \sin\theta\,\hat{\theta}\right) + F_y\left(\sin\theta\,\hat{r} + \cos\theta\,\hat{\theta}\right) \\ &= \left(F_x\,\cos\theta + F_y\,\sin\theta\right)\hat{r} + \left(F_y\,\cos\theta - F_x\,\sin\theta\right)\hat{\theta} \\ &:= F_r\,\hat{r} + F_\theta\,\hat{\theta}\,. \end{aligned}$$

Interestingly, this definition for the polar components of $\vec{F}_p$ can also be written using the rotation matrix, as follows:

$$\begin{pmatrix} F_r \\ F_\theta \end{pmatrix} = R_\theta \begin{pmatrix} F_x \\ F_y \end{pmatrix}. \tag{10}$$

By applying the gradient operator (8) directly to $\vec{F}$, one finds that

$$\begin{aligned} \nabla \cdot \vec{F} &= \nabla \cdot \left(\vec{F}_p + F_z\,\hat{k}\right) \\ &= \hat{r} \cdot \frac{\partial \vec{F}_p}{\partial r} + \hat{r} \cdot \hat{k}\,\frac{\partial F_z}{\partial r} + \frac{1}{r}\,\hat{\theta} \cdot \left(\frac{\partial \vec{F}_p}{\partial \theta} + \frac{\partial F_z}{\partial \theta}\,\hat{k}\right) + \hat{k} \cdot \left(\frac{\partial \vec{F}_p}{\partial z} + \frac{\partial F_z}{\partial z}\,\hat{k}\right). \end{aligned} \tag{11}$$

*In general, the component functions F_r, F_θ, and F_z are functions of the three coordinates r, θ, and z. The subscripts only indicate to which of the unit vectors $\hat{r}$, $\hat{\theta}$, or $\hat{k}$ the component function is attached.

Since the unit vectors $\hat{r}$, $\hat{\theta}$, and $\hat{k}$ are orthogonal, it follows that $\vec{F}_p = F_r\,\hat{r} + F_\theta\,\hat{\theta}$ is orthogonal to $\hat{k}$. Equation (11) then simplifies to

$$\nabla \cdot \vec{F} = \hat{r} \cdot \frac{\partial \vec{F}_p}{\partial r} + \frac{1}{r}\,\hat{\theta} \cdot \frac{\partial \vec{F}_p}{\partial \theta} + \frac{\partial F_z}{\partial z}\,. \tag{12}$$

Using the product rule for derivatives, the radial term becomes

$$\hat{r} \cdot \frac{\partial \vec{F}_p}{\partial r} = \hat{r} \cdot \left(\hat{r}\,\frac{\partial F_r}{\partial r} + F_r\,\frac{\partial \hat{r}}{\partial r} + \hat{\theta}\,\frac{\partial F_\theta}{\partial r} + F_\theta\,\frac{\partial \hat{\theta}}{\partial r} \right) = \frac{\partial F_r}{\partial r}\,, \tag{13}$$

since $\hat{r}$ and $\hat{\theta}$ are independent of r. Since $\partial \hat{r}/\partial \theta = \hat{\theta}$ and $\partial \hat{\theta}/\partial \theta = -\hat{r}$, the azimuthal term becomes

$$\begin{aligned} \frac{\hat{\theta}}{r} \cdot \frac{\partial \vec{F}_p}{\partial \theta} &= \frac{\hat{\theta}}{r} \cdot \left(\hat{r}\,\frac{\partial F_r}{\partial \theta} + F_r\,\frac{\partial \hat{r}}{\partial \theta} + \hat{\theta}\,\frac{\partial F_\theta}{\partial \theta} + F_\theta\,\frac{\partial \hat{\theta}}{\partial \theta} \right) \\ &= \frac{\hat{\theta}}{r} \cdot \left(\hat{r}\,\frac{\partial F_r}{\partial \theta} + F_r\,\hat{\theta} + \hat{\theta}\,\frac{\partial F_\theta}{\partial \theta} - F_\theta\,\hat{r} \right) \\ &= \frac{1}{r} \left(F_r + \frac{\partial F_\theta}{\partial \theta} \right). \end{aligned} \tag{14}$$

On substituting these results into (12), we finally arrive at

$$\begin{aligned} \nabla \cdot \vec{F} &= \frac{\partial F_r}{\partial r} + \frac{1}{r} \left(F_r + \frac{\partial F_\theta}{\partial \theta} \right) + \frac{\partial F_z}{\partial z} \\ &= \frac{1}{r}\,\frac{\partial (r\,F_r)}{\partial r} + \frac{1}{r}\,\frac{\partial F_\theta}{\partial \theta} + \frac{\partial F_z}{\partial z}\,. \end{aligned} \tag{15}$$

Interestingly, the equations for conversion from Cartesian to cylindrical coordinates are analogous to those for conversion from cylindrical to spherical coordinates (ρ, ϕ, θ), namely (see, e.g., Larson [**3**, Sec. 11.7])

$$\begin{cases} z = \rho \cos \phi \\ r = \rho \sin \phi \\ \theta = \theta \end{cases} \tag{16}$$

where ρ is the distance to the origin O, ϕ is the polar angle (between the positive z-axis and the ray through the origin), and θ is the azimuthal angle (between the positive x-axis and the ray made by dropping a perpendicular from the point to the xy-plane). This suggests that our approach can be extended to spherical coordinates. In order to show that this is indeed the case, we rewrite (8) as

$$\nabla f = \left(\frac{\partial f}{\partial z}\,\hat{k} + \frac{\partial f}{\partial r}\,\hat{r} \right) + \frac{1}{r}\,\frac{\partial f}{\partial \theta}\,\hat{\theta}\,. \tag{17}$$

When x and y are replaced by r and θ in ∇f,

$$\frac{\partial f}{\partial x}\,\hat{i} + \frac{\partial f}{\partial y}\,\hat{j} \quad \text{becomes} \quad \frac{\partial f}{\partial r}\,\hat{r} + \frac{1}{r}\,\frac{\partial f}{\partial \theta}\,\hat{\theta}\,,$$

while $\partial f/\partial z\,\hat{k}$ is unchanged. Therefore, on replacing z and r by ρ and ϕ, respectively,

$$\frac{\partial f}{\partial z}\,\hat{k} + \frac{\partial f}{\partial r}\,\hat{r} \quad \text{should become} \quad \frac{\partial f}{\partial \rho}\,\hat{\rho} + \frac{1}{\rho}\,\frac{\partial f}{\partial \phi}\,\hat{\phi}\,,$$

while $\frac{1}{r}\,\frac{\partial f}{\partial \theta}\,\hat{\theta}$ is unchanged, except that we replace r by $\rho\,\sin\phi$. This gives

$$\nabla f = \left(\frac{\partial f}{\partial \rho}\,\hat{\rho} + \frac{1}{\rho}\,\frac{\partial f}{\partial \phi}\,\hat{\phi}\right) + \frac{1}{\rho\,\sin\phi}\,\frac{\partial f}{\partial \theta}\,\hat{\theta}\,, \tag{18}$$

which is the correct expression (see, e.g., Colley [**2**, Thm. 4.6]). Incidentally, the unit vectors in spherical coordinates themselves can be related to the (fixed) Cartesian unit vectors using a rotation matrix, as follows. From (1), we know that

$$\begin{pmatrix} \hat{r} \\ \hat{\theta} \end{pmatrix} = R_\theta \begin{pmatrix} \hat{\imath} \\ \hat{\jmath} \end{pmatrix},$$

so

$$\begin{aligned} \begin{pmatrix} \hat{\rho} \\ \hat{\phi} \end{pmatrix} = R_\phi \begin{pmatrix} \hat{k} \\ \hat{r} \end{pmatrix} &= \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} \hat{k} \\ \cos\theta\,\hat{\imath} + \sin\theta\,\hat{\jmath} \end{pmatrix} \\ &= \begin{pmatrix} \sin\phi\,\cos\theta\,\hat{\imath} + \sin\phi\,\sin\theta\,\hat{\jmath} + \cos\phi\,\hat{k} \\ \cos\phi\,\cos\theta\,\hat{\imath} + \cos\phi\,\sin\theta\,\hat{\jmath} - \sin\phi\,\hat{k} \end{pmatrix} \end{aligned} \tag{19}$$

and $\hat{\theta} = -\sin\theta\,\hat{\imath} + \cos\theta\,\hat{\jmath}$ is inherited from polar coordinates.

From the result in equation (18), we can derive the expression for the divergence operator in spherical coordinates by following the same procedure developed above for cylindrical coordinates. We begin with

$$\begin{aligned} \nabla \cdot \vec{F} &= \left(\hat{\rho}\,\frac{\partial}{\partial \rho} + \frac{\hat{\phi}}{\rho}\,\frac{\partial}{\partial \phi} + \frac{\hat{\theta}}{\rho\,\sin\phi}\,\frac{\partial}{\partial \theta}\right) \cdot \left(F_\rho\,\hat{\rho} + F_\phi\,\hat{\phi} + F_\theta\,\hat{\theta}\right) \\ &= \hat{\rho} \cdot \frac{\partial}{\partial \rho}\left(F_\rho\,\hat{\rho} + F_\phi\,\hat{\phi} + F_\theta\,\hat{\theta}\right) + \frac{1}{\rho}\,\hat{\phi} \cdot \frac{\partial}{\partial \phi}\left(F_\rho\,\hat{\rho} + F_\phi\,\hat{\phi} + F_\theta\,\hat{\theta}\right) \\ &\qquad + \frac{1}{\rho\,\sin\phi}\,\hat{\theta} \cdot \frac{\partial}{\partial \theta}\left(F_\rho\,\hat{\rho} + F_\phi\,\hat{\phi} + F_\theta\,\hat{\theta}\right). \end{aligned} \tag{20}$$

We again have to apply the product rule for derivatives on each term, in general. However, since the unit vectors $\hat{\rho}$, $\hat{\phi}$, and $\hat{\theta}$ are all independent of ρ, as can be seen in equation (19), the first term promptly reduces to $\partial F_\rho/\partial \rho$ only. The second term in equation (20) can be reduced by noting that $\partial \hat{\rho}/\partial \phi = \hat{\phi}$, $\partial \hat{\phi}/\partial \phi = -\hat{\rho}$, and $\hat{\theta}$ is independent of ϕ, which results in

$$\frac{F_\rho}{\rho} + \frac{1}{\rho}\,\frac{\partial F_\phi}{\partial \phi}\,. \tag{21}$$

Finally, on noting that

$$\frac{\partial \hat{\rho}}{\partial \theta} = \sin\phi\,\hat{\theta}, \quad \frac{\partial \hat{\phi}}{\partial \theta} = \cos\phi\,\hat{\theta}, \quad \text{and} \quad \frac{\partial \hat{\theta}}{\partial \theta} = -\sin\phi\,\hat{\rho} - \cos\phi\,\hat{\phi},$$

the third term in (20) simplifies to

$$\frac{F_\rho}{\rho} + \frac{\cos\phi\,F_\phi}{\rho\,\sin\phi} + \frac{1}{\rho\,\sin\phi}\,\frac{\partial F_\theta}{\partial \theta}\,. \tag{22}$$

On summing up these three results, one finds

$$\begin{aligned}\boldsymbol{\nabla}\cdot\vec{F} &= 2\,\frac{F_\rho}{\rho}+\frac{\partial F_\rho}{\partial\rho}+\frac{1}{\rho}\left(\frac{\partial F_\phi}{\partial\phi}+\frac{\cos\phi}{\sin\phi}\,F_\phi\right)+\frac{1}{\rho\,\sin\phi}\,\frac{\partial F_\theta}{\partial\theta}\\ &= \frac{1}{\rho^2}\,\frac{\partial}{\partial\rho}\left(\rho^2\,F_\rho\right)+\frac{1}{\rho\,\sin\phi}\,\frac{\partial}{\partial\phi}\left(\sin\phi\,F_\phi\right)+\frac{1}{\rho\,\sin\phi}\,\frac{\partial F_\theta}{\partial\theta}\,,\end{aligned}\tag{23}$$

as expected (see, e.g., Colley [**2**, Thm. 4.6]).

Of course, the correct expressions for the *Laplacian* operator $\nabla^2 f := \boldsymbol{\nabla}\cdot(\boldsymbol{\nabla} f)$ in cylindrical coordinates, namely

$$\nabla^2 f = \frac{1}{r}\,\frac{\partial}{\partial r}\left(r\,\frac{\partial f}{\partial r}\right)+\frac{1}{r^2}\,\frac{\partial^2 f}{\partial\theta^2}+\frac{\partial^2 f}{\partial z^2}\,,$$

as well as in spherical coordinates

$$\nabla^2 f = \frac{1}{\rho^2}\,\frac{\partial}{\partial\rho}\left(\rho^2\,\frac{\partial f}{\partial\rho}\right)+\frac{1}{\rho^2\,\sin\phi}\,\frac{\partial}{\partial\phi}\left(\sin\phi\,\frac{\partial f}{\partial\phi}\right)+\frac{1}{\rho^2\,\sin^2\phi}\,\frac{\partial^2 f}{\partial\theta^2}\,,$$

as seen, for example, in exercises 32 and 33 of Section 2.5 of Colley [**2**], which are special cases of our equations (15) and (23), respectively, can be derived effortlessly by setting $\vec{F} = \boldsymbol{\nabla} f$.

Acknowledgments We thank the anonymous referees for the hints about the divergence operator and for suggesting the extension to spherical coordinates. PGFJ thanks the National Council of Research (CNPq) for financial support in the form of a fellowship.

REFERENCES

[1] Anton, H., Rorres, C. (2014). *Elementary Linear Algebra*, 11th ed. New York: Wiley.
[2] Colley, S. J., (2012). *Vector Calculus*, 4th ed. New York: Pearson Education.
[3] Larson, R., Edwards, B. (2018). *Calculus*, 11th ed. Boston, MA: Cengage.

Summary. We show that the correct expressions for the gradient and divergence differential operators in cylindrical and spherical coordinates can be derived using rotation matrices. The method is intuitive and less strenuous than what is usually found in calculus textbooks.

FÁBIO M. S. LIMA (MR Author ID: 134223) is a professor of physics at the University of Brasília, where he earned his Ph.D. in theoretical physics in 2003. His research interests in mathematics include classical analysis, special functions, and analytic number theory, in particular the properties of the Euler gamma and Riemann zeta functions.

PEDRO G. F. JORDÃO is a graduate student of mathematics at University of Brasília, where he received his Licentiate Degree in Mathematics in 2018. His research interests in mathematics include number theory and classical analysis.

Another Proof of the Chu–Vandermonde Formula

WENCHANG CHU
Zhoukou Normal University
Henan, Zhoukou, P. R. China
chu.wenchang@unisalento.it

The Chu–Vandermonde convolution formula on binomial coefficients is well-known in combinatorics and number theory. It states that for a nonnegative integer n and two variables x and y, we have the formula

$$\sum_{k=0}^{n}\binom{x}{k}\binom{y}{n-k}=\binom{x+y}{n}.$$

There exist several different proofs of this formula. We provide a novel proof by means of finite differences.

Let Δ be the forward difference operator of unit increment with respect to the variable λ. Under the replacements

$$x\to x+\lambda \qquad \text{and} \qquad y\to y-\lambda,$$

the Chu–Vandermonde formula can be reformulated equivalently as

$$\sum_{k=0}^{n}\binom{x+\lambda}{k}\binom{y-\lambda}{n-k}=\binom{x+y}{n}.$$

Denote by $\mathcal{S}(\lambda)$ the binomial sum on the left. Then we have

$$\begin{aligned}\Delta\mathcal{S}(\lambda)=&\mathcal{S}(\lambda+1)-\mathcal{S}(\lambda)\\ =&\sum_{k=0}^{n}\binom{x+\lambda+1}{k}\binom{y-\lambda-1}{n-k}-\sum_{k=0}^{n}\binom{x+\lambda}{k}\binom{y-\lambda}{n-k}.\end{aligned}$$

By making use of the recurrence relations

$$\begin{aligned}\binom{x+\lambda+1}{k}&=\binom{x+\lambda}{k}+\binom{x+\lambda}{k-1},\\ \binom{y-\lambda}{n-k}&=\binom{y-\lambda-1}{n-k}+\binom{y-\lambda-1}{n-k-1},\end{aligned}$$

we can show that $\Delta\mathcal{S}(\lambda)$ vanishes as follows:

$$\begin{aligned}\Delta\mathcal{S}(\lambda)=&\sum_{k=0}^{n}\left\{\binom{x+\lambda}{k}+\binom{x+\lambda}{k-1}\right\}\binom{y-\lambda-1}{n-k}\\ &-\sum_{k=0}^{n}\binom{x+\lambda}{k}\left\{\binom{y-\lambda-1}{n-k}+\binom{y-\lambda-1}{n-k-1}\right\}\end{aligned}$$

Math. Mag. **95** (2022) 143–144. doi:10.1080/0025570X.2022.2022959 © Mathematical Association of America
MSC: 05A10

$$= \sum_{k=1}^{n} \binom{x+\lambda}{k-1}\binom{y-\lambda-1}{n-k} - \sum_{k=0}^{n-1} \binom{x+\lambda}{k}\binom{y-\lambda-1}{n-k-1} = 0,$$

where the last line is justified by making the change $k \to k-1$ for the second sum. Therefore, $\mathcal{S}(\lambda)$ is independent of λ and can be evaluated by assigning a particular value for λ. Taking $\lambda = y$, we find that $\mathcal{S}(y) = \binom{x+y}{n}$ which corresponds to the only surviving term with $k = n$ in the binomial sum $\mathcal{S}(\lambda)$. This completes the proof.

Before concluding the paper, we would like to emphasize that there are different methods to prove the Chu-Vandermonde formula. Three typical ones are the generating function approach as in Comtet [**1**, §1.13], combinatorial interpretations such as provided by Graham, Knuth, and Patashnik [**2**, §5.1] and the WZ-method of Petkovšek, Wilf, and Zeilberger [**3**, §7.3]. For the reader's pleasure, we now outline the generating function proof.

Let τ be a variable subject to $|\tau| < 1$ and $[\tau^n]\phi(\tau)$ stand for the coefficient of τ^n in the power series $\phi(\tau)$. Consider the binomial series

$$(1+\tau)^x = \sum_{k=0}^{\infty} \binom{x}{k}\tau^k \quad \text{and} \quad (1+\tau)^y = \sum_{k=0}^{\infty} \binom{y}{k}\tau^k.$$

Multiply these two series and extract the coefficient of τ^n. On one side, we have the binomial coefficient

$$[\tau^n](1+\tau)^x(1+\tau)^y = [\tau^n](1+\tau)^{x+y} = \binom{x+y}{n}.$$

On another side, we get the binomial convolution

$$\begin{aligned}
[\tau^n](1+\tau)^x(1+\tau)^y &= [\tau^n]\sum_{k=0}^{\infty}\binom{x}{k}\tau^k \sum_{j=0}^{\infty}\binom{y}{j}\tau^j \\
&= [\tau^n]\sum_{m=0}^{\infty}\tau^m \sum_{k=0}^{m}\binom{x}{k}\binom{y}{m-k} \\
&= \sum_{k=0}^{n}\binom{x}{k}\binom{y}{n-k}.
\end{aligned}$$

Equating the two expressions confirms again the Chu-Vandermonde formula.

REFERENCES

[1] Comtet, L. (1974). *Advanced Combinatorics.* Dordrecht, The Netherlands: North-Holland.

[2] Graham, R. L., Knuth, D. E., Patashnik, O. (1994). *Concrete Mathematics,* 2nd ed. Upper Saddle River, NJ: Addison-Wesley.

[3] Petkovšek, M., Wilf, H. S., Zeilberger, D. (1996). $A = B$. Boca Raton, FL: CRC Press. http://www.math.upenn.edu/~wilf/AeqB.html

Summary. The Chu-Vandermonde convolution formula is a classic result on binomial coefficients with many known proofs. We provide a short and novel proof by using the method of finite differences.

WENCHANG CHU (MR Author ID: 213991) received his PhD in mathematics from Dalian University of Technology. After that he migrated to Europe and became a professor of mathematics at the University of Salento in Lecce, Italy. His research interests include combinatorial analysis and special functions, highlighted by hundreds of publications in both classical hypergeometric series and q-series.

Report on the 62nd Annual International Mathematical Olympiad

BÉLA BAJNOK
Gettysburg College
Gettysburg, PA 17325
bbajnok@gettysburg.edu

The International Mathematical Olympiad (IMO) is the world's leading mathematics competition for high school students and is organized annually by different host countries. The competition consists of three problems each on two consecutive days, with an allowed time of four and a half hours both days. In the recent years, more than one hundred countries have sent teams of up to six students to compete.

The 62nd IMO was organized by Russia; due to the COVID-19 pandemic, the competition was held online. The six members of the US team participated virtually on July 19 and July 20, 2021, at universally coordinated times in four different time zones.

The members of the US team are chosen during the Math Olympiad Program (MOP) each year, a year-long endeavor organized by the MAA's American Mathematics Competitions (AMC) program. Students gain admittance to MOP based on their performance on a series of examinations, culminating in the USA Mathematical Olympiad (USAMO). A report on the 2021 USAMO can be found in the February 2022 issue of this *Magazine*; a similar report on the 2021 USA Junior Mathematical Olympiad appeared in the January 2022 issue of the *College Mathematics Journal*. More information on the American Mathematics Competitions program can be found online [**1**].

The members of the 2021 US team were Ankit Bisain (12th grade, Canyon Crest Academy, San Diego, CA); Quanlin Chen (12th grade, Princeton International School of Mathematics and Science, Princeton, NJ); Maxim Li (12th grade, Okemos High School, Okemos, MI); Luke Robitaille (11th grade, Robitaille Homeschool, Euless, TX); Noah Walsh (12th grade, Walsh Academy, Hillsboro, OR); and Zifan Wang (11th grade, Princeton International School of Mathematics and Science, Princeton, NJ). Li, Robitaille, Walsh, and Wang each earned Gold Medals, and Bisain and Chen each earned a Silver Medal. In the unofficial ranking of countries, the United States finished fourth after China (first), Russia (second), and South Korea (third).

Below we present the problems and solutions of the 62nd IMO. The solutions are those of the current author, utilizing some of the various sources already available. (The online videos by RedPig [**4**] were particularly helpful with Problems 2 and 3). Each problem was worth 7 points. The nine-tuple $(a_7, a_6, a_5, a_4, a_3, a_2, a_1, a_0;$ **a**$)$ states the number of students who scored $7, 6, \ldots, 0$ points, respectively, followed by the mean score achieved for the problem. The combined mean total score of all participants was 11.59, making this the most difficult competition in the history of the IMO.

Problem 1 (286, 36, 38, 41, 10, 41, 36, 131; **4.39**); *proposed by Australia.* Let $n \geqslant 100$ be an integer. Ivan writes the numbers $n, n+1, \ldots, 2n$ each on different cards. He then shuffles these $n+1$ cards, and divides them into two piles. Prove that at least one of the piles contains two cards such that the sum of their numbers is a perfect square.

Math. Mag. **95** (2022) 145–153. doi:10.1080/0025570X.2022.2026093 © Mathematical Association of America
MSC: 00-01

SOLUTION. To prove our claim, it suffices to show that there exist three cards such that the sum of the numbers on any two of them is a perfect square. If this is the case, then the pigeonhole principle implies that Ivan must have placed at least two of these three cards in the same pile, as needed.

Therefore, we want to prove that there exist $x, y, z \in \{n, n+1, \ldots, 2n\}$ such that $x + y = a^2$, $y + z = b^2$, and $z + x = c^2$ for some integers a, b, and c. To facilitate having x, y, and z all fall within the given range, we aim for a, b, and c to be consecutive integers, and since $a^2 + b^2 + c^2 = 2(x + y + z)$ is even, we need the middle square to be even. So let us assume that $a = 2t - 1$, $b = 2t$, and $c = 2t + 1$ for some positive integer t. These equations then yield $x = 2t^2 + 1$, $y = 2t^2 - 4t$, and $z = 2t^2 + 4t$. Therefore, our method is successful if for every integer $n \geqslant 100$ we can find an integer t such that

$$n \leq 2t^2 - 4t < 2t^2 + 1 < 2t^2 + 4t \leq 2n,$$

or, equivalently,

$$\sqrt{\frac{n}{2} + 1} + 1 \leq t \leq \sqrt{n+1} - 1.$$

We can verify that the greatest integer satisfying the second inequality also satisfies the first. Indeed, if $t = \lfloor \sqrt{n+1} \rfloor - 1$, then $t > \sqrt{n+1} - 2$, and thus

$$\frac{n}{2} + 1 = \frac{1}{2}(n+1) + \frac{1}{2} < \frac{1}{2}(t+2)^2 + \frac{1}{2} = (t-1)^2 - \frac{1}{2}(t+1)(t-9) - 3.$$

Since $n \geq 100$ implies that $t \geq 9$, we have $\frac{n}{2} + 1 < (t-1)^2$, from which $\sqrt{\frac{n}{2} + 1} + 1 \leq t$, as claimed.

Problem 2 (16, 2, 1, 3, 2, 12, 61, 522; **0.38**); *proposed by Canada.* Show that the inequality

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \sqrt{|x_i - x_j|} \leq \sum_{i=1}^{n} \sum_{j=1}^{n} \sqrt{|x_i + x_j|}$$

holds for all real numbers $x_1, x_2, \ldots, x_n$.

SOLUTION. For given real numbers $x_1, x_2, \ldots, x_n$, we let $X = \{x_1, \ldots, x_n\}$, and set

$$L(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sqrt{|x_i - x_j|}$$

and

$$R(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sqrt{|x_i + x_j|}.$$

The inequality $L(X) \leq R(X)$ clearly holds when $n = 0$ or $n = 1$ since for $X = \emptyset$ we have $L(X) = R(X) = 0$, and for $X = \{x_1\}$ we have $L(X) = 0$ and $R(X) = \sqrt{|2x_1|}$. We proceed by induction, and prove that if the inequality $L(X) \leq R(X)$ holds for all sets X with $n - 2$ or $n - 1$ elements, then it also holds when $|X| = n$.

We first consider the case when X contains 0, say $x_1 = 0$. In this case,

$$L(X) = 2\sum_{i=2}^{n} \sqrt{|x_i|} + L(X \setminus \{0\})$$

and

$$R(X) = 2\sum_{i=2}^{n} \sqrt{|x_i|} + R(X \setminus \{0\}),$$

and thus $L(X) \leq R(X)$ by induction.

The situation is only slightly more complicated when X contains two elements that add to 0, say $x_1 + x_2 = 0$. We see that

$$L(X) = 2\sqrt{|2x_1|} + 2\sum_{i=3}^{n} \sqrt{|x_1 - x_i|} + 2\sum_{i=3}^{n} \sqrt{|x_2 - x_i|} + L(X \setminus \{x_1, x_2\})$$

and

$$R(X) = 2\sqrt{|2x_1|} + 2\sum_{i=3}^{n} \sqrt{|x_1 + x_i|} + 2\sum_{i=3}^{n} \sqrt{|x_2 + x_i|} + R(X \setminus \{x_1, x_2\}).$$

Since $|x_1 - x_i| = |x_2 + x_i|$ and $|x_2 - x_i| = |x_1 + x_i|$ for each $3 \leq i \leq n$, our inequality $L(X) \leq R(X)$ follows again by induction.

The two cases we have examined so far can be combined by saying that there are not-necessarily-distinct indices $1 \leq i, j \leq n$ for which $x_i + x_j = 0$; we will refer to such sets X as *reducible*.

Before proceeding further, we note that the graph of the function $f(x) = \sqrt{x}$ is *concave down* over its entire domain. That is, for all nonnegative real numbers A and B, the secant line connecting the points $(A, \sqrt{A})$ and $(B, \sqrt{B})$ lies under the curve between these two points. We can express this fact algebraically to say that for all nonnegative real numbers A and B, and for all nonnegative weights λ_1 and λ_2 that add to 1, we have

$$\lambda_1\sqrt{A} + \lambda_2\sqrt{B} \leq \sqrt{\lambda_1 A + \lambda_2 B}. \tag{1}$$

(One can readily verify this inequality by noting that the square of the left-hand side can be written as

$$\lambda_1(1 - \lambda_2)A + \lambda_2(1 - \lambda_1)B + 2\lambda_1\lambda_2\sqrt{AB},$$

which equals

$$\lambda_1 A + \lambda_2 B - \lambda_1\lambda_2(\sqrt{A} - \sqrt{B})^2,$$

and this quantity is no greater than the square of the right-hand side. (This inequality is a special case of *Jensen's Inequality*; see, for example, Section 3.5 in the classic book by Hardy, Littlewood, and Polya [3].)

Suppose now that X is any set of n real numbers $x_1, x_2, \ldots, x_n$. We may assume that X is not reducible, and thus $x_i + x_j$ is nonzero for all $1 \leq i, j \leq n$. We introduce the set P of ordered pairs (i, j) for which $x_i + x_j$ is positive and the set N of ordered pairs (i, j) for which $x_i + x_j$ is negative $(1 \leq i, j \leq n)$. We then define d_P to be the minimum value of $x_i + x_j$ as $(i, j) \in P$ and d_N to be the minimum value of $-(x_i + x_j)$ as $(i, j) \in N$.

Observe that for all $(i, j) \in P$, we have $x_i + x_j - d_P \geq 0$ and $x_i + x_j + d_N \geq 0$, while for all $(i, j) \in N$, we have $x_i + x_j - d_P \leq 0$ and $x_i + x_j + d_N \leq 0$. Therefore,

$$R\left(X - \frac{d_P}{2}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sqrt{|x_i + x_j - d_P|}$$
$$= \sum_{(i,j)\in P}\sqrt{|x_i + x_j| - d_P} + \sum_{(i,j)\in N}\sqrt{|x_i + x_j| + d_P},$$

and

$$R\left(X + \frac{d_N}{2}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sqrt{|x_i + x_j + d_N|}$$
$$= \sum_{(i,j)\in P}\sqrt{|x_i + x_j| + d_N} + \sum_{(i,j)\in N}\sqrt{|x_i + x_j| - d_N}.$$

(As usual, $X + r$ denotes the translation of X by the real number r, defined as the collection $\{x + r \mid x \in X\}$.)

We now use (1) to find that

$$\frac{d_N}{d_N + d_P} \cdot \sqrt{|x_i + x_j| - d_P} + \frac{d_P}{d_N + d_P} \cdot \sqrt{|x_i + x_j| + d_N} \leq \sqrt{|x_i + x_j|}$$

holds for each $(i, j) \in P$, and that

$$\frac{d_N}{d_N + d_P} \cdot \sqrt{|x_i + x_j| + d_P} + \frac{d_P}{d_N + d_P} \cdot \sqrt{|x_i + x_j| - d_N} \leq \sqrt{|x_i + x_j|}$$

holds for each $(i, j) \in N$. Therefore,

$$\frac{d_N}{d_N + d_P} \cdot R\left(X - \frac{d_P}{2}\right) + \frac{d_P}{d_N + d_P} \cdot R\left(X + \frac{d_N}{2}\right) \leq R(X). \tag{2}$$

Since both $X - \frac{d_P}{2}$ and $X + \frac{d_N}{2}$ are reducible by definition (when they are not empty), we have

$$L\left(X - \frac{d_P}{2}\right) \leq R\left(X - \frac{d_P}{2}\right)$$

and

$$L\left(X + \frac{d_N}{2}\right) \leq R\left(X + \frac{d_N}{2}\right),$$

and since $L(X) = L(X + r)$ for any $r \in \mathbb{R}$, from (2) we get $L(X) \leq R(X)$, completing our proof.

Problem 3 (15, 0, 0, 1, 1, 4, 110, 488; **0.37**) *proposed by Ukraine.* Let D be an interior point of the acute triangle ABC with $AB > AC$ so that $\angle DAB = \angle CAD$. The point E on the segment AC satisfies $\angle ADE = \angle BCD$, the point F on the segment AB satisfies $\angle FDA = \angle DBC$, and the point X on the line AC satisfies $CX = BX$. Let O_1 and O_2 be the circumcenters of the triangles ADC and EXD respectively. Prove that the lines BC, EF, and O_1O_2 are concurrent.

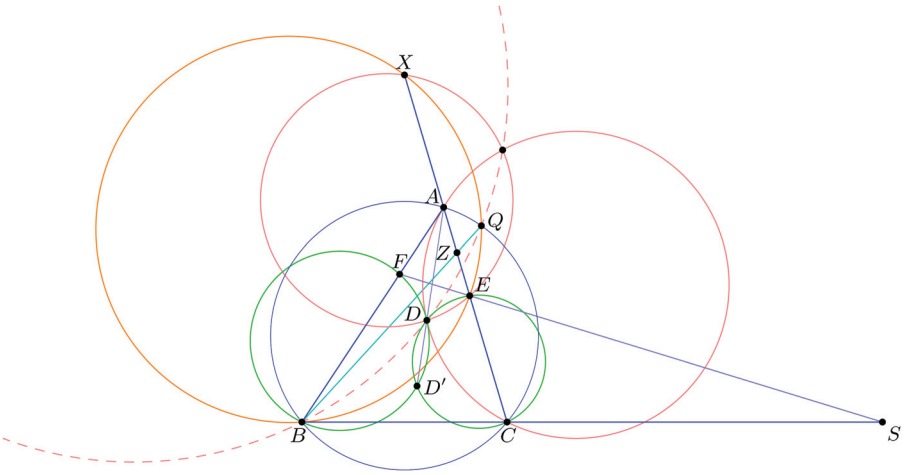


Figure 1 Some of the many circles of Problem 3

Solution. Denote the intersection of lines BC and EF by S. Our goal is to prove that the points O_1, O_2, and S are collinear. (Figure 1 will be helpful in following the argument.)

Recall that if we reflect the lines AD, BD, and CD about the angle bisectors of triangle ABC at vertices A, B, and C, respectively, then the three reflection lines meet at the *isogonal conjugate* of D, denoted here by D'. Since AD is the angle bisector at A, the points A, D, and D' are collinear. (For an excellent introduction to some of the more advanced topics and results in geometry used in this solution, such as isogonal conjugates, the radical axes concurrence theorem, Miquel's quadrilateral theorem, and circular inversion, see Chen [2].)

We have $\angle DBC = \angle D'BA$, so the condition that $\angle FDA = \angle DBC$ implies that $\angle FDD'$ and $\angle FBD'$ are supplementary, and thus that $BD'DF$ is a cyclic quadrilateral. The power of A to the circumcircle of $BD'DF$ equals

$$AF \cdot AB = AD \cdot AD'.$$

Similarly, $CD'DE$ is a cyclic quadrilateral, and thus $AE \cdot AC = AD \cdot AD'$. But then

$$AF \cdot AB = AE \cdot AC,$$

which implies that $BCEF$ is also a cyclic quadrilateral.

Our next goal is to show that DS is the tangent line both to the circumcircle of $\triangle BCD$ and to the circumcircle of $\triangle EFD$.

First, we prove that the two circles are tangent to one another at D. Let G be the intersection point of lines AD and EF, and let H be the intersection point of lines AD and BC.

Since $BCEF$ is a cyclic quadrilateral, we have that $\angle FBC$ and $\angle FEA$ are equal. Considering $\triangle ABH$ and $\triangle AGE$, we see that they have two pairs of angles that are equal. Therefore, they are similar, and thus $\angle AGE = \angle BHA$. By the same reasoning about $\triangle FGD$ and $\triangle BHD$, we have $\angle DFG = \angle HDB$. Now

$$\begin{aligned}\angle EDC &= 180^\circ - \angle ADE - \angle HDC \\ &= 180^\circ - \angle HCD - \angle HDC = \angle DHC,\end{aligned}$$

and

$$\begin{aligned}\angle DHC &= 180^\circ - \angle BHD \\ &= \angle DBH + \angle HDB = \angle DBH + \angle DFG.\end{aligned}$$

Therefore, we have

$$\angle EDC = \angle DBH + \angle DFG.$$

But $\angle DBH = \angle DBC$ is the angle between DC and the tangent line to the circumcircle of $\triangle BCD$ at D, and $\angle DFG = \angle DFE$ is the angle between DE and the tangent line to the circumcircle of $\triangle EFD$ at D. Since the sum of these two angles equals $\angle EDC$, the two tangent lines must coincide. We let ℓ be this common tangent line.

Recall now that, given two nonconcentric circles, the collection of points that have equal power with respect to the two circles is a line, called the *radical axis* of the circles. Furthermore, if the two circles intersect at a single point, then their radical axis is the common tangent line at that point, and if the two circles intersect in two points, then their radical axis is the line through these two points. Therefore, the radical axis of the circumcircles of $\triangle DBC$ and $\triangle DEF$ is ℓ, the radical axis of the circumcircles of triangle DBC and quadrilateral $BCEF$ is BC, and the radical axis of the circumcircles of triangle DEF and quadrilateral $BCEF$ is EF. By the radical axis concurrence theorem, the three pairwise radical axes of three circles concur. That is, ℓ, BC, and EF all meet at S.

Next, we recall *Miquel's Quadrilateral Theorem* which, applied to our situation, says that the circumcircles of $\triangle ABC$, $\triangle AEF$, $\triangle SBF$, and $\triangle SCE$ intersect at a single point, called the *Miquel point* of quadrilateral $BCEF$. We denote this point by Q.

Since $AFEQ$ is a cyclic quadrilateral, angles $\angle AQE$ and $\angle AFE$ are supplementary, and thus $\angle AQE = \angle BFE$. Since $\angle BFE$ and $\angle BCE$ are supplementary, $\angle BFE = \angle ECS$. But quadrilateral $CSQE$ is also cyclic, so $\angle ECS$ and $\angle EQS$ are supplementary as well, and therefore points A, Q, and S are collinear.

Next, we prove that quadrilateral $XQEB$ is cyclic, adding to our considerable collection of cyclic quadrilaterals. Since A, Q, and S are collinear, we can write

$$\angle QSC = \angle ASC = 180^\circ - \angle ACS - \angle SAC = \angle ACB - \angle QAC.$$

The assumption that $CX = BX$ implies then that

$$\angle QSC = \angle XBC - \angle QAC.$$

Since $QECS$ is cyclic, the angles $\angle QSC$ and $\angle QEC$ are supplementary, and thus $\angle QSC = \angle XEQ$; and since $ABCQ$ is cyclic, we have $\angle QAC = \angle QBC$. Therefore,

$$\angle XEQ = \angle XBC - \angle QBC,$$

and thus $\angle XEQ = \angle XBQ$, implying that quadrilateral $XQEB$ is indeed cyclic.

Let us now consider the two circles in the problem statement: the circumcircles of $\triangle ADC$ and $\triangle EXD$, which we denote by ω_1 and ω_2, respectively. The two circles meet at D. Let R be their second intersection point. Also, let Z be the intersection point of segments BQ and AC. Since quadrilaterals $ABCQ$ and $XQEB$ are cyclic, we have

$$ZA \cdot ZC = ZB \cdot ZQ \qquad \text{and} \qquad ZB \cdot ZQ = ZE \cdot ZX.$$

Therefore,

$$ZA \cdot ZC = ZE \cdot ZX,$$

meaning that the power of Z to ω_1 equals the power of Z to ω_2. This means that Z lies on the radical axis DR of ω_1 and ω_2. This, in turn, means that the power of Z to each o f the two circles can also be written as $ZD \cdot ZR$. Since

$$ZD \cdot ZR = ZB \cdot ZQ,$$

we see that $BDQR$ is yet another cyclic quadrilateral.

Our final step is to use circular inversion. Note first that, since the line SD is tangent at D both to the circumcircle of $\triangle BCD$ and to the circumcircle of $\triangle EFD$, we have

$$SD^2 = SB \cdot SC = SF \cdot SE.$$

Since $AFEQ$ is also cyclic, we have

$$SF \cdot SE = SA \cdot SQ.$$

Let τ be the circular inversion centered at S and of radius SD. Our equations then imply that $\tau(B) = C$ and $\tau(Q) = A$. Obviously $\tau(D) = D$. Since the circular inversion maps a circle (not passing through the center of the inversion) to a circle, we thus find that τ maps the circumcircle of $\triangle BQD$ to the circumcircle of $\triangle CAD$.

Recall that quadrilateral $BDQR$ is cyclic, so R lies on the circumcircle of $\triangle BDQ$, and R is on the circumcircle of $\triangle CAD$ by definition. Therefore, R is a fixed point of the inversion, which means that $SR = SD$, and thus S lies on the perpendicular bisector of DR, which is the line O_1O_2. This proves that points O_1, O_2, and S are collinear, as claimed.

Problem 4 (309, 5, 1, 12, 2, 39, 33, 218; **3.82**) *proposed by Poland.* Let Γ be a circle with center I, and $ABCD$ a convex quadrilateral such that each of the segments AB, BC, CD, and DA is tangent to Γ. Let Ω be the circumcircle of the triangle AIC. The extension of BA beyond A meets Ω at X, and the extension of BC beyond C meets Ω at Z. The extensions of AD and CD beyond D meet Ω at Y and T, respectively. Prove that

$$AD + DT + TX + XA = CD + DY + YZ + ZC. \tag{3}$$

Solution. Consult Figure 2 while following the solution. Since $XYIA$ is a cyclic quadrilateral, we have that

$$\angle YXI = \angle YAI \qquad \text{and} \qquad \angle XYI = \angle IAB.$$

Since AI is the angle bisector of $\angle YAB$, we also have $\angle YAI = \angle IAB$. Therefore, $\angle YXI = \angle XYI$, and thus $\triangle XYI$ is an isosceles triangle. By an analogous argument, $\triangle ZTI$ is an isosceles triangle as well, and thus $XT = YZ$. This makes (3) equivalent to

$$AD + DT + XA = CD + DY + ZC. \tag{4}$$

Now let P, Q, R, and S be the points where lines AB, BC, CD, and DA meet Γ, respectively. Since $IX = IY$ and $IP = IS$, right triangles XPI and YSI are congruent, and so $XP = YS$. Similarly, we find that $TR = ZQ$. Adding these two equations, we get

$$XP + TR = YS + ZQ,$$

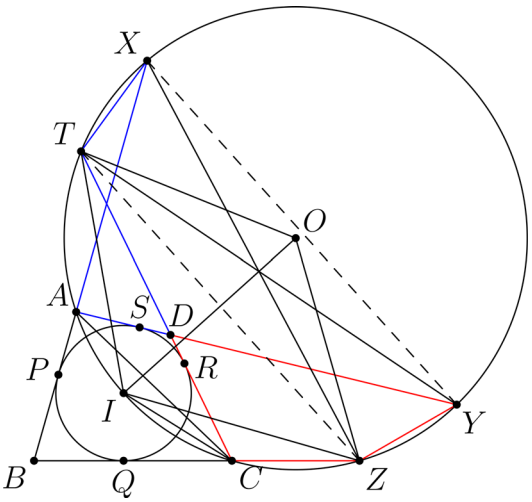


Figure 2 Illustration for Problem 4

or

$$XA + AP + TD + DR = YD + DS + ZC + CQ.$$

Now $AP = AS$, $DR = DS$, and $CQ = CR$, so this equation is equivalent to

$$XA + AS + SD + TD = YD + DR + RC + ZC,$$

which yields (4).

Problem 5 (175, 4, 5, 2, 4, 13, 12, 404; **2.15**); *proposed by Spain.* Two squirrels, Bushy and Jumpy, have collected 2021 walnuts for the winter. Jumpy numbers the walnuts from 1 through 2021, and digs 2021 little holes in a circular pattern in the ground around their favorite tree. The next morning Jumpy notices that Bushy had placed one walnut into each hole, but had paid no attention to the numbering. Unhappy, Jumpy decides to reorder the walnuts by performing a sequence of 2021 moves. In the k-th move, Jumpy swaps the positions of the two walnuts adjacent to walnut k. Prove that there exists a value of k such that, on the k-th move, Jumpy swaps some walnuts a and b such that $a < k < b$.

SOLUTION. For a given arrangement of the walnuts, and for each $1 \leq k \leq 2021$, we let n_k be the number of pairs among walnuts numbered $1, 2, \ldots, k$ that are immediate neighbors. Clearly, for any arrangement of the 2021 walnuts, we have $n_1 = 0$ and $n_{2021} = 2021$.

Suppose now that Jumpy is about to swap the neighbors a and b of walnut k. If a and b are both less than k, then we have $n_k = n_{k-1} + 2$ both before and after the swap, and if a and b are both greater than k, then we have $n_k = n_{k-1}$, again both before and after the swap. Therefore, if we were in one of these two situations at every stage, then n_0 and n_{2021} would have the same parity. Since that is not the case, our claim holds.

Problem 6 (37, 0, 2, 1, 3, 2, 12, 562; **0.48**); *proposed by Austria.* Let $m \geq 2$ be an integer, A be a finite set of (not necessarily positive) integers, and $B_1, B_2, B_3, \ldots, B_m$ be subsets of A. Assume that for each $k = 1, 2, \ldots, m$ the sum of the elements of B_k is m^k. Prove that A contains at least $m/2$ elements.

SOLUTION. Let n be the number of elements of A. We denote these elements (in any order) by $a_1, a_2, \ldots, a_n$. For each $i = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, m$, we let $\epsilon_{i,k}$ be the indicator variable that equals 1 if $a_i \in B_k$ and 0 otherwise. This allows us to take the condition that the elements of B_k sum to m^k and rewrite it algebraically as

$$\epsilon_{1,k}a_1 + \epsilon_{2,k}a_2 + \cdots + \epsilon_{n,k}a_n = m^k. \tag{5}$$

Recall now that every nonnegative integer has a unique base-m representation. In particular, for each nonnegative integer N that is less than m^{m+1} and is divisible by m we have unique nonnegative integers $d_1, d_2, \ldots, d_m$, all less than m, such that

$$N = d_m m^m + d_{m-1}m^{m-1} + \cdots + d_1 m.$$

Using (5) for each $k = 1, 2, \ldots, m$ allows us to write N in the form

$$N = c_1a_1 + c_2a_2 + \cdots + c_na_n,$$

where

$$c_i = \epsilon_{i,1}d_1 + \epsilon_{i,2}d_2 + \cdots + \epsilon_{i,m}d_m$$

for $i = 1, 2, \ldots, n$. Here each c_i is a nonnegative integer that is at most $m(m-1)$. It follows that there are at most $(m^2 - m + 1)^n$ integers that can be written in this form. But N was any of the m^m nonnegative integers divisible by m and less than m^{m+1}, so we must have

$$m^m \leq (m^2 - m + 1)^n \leq m^{2n},$$

implying $n \geq m/2$, as desired.

Acknowledgments I am very grateful to Jerrold Grossman, Chris Jeuell, Enrique Treviño, and Paul Zeitz for their willingness to proofread parts of this article and for doing it so carefully. Credits for Figures 1 and 2 go to Eric Shen and Rey Li, respectively.

REFERENCES

[1] American Mathematics Competitions. https://www.maa.org/math-competitions
[2] Chen, E. (2016). *Euclidean Geometry in Mathematical Olympiads*. Washington, DC: Mathematical Association of America.
[3] Hardy, G. H., Littlewood, J. E., Pólya, G. (1952). *Inequalities*, 2nd ed. London: Cambridge University Press.
[4] Maths RedPig Olympiad. https://www.youtube.com/c/redpigmathsolympiads

Summary. We present the problems and solutions to the 62nd International Mathematical Olympiad.

BÉLA BAJNOK (MR Author ID: 314851) is a professor of mathematics at Gettysburg College. He is the Director of the American Mathematics Competitions program of the MAA.

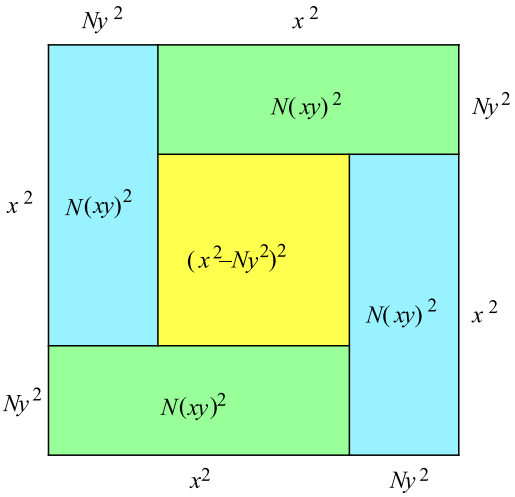
Solutions to the Pell Equation $x^2 - Ny^2 = 1$

ROGER B. NELSEN
Lewis & Clark College
Portland, OR 97219
rnelsen@lclark.edu

The *Pell equation* $x^2 - Ny^2 = 1$ is a quadratic *Diophantine equation* (an equation with integer coefficients in which the unknowns are also integers) in which N is a positive non-square integer. In 1768, Joseph-Louis Lagrange proved that this equation has a solution in positive integers for every N [2]. The following theorem then establishes that Pell equations have infinitely many solutions in positive integers. This wordless proof employs a frequently used diagram [1, 3–5].

Theorem 1. *If $x^2 - Ny^2 = 1$, where N is positive and non-square, has a solution in positive integers, then it has infinitely many.*

Proof. We show that if (x, y) is a solution, then so is $(x^2 + Ny^2, 2xy)$.



$$(x^2 - Ny^2)^2 = (x^2 + Ny^2)^2 - 4N(xy)^2;$$

$$\therefore\ x^2 - Ny^2 = 1 \Longrightarrow (x^2 + Ny^2)^2 - N(2xy)^2 = 1.$$

Example When $N = 3$, we have that $(x, y) = (2, 1)$ is a solution, and the recursion in the proof of the theorem yields the following infinite sequence of solutions:

$$(2, 1) \to (7, 4) \to (97, 56) \to (18817, 10864) \to \cdots.$$

However, this sequence does not contain all of the solutions to $x^2 - 3y^2 = 1$. ■

Math. Mag. **95** (2022) 154–155. doi:10.1080/0025570X.2022.2023308 © Mathematical Association of America
MSC: 11D09
This article was originally published with errors, which have now been corrected in the online version. Please see Correction (http://dx.doi.org/10.1080/0025570X.2022.2065164).

REFERENCES

[1] Alsina, C., Nelsen, R. B. (2011). Inequalities for two numbers whose sum is one. *Math. Mag.* 84(3): 228. doi.org/10.4169/math.mag.84.3.228
[2] Lenstra Jr., H. W. (2002). Solving the Pell equation. *Notices Amer. Math. Soc.* 49: 182–192.
[3] Nelsen, R. B. (1994). Proof without words: The sum of a positive number and its reciprocal is at least two (four proofs). *Math. Mag.* 67(5): 374. doi.org/10.1080/0025570X.1994.11996253
[4] Nelsen, R. B. (2018). Periodic continued fractions via a proof without words. *Math. Mag.* 91(5): 364–365. doi.org/10.1080/0025570X.2018.1456151
[5] Schattschneider, D. (1986). Proof without words: The arithmetic mean-geometric mean inequality. *Math. Mag.* 59(1): 11. doi.org/10.1080/0025570X.1986.11977213

Summary. We show that if the Pell equation $x^2 - Ny^2 = 1$ has a solution in positive integers, then it has infinitely many.

ROGER B. NELSEN (MR Author ID: 237909, ORCID: 0000-0002-3632-6577) is a professor emeritus at Lewis & Clark College, where he taught mathematics and statistics for 40 years.

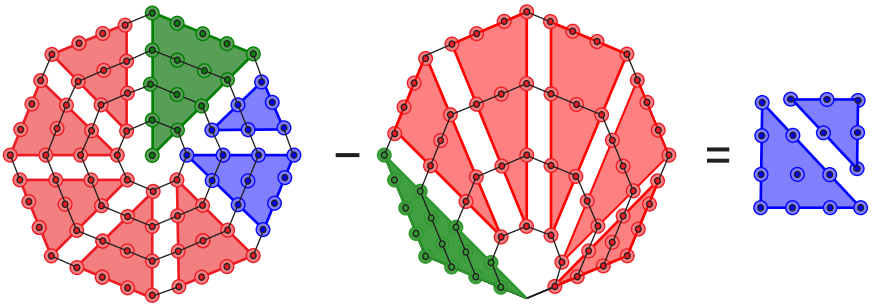
A Polygonal Number Identity

GÜNHAN CAGLAYAN
New Jersey City University
Jersey City, NJ 07305
gcaglayan@ncju.edu

Theorem 1. *Let $C_{k,n}$ denote the centered k-gonal number of side length n. Let $P_{k,n}$ denote the regular k-gonal number of side length n [1]. Then we have*

$$C_{k,n} - P_{k,n} = (n-1)^2.$$

Proof. The proof is shown for $n = 5$ and $k = 8$.



■

In general, expressing $C_{k,n}$ and $P_{k,n}$ in terms of triangular numbers, we have

$$\begin{aligned} C_{k,n} - P_{k,n} &= [T_n + (k-2)T_{n-1} + T_{n-2}] - [T_n + (k-3)T_{n-1}] \\ &= T_{n-1} + T_{n-2} = (n-1)^2. \end{aligned}$$

Acknowledgments The author wishes to thank Mathematics Magazine editorial board and anonymous reviewers for their helpful comments on earlier drafts.

REFERENCE

[1] Deza, E., Deza, M. (2012). *Figurate Numbers*. Singapore: World Scientific.

Summary. Let $C_{k,n}$ and $P_{k,n}$ represent, respectively, the centered k-gonal number of side length n, and the regula k-gonal number of side n. We provide a visual proof that $C_{k,n} - P_{k,n}$ is always a perfect square.

GÜNHAN CAGLAYAN (MR Author ID: 1116420) teaches mathematics at New Jersey City University. His main interests are visual mathematics and student learning through modeling and visualization.

*The online version of this article includes color diagrams.

PROBLEMS

LES REID, *Editor*
Missouri State University

EUGEN J. IONAŞCU, *Proposals Editor*
Columbus State University

RICHARD BELSHOFF, Missouri State University; *MAHYA GHANDEHARI*, University of Delaware; *EYVINDUR ARI PALSSON*, Virginia Tech; *GAIL RATCLIFF*, East Carolina University; *ROGELIO VALDEZ*, Centro de Investigación en Ciencias, UAEM, Mexico; *Assistant Editors*

Proposals

To be considered for publication, solutions should be received by September 1, 2022.

2141. *Proposed by Paul Bracken, University of Texas Rio Grande Valley, Edinburg, TX.*

Evaluate

$$\int_0^\infty \ln\left(1 + 2x^{-2}\cos\varphi + x^{-4}\right)\, dx.$$

2142. *Proposed by Roger Izard, Dallas, TX.*

Given a parabola in the plane, find its axis and focus using compass and straightedge.

2143. *Proposed by Florin Stănescu, Şerban Cioculescu School, Găeşti, Romania.*

Evaluate

$$\lim_{n\to\infty} \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{(-1)^{i+j}}{2i+2j+1} \binom{n+i}{n-i} \binom{n+j}{n-j}.$$

2144. *Proposed by Souvik Dey (graduate student), University of Kansas, Lawrence, KS.*

Let R be a finite commutative ring with unity such that distinct ideals of R have distinct orders. Show that R is a principal ideal ring.

Math. Mag. **95** (2022) 157–166. doi:10.1080/0025570X.2022.2026096 © Mathematical Association of America

We invite readers to submit original problems appealing to students and teachers of advanced undergraduate mathematics. Proposals must always be accompanied by a solution and any relevant 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.

Proposals and solutions should be written in a style appropriate for this MAGAZINE.

Authors of proposals and solutions should send their contributions using the Magazine's submissions system hosted at `http://mathematicsmagazine.submittable.com`. *More detailed instructions are available there. We encourage submissions in PDF format, ideally accompanied by LaTeX source. General inquiries to the editors should be sent to* `mathmagproblems@maa.org`.

2145. *Proposed by the Missouri State University Problem Solving Group, Missouri State University, Springfield, MO.*

Given a set of points S, let $L(S)$ be the set of all points lying on any line connecting two distinct points in S. For example, if S is the disjoint union of a closed line segment and a point not lying on the line containing the segment, then $L(S)$ consists of two vertical angles, their interiors, and the line containing the segment. In this case, $L(L(S))$ is the entire plane.

Determine $L(L(S))$ when S consists of the vertices of a regular tetrahedron.

Quickies

1119. *Proposed by George Stoica, Saint John, NB, Canada.*

A standard deck of 52 playing cards is shuffled. The cards are turned over one by one from the top of the deck until the third ace appears. What is the expected value of its position?

1120. *Proposed by Hideyuki Ohtsuka, Saitama, Japan.*

Given $\triangle ABC$, let $a = BC$, $b = AC$, $c = AB$, and r be the triangle's inradius. If X is a point in the triangle's interior, let

$$R_1 = AX,\ R_2 = BX, \quad \text{and} \quad R_3 = CX.$$

Prove that

$$\frac{aR_1 + bR_2 + cR_3}{a + b + c} \geq 2r.$$

When does equality hold?

Solutions

Evaluating an improper integral — April 2021

2116. *Proposed by Fook Sung Wong, Temasek Polytechnic, Singapore.*

Evaluate

$$\int_0^\infty \frac{e^{\cos x}\cos(\alpha x + \sin x)}{x^2 + \beta^2}\,dx,$$

where α and β are positive real numbers.

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

We claim the answer is $\dfrac{\pi}{2\beta}\exp(e^{-\beta} - \alpha\beta)$.

Consider the meromorphic function

$$F(z) = \frac{g(z)}{z^2 + \beta^2}, \quad \text{where } g(z) = \exp(e^{iz} + i\alpha z).$$

If $z = x + iy$ with $x, y \in \mathbb{R}$ and $y \geq 0$, then

$$|g(z)| = \exp\left(\mathrm{Re}(e^{iz} + i\alpha z)\right) = \exp\left(e^{-y}\cos(x) - \alpha y\right) \leq \exp\left(e^{-y} - \alpha y\right) \leq e.$$

For $R > \beta$, consider the closed contour Γ_R consisting of the line segment $[-R, R]$ followed by the semicircle γ_R parametrized by $\theta \mapsto Re^{i\theta}$ for $\theta \in [0, \pi]$. The only singularity that F has inside the domain bounded by Γ_R is a simple pole at $z = i\beta$ with residue

$$\mathrm{Res}\,(F, i\beta) = \frac{g(i\beta)}{2i\beta} = \frac{\exp(e^{-\beta} - \alpha\beta)}{2i\beta}.$$

By the residue theorem we have

$$\int_{\Gamma_R} F(z)dz = 2i\pi\,\mathrm{Res}\,(F, i\beta) = \frac{\pi}{\beta}\exp(e^{-\beta} - \alpha\beta).$$

But

$$\begin{aligned}\int_{\Gamma_R} F(z)dz &= \int_{-R}^{R} F(x)dx + \int_{\gamma_R} F(z)dz \\ &= 2\int_0^R \frac{e^{\cos x}\cos(\alpha x + \sin x)}{x^2 + \beta^2}dx + \epsilon_R,\end{aligned}$$

where

$$\epsilon_R = \int_{\gamma_R} F(z)dz.$$

Since $R > \beta$, we have

$$|\epsilon_R| \leq \pi R \sup_{\theta\in[0,\pi]} |F(Re^{i\theta})| \leq \pi e \frac{R}{R^2 - \beta^2}.$$

Thus $\lim_{R\to\infty} \epsilon_R = 0$. Therefore

$$2\int_0^\infty \frac{e^{\cos x}\cos(\alpha x + \sin x)}{x^2 + \beta^2}dx = \frac{\pi}{\beta}\exp(e^{-\beta} - \alpha\beta),$$

as claimed.

Also solved by Khristo N. Boyadzhiev, Hongwei Chen, John Fitch, G. C. Greubel, Eugene A. Herman, Rafe Jones, Kee-Wai Lau (Hong Kong), Kelly D. McLenithan, Raymond Mortini (France) & Rudolf Rupp (Germany), Moubinool Omarjee (France), Didier Pinchon (France), Ahmad Sabihi (Iran), Albert Stadler (Switzerland), Seán M. Stewart (Australia), and the proposer. There were two incomplete or incorrect solutions.

A factorial Diophantine equation **April 2021**

2117. *Proposed by Ahmad Sabihi, Isfahan, Iran.*

Find all positive integer solutions to the equation

$$(m+1)^n = m! + 1.$$

Solution by Michael Kardos (student), East Carolina University, Greenville, NC.
Note that for any solution with $m \geq 2$, we have m even. This follows from the fact that $m!$ is even for $m \geq 2$ and so $m! + 1 = (m+1)^n$ is odd. Thus $(m+1)$ must be odd and m even.

We can reduce the pool of possible solutions by showing that $m \leq 4$. Clearly a solution with $m > 4$ and m even implies $2 < m/2 < m$, so

$$2\left(\frac{m}{2}\right) m \big| m! \Rightarrow m^2 \big| \left((m+1)^n - 1\right) \Rightarrow m^2 \big| m \sum_{k=1}^{n} \binom{n}{k} m^{k-1}.$$

Thus

$$m \Big| \left(n + m \sum_{k=2}^{n} \binom{n}{k} m^{k-2} \right),$$

so m divides n and hence $n \geq m$.

We will now show that $n \geq m$ and $m > 4$ yields no solutions. In that case,

$$m! + 1 < m^{m-1} + 1 < (m+1)m^{m-1} < (m+1)^m \leq (m+1)^n.$$

Thus, there are no positive integer solutions with $m > 4$. Now we can find all solutions using the previously gathered information about m. For each possible m we have the following.

$$m = 1 \Rightarrow 2^n = 2, \text{ so } n = 1,$$
$$m = 2 \Rightarrow 3^n = 3, \text{ so } n = 1,$$
$$m = 4 \Rightarrow 5^n = 25, \text{ so } n = 2.$$

Also solved by John Christopher, Michael P. Cohen, Charles Curtis & Jacob Boswell, Eagle Problem Solvers (Georgia Southern University), John Fitch, Khaled Halaoua (Syria), Walther Janous (Austria), Rafe Jones, Koopa Tak Lun Koo (Hong Kong), Seungheon Lee (South Korea), Graham Lord, Kelly D. McLenithan, Stephen Meskin, Raymond Mortini (France) & Rudolf Rupp (Germany) & Amol Sasane (UK), Sonebi Omar (Morocco), Didier Pinchon (France), Henry Ricardo, Celia Schacht, Albert Stadler (Switzerland), Wong Fook Sung (Singapore), and the proposer. There were two incomplete or incorrect solutions.

Does the series converge or diverge? **April 2021**

2118. *Proposed by Moubinool Omarjee, Lycée Henri IV, Paris, France.*

It is well known that the series

$$\sum_{k=1}^{\infty} \frac{\sin k}{k}$$

converges. Does the series

$$\sum_{k=1}^{\infty} e^{-\lfloor \ln k \rfloor} \sin k$$

converge or diverge?

Solution by the Northwestern University Math Problem Solving Group, Northwestern University, Evanston, IL.
Both series can be shown to be convergent using the following well-known result.

Dirichlet's test: If a_n is a monotonic sequence of real numbers that tends to zero, and b_n is a sequence of complex numbers such that, for some M, $\left|\sum_{n=1}^{N} b_n\right| \le M$ for every positive integer N, then the series $\sum_{n=1}^{\infty} a_n b_n$ converges.

For this problem, we use $a_k = e^{-\lfloor \ln k \rfloor}$, which tends monotonically to zero, and $b_k = \sin k$, whose partial sums are

$$\begin{aligned}
\sum_{k=1}^{N} \sin k &= \sum_{k=1}^{N} \frac{1}{2i}(e^{ik} - e^{-ik}) = \frac{1}{2i}\left(\frac{e^{i(N+1)} - e^{i}}{e^{i} - 1} - \frac{e^{-i(N+1)} - e^{-i}}{e^{-i} - 1}\right) \\
&= \frac{1}{2i}\left(\frac{e^{i(N+1/2)} - e^{i/2} + e^{-i(N+1/2)} - e^{-i/2}}{e^{i/2} - e^{-i/2}}\right) \\
&= \frac{1}{2}\left(\frac{\left(e^{i/2} + e^{-i/2}\right)/2 - \left(e^{i(N+1/2)} + e^{-i(N+1/2)}\right)/2}{\left(e^{i/2} - e^{-i/2}\right)/(2i)}\right) \\
&= \frac{1}{2\sin\frac{1}{2}}\left(\cos\tfrac{1}{2} - \cos\left(N + \tfrac{1}{2}\right)\right)
\end{aligned}$$

hence

$$\left|\sum_{k=1}^{N} \sin k\right| \le \frac{1}{\sin\frac{1}{2}}$$

So, by Dirichlet's test, the series converges.

Also solved by Hongwei Chen, Richard Daquila, Eagle Problem Solvers (Georgia Southern University), John Fitch, Russell Gordon, Eugene A. Herman, Walther Janous (Austria), Mark Kaplan & Michael Goldenberg, Raymond Mortini (France), Didier Pinchon (France), Omar Sonebi (Morocco), Albert Stadler (Switzerland), Seán Stewart (Australia), and the proposer. There were three incomplete or incorrect solutions.

Find the side length of the regular n-simplex — April 2021

2119. *Proposed by Viktors Berstis, Portland, OR.*

A point in the plane is a distance of a, b, and c units from the vertices of an equilateral triangle in the plane. Denote the side length of the equilateral triangle by s.

(a) Find a polynomial relation between a, b, c, and s.

(b) Give a simple compass and straightedge construction of a segment of length s given segments of lengths a, b, and c.

(c) Generalize part (a) to the case of a point at a distance of a_i units, $i = 1, \ldots, n+1$, from the vertices of a regular n-dimensional simplex having sides of length s.

Solution by Didier Pinchon, Toulouse, France.

(a) Given $s > 0$, the points

$$A = \left(0, \frac{\sqrt{3}}{3}s\right), \; B = \left(-\frac{1}{2}s, -\frac{\sqrt{3}}{6}s\right), \text{ and } C = \left(\frac{1}{2}s, -\frac{\sqrt{3}}{6}s\right)$$

are the vertices of an equilateral triangle with side length s. For a point $P = (x, y)$, the relations $PA^2 = a^2$, $PB^2 = b^2$ and $PC^2 = c^2$ give the equations

$$E_1 : x^2 + \left(y - \frac{\sqrt{3}}{3}s\right)^2 - a^2 = 0,$$

$$E_2 : \left(x + \frac{1}{2}s\right)^2 + \left(y + \frac{\sqrt{3}}{6}s\right)^2 - b^2 = 0,$$

$$E_3 : \left(x - \frac{1}{2}s\right)^2 + \left(y + \frac{\sqrt{3}}{6}s\right)^2 - c^2 = 0.$$

From $E_2 - E_3$, we get $x = (b^2 - c^2)/(2s)$, and substituting this value into $E_1 - E_2$, we get $y = \sqrt{3}(b^2 + c^2 - 2a^2)/(6s)$. Finally, substituting these values into equation E_1, we obtain

$$s^4 - (a^2 + b^2 + c^2)s^2 + a^4 + b^4 + c^4 - a^2b^2 - a^2c^2 - b^2c^2 = 0.$$

(b) Note that s^2 satisfies a second-degree polynomial with discriminant

$$\begin{aligned}\Delta &= (a^2 + b^2 + c^2)2 - 4(a^4 + b^4 + c^4 - a^2b^2 - a^2c^2 - b^2c^2) \\ &= 3(a + b + c)(a + b - c)(a + c - b)(b + c - a).\end{aligned}$$

Given positive real numbers a, b and c, $\Delta \geq 0$ if and only if $c \leq a + b$, $b \leq a + c$ and $a \leq b + c$. Indeed, when two factors of Δ are negative, say for example $a + c \leq b$, $b + c \leq a$, then $c \leq 0$, which is impossible. Hence, $\Delta \geq 0$ if and only a, b, and c are the lengths of the sides of a triangle. Note that the triangle is degenerate if and only if $\Delta = 0$. If a, b, and c are not all equal, then

$$a^4 + b^4 + c^4 - a^2b^2 - a^2c^2 - b^2c^2 = \frac{1}{2}\left[(a^2 - b^2)^2 + (b^2 - c^2)^2 + (c^2 - a^2)^2\right] > 0.$$

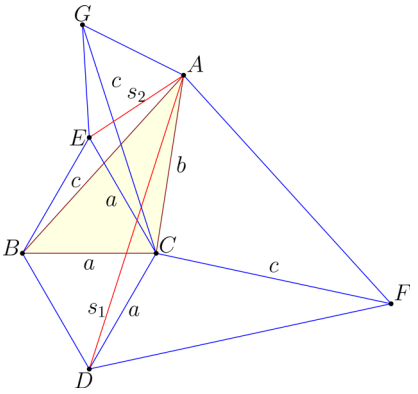
Therefore, the equation in s^2 has two different positive solutions, denoted by s_1 and s_2, if $\Delta > 0$ and a, b, and c are not all equal, and one positive solution otherwise.

The two solutions will now be constructed using a compass and straightedge. It is straightforward to construct a triangle ABC with side lengths a, b and c. The two circles of centers B and C and radius a intersect in two points D and E such that the triangles BCD and BCE are equilateral, with D and A being on opposite sides of line BC. Because ABC is not an equilateral triangle, point E is distinct from A.

We claim the lengths of the segments AD and AE are the solutions s_1 and s_2. The images of the points B and A by the rotation of center D and angle $-\pi/3$ are C and

F, and thus $BA = CF = c$. In a similar way, the images of A and B by the rotation of center E and angle $\pi/3$ are C and G, and thus $AB = c = CG$.

When $a = b = c$, then the first part of the construction is possible, and the unique solution is $s = DA = a\sqrt{3}$, and C is the center of equilateral triangle ADF. When $\Delta = 0$, A, B, and C are collinear, so A is equidistant from D and E and there is only one solution.



(c) *Editor's Note.* The solver uses the fact that if the distance between the ith and jth vertices of an n-simplex is $d_{i,j}$, then the volume of the simplex is

$$V^2 = \frac{(-1)^{n+1}}{2^n (n!)^2} \begin{vmatrix} 0 & 1 & 1 & 1 & \dots & 1 \\ 1 & 0 & d_{1,2}^2 & d_{1,3}^2 & \dots & d_{1,n+1}^2 \\ 1 & d_{1,2}^2 & 0 & d_{2,3}^2 & \dots & d_{2,n+1}^2 \\ \vdots & \vdots & & \ddots & & \vdots \\ 1 & d_{1,n}^2 & d_{2,n}^2 & & 0 & d_{n,n+1}^2 \\ 1 & d_{1,n+1}^2 & d_{2,n+1}^2 & \dots & d_{n,n+1}^2 & 0 \end{vmatrix}.$$

He applies this formula to the degenerate $(n+1)$-simplex whose vertices are the vertices of the regular n-simplex along with the additional point and performs a series of row and column operations to derive the result.

Here is an alternative derivation. Let the vertices of the regular n-simplex be

$$(s/\sqrt{2}, 0, 0, \dots, 0), (0, s/\sqrt{2}, 0, \dots, 0), \dots, (0, 0, \dots, s/\sqrt{2})$$

in $\mathbb{R}^{n+1}$. Note that these vertices lie in the hyperplane whose equation is

$$\sum_{i=1}^{n+1} x_i = \frac{s}{\sqrt{2}}.$$

Let $(x_1, \dots, x_{n+1})$ be a point in this hyperplane. We have

$$a_i^2 = -\sqrt{2} s x_i + \frac{s^2}{2} + \sum_{i=1}^{n+1} x_i^2. \tag{1}$$

Expanding and summing these equations as $i = 1, \dots, n+1$, we obtain

$$\sum_{i=1}^{n+1} a_i^2 = -\sqrt{2} s \sum_{i=1}^{n+1} x_i + (n+1)\frac{s^2}{2} + (n+1) \sum_{i=1}^{n+1} x_i^2$$

$$= -\sqrt{2}s\left(\frac{s}{\sqrt{2}}\right) + (n+1)\frac{s^2}{2} + (n+1)\sum_{i=1}^{n+1} x_i^2$$

$$= (n-1)\frac{s^2}{2} + (n+1)\sum_{i=1}^{n+1} x_i^2.$$

Therefore

$$\sum_{i=1}^{n+1} x_i^2 = -\frac{n-1}{2(n+1)}s^2 + \frac{1}{n+1}\sum_{i=1}^{n+1} a_i^2. \tag{2}$$

Substituting into (1), we have

$$a_i^2 = -\sqrt{2}sx_i + \frac{s^2}{2} - \frac{n-1}{2(n+1)}s^2 + \frac{1}{n+1}\sum_{i=1}^{n+1} a_i^2$$

$$= -\sqrt{2}sx_i + \frac{1}{n+1}s^2 + \frac{1}{n+1}\sum_{i=1}^{n+1} a_i^2.$$

Solving for x_i, we find that

$$x_i = \frac{1}{\sqrt{2}s}\left(-a_i^2 + \frac{1}{n+1}s^2 + \frac{1}{n+1}\sum_{i=1}^{n+1} a_i^2\right).$$

Substituting into (2) and letting $s = a_0$, we have

$$-\frac{n-1}{2(n+1)}a_0^2 + \frac{1}{n+1}\sum_{i=1}^{n+1} a_i^2 = \frac{1}{2a_0^2}\sum_{i=1}^{n+1}\left(-a_i^2 + \frac{1}{n+1}\sum_{i=0}^{n+1} a_i^2\right)^2$$

$$2a_0^2\left(-\frac{1}{2}a_0^2 + \frac{1}{n+1}\sum_{i=0}^{n+1} a_i^2\right) = \sum_{i=1}^{n+1}\left(-a_i^2 + \frac{1}{n+1}\sum_{i=0}^{n+1} a_i^2\right)^2.$$

Letting $T = \sum_{i=0}^{n+1} a_i^2/(n+1)$, we have

$$-a_0^4 + 2a_0^2T = \sum_{i=1}^{n+1} a_i^4 - 2T\sum_{i=1}^{n+1} a_i^2 + (n+1)T^2$$

$$0 = \sum_{i=0}^{n+1} a_i^4 - 2T\sum_{i=0}^{n+1} a_i^2 + (n+1)T^2$$

$$0 = \sum_{i=0}^{n+1} a_i^4 - 2T(n+1)T + (n+1)T^2$$

$$\sum_{i=0}^{n+1} a_i^4 = (n+1)T^2 = \frac{1}{n+1}\left(\sum_{i=0}^{n+1} a_i^2\right)^2$$

$$(n+1)\sum_{i=0}^{n+1} a_i^4 = \sum_{i=0}^{n+1} a_i^4 + 2\sum_{0\le i<j\le n+1} a_i^2 a_j^2$$

$$n\sum_{i=0}^{n+1} a_i^4 = 2\sum_{0\le i<j\le n+1} a_i^2 a_j^2,$$

which is the desired relation.

Also solved by Elton Bojaxhiu (Germany) & Enkel Hysnelaj (Australia), Albert Stadler (Switzerland), and the proposer. There were two incomplete or incorrect solutions.

Find the normalizer **April 2021**

2120. *Proposed by Gregory Dresden, Jackson Gazin (student), and Kathleen McNeill (student), Washington & Lee University, Lexington, VA.*

Recall that the normalizer of a subgroup H of G is defined as

$$N_G(H) = \left\{ g \in G | ghg^{-1} \in H \text{ for all } h \in H \right\}.$$

Determine $N_G(H)$, when $G = GL_2(\mathbb{R})$, the group of all invertible 2×2 matrices with real entries, and

$$H = SO_2(\mathbb{R}) = \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \middle| \, \theta \in \mathbb{R} \right\}.$$

Solution by Eugene A. Herman, Grinnell College, Grinnell, IA.
More generally, for any $n \ge 1$, let $G = GL_n(\mathbb{R})$ and $H = SO_n(\mathbb{R})$, the subgroup of $O_n(\mathbb{R})$, the group of orthogonal matrices, consisting of matrices whose determinant is 1. We will show that

$$N_G(H) = \{aU \,|a \in \mathbb{R} - \{0\}, U \in O_n(\mathbb{R})\}.$$

Suppose $A = aU$, where $a \neq 0$ and U is orthogonal. Then for any $M \in SO_n(\mathbb{R})$,

$$AMA^{-1} = aUM\,\frac{1}{a}\,U^{-1} = UMU^{-1}.$$

Since

$$\det(UMU^{-1}) = \det(U)\det(M)/\det(U) = 1,$$

and the product of orthogonal matrices is orthogonal, we see that $AMA^{-1} \in SO_n(\mathbb{R})$.

For the converse, we use a polar decomposition. For $A \in N_G(H)$, write $A = PU$, where P is positive-definite and U is orthogonal. For any $M \in SO_n(\mathbb{R})$, let $N = U^{-1}MU$. Then $N \in SO_n(\mathbb{R})$, so $ANA^{-1} \in SO_n(\mathbb{R})$. But

$$ANA^{-1} = P(UNU^{-1})P^{-1} = PMP^{-1},$$

so $P \in N_G(H)$. Therefore, it remains only to determine which positive-definite matrices are in the normalizer. Now every positive-definite matrix can be written as $P = VDV^{-1}$, where $D = \text{diag}(d_1, \ldots, d_n)$ is a diagonal matrix with $d_i > 0$ and $V \in O_n(\mathbb{R})$. For any $M \in SO_n(\mathbb{R})$, let $N = VMV^{-1}$. Then $B = PNP^{-1} \in SO_n(\mathbb{R})$ and

$$B = VDMD^{-1}V^{-1} \in SO_n(\mathbb{R}), \text{ so } DMD^{-1} = V^{-1}BV \in SO_n(\mathbb{R}).$$

Therefore, $D \in N_G(H)$.

For $k > 1$, let $M_k = [m_{ij}]$, where

$$m_{11} = 0, m_{1k} = -1, m_{k1} = 1, m_{kk} = 0, m_{ii} = 1 \ (i \neq 1, k), \quad \text{and} \quad m_{i,j} = 0 \ \text{otherwise.}$$

Then $R \in SO_n(\mathbb{R})$ and the first column of DRD^{-1} consists of zeros except the kth entry, which is d_k/d_1. Since DRD^{-1} is orthogonal, this column must have length 1, which means that $d_k = d_1$ for all $k > 1$. Therefore D is a positive multiple of the identity, and so A is a multiple of an orthogonal matrix.

Note: The same proof works for the complex version. In that case, $G = GL_n(\mathbb{C})$ and $H = SU_n(\mathbb{C})$, where the latter is the group of $n \times n$ unitary matrices whose determinant equals 1. Then $N_G(H)$ is the group of all nonzero complex multiples of $n \times n$ unitary matrices.

Also solved by Elton Bojaxhiu (Germany) & Enkel Hysnelaj (Australia), Robert Calcaterra, Eagle Problem Solvers (Georgia Southern University), John Fitch, Dmitry Fleischman, Mark Kaplan & Michael Goldenberg, Koopa Tak Lun Koo (Hong Kong), Didier Pinchon (France), Albert Stadler (Switzerland) and the proposers. There were two incomplete or incorrect solutions.

Answers

Solutions to the Quickies from page 158.

A1119. The aces divide the 48 other cards into 5 "urns", with a, b, c, d, and e non-aces in them, respectively. The position of the third ace is equal to $a + b + c + 3$, so the expected value of its position is $E[a + b + c + 3]$. By linearity of expectation, this is $E[a] + E[b] + E[c] + 3$. Because a non-ace is equally likely to be placed in any of the five "urns", $E[a] = \ldots = E[e]$. Since $E[a + b + c + d + e] = 48$, we have $E[a] = \ldots = E[e] = \frac{48}{5}$.

Therefore the expected value is

$$3 \cdot \frac{48}{5} + 3 = \frac{159}{5}.$$

A1120. Let S, S_1, S_2, and S_3 be the areas of $\triangle ABC$, $\triangle XBC$, $\triangle XCA$, and $\triangle XAB$, respectively. Let h_2 and h_3 be the heights of $\triangle XCA$ and $\triangle XAB$ with $\overline{AX}$ as base. Let θ be the angle between $\overleftrightarrow{AX}$ and $\overleftrightarrow{BC}$. Then

$$S_2 + S_3 = \frac{1}{2}(h_2 + h_3) R_1 = \frac{1}{2} a \sin\theta R_1 \leq \frac{1}{2} a R_1.$$

Similar arguments give

$$S_3 + S_1 \leq \frac{1}{2} b R_2 \ \text{ and } \ S_1 + S_2 \leq \frac{1}{2} c R_3.$$

Therefore

$$\frac{1}{2} a R_1 + \frac{1}{2} b R_2 + \frac{1}{2} c R_3 \geq (S_2 + S_3) + (S_3 + S_1) + (S_1 + S_2) = 2S = r(a + b + c)$$

and the result follows.

Equality holds if and only if the line through a vertex and X and the line containing the side opposite the vertex are perpendicular. In other words, X must be the orthocenter of the triangle, which must be acute in order for X to lie in its interior.

Note. Let O and R be the circumcenter and the circumradius for a given acute triangle. Since $R_1 = R_2 = R_3 = R$, we obtain Euler's inequality $R \geq 2r$.

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.

Singmaster, David, *Adventures in Recreational Mathematics*, 2 vols., World Scientific, 2022; xiv + 300 pp, $78, $38(P), ISBN 978-981-122-600-7, -650-2; xiv + 164 pp, $68, $28(P), ISBN -603-8, -651-9. Both together: $108, $48(P), $38(ebook), ISBN -564-2, -630-4, 978-981-12-5162-7.

What makes for good recreational mathematics? According to David Singmaster, it must be fun and understandable to an interested person. For many years—particularly since 1978, when he became fascinated with Rubik's cube—he has been tracing and documenting sources for the history of recreational mathematics, leading to a bibliography of more than a thousand pages. The two volumes noted here are not that work but are fruits of that labor, drawing together in one place (with updates) works by Singmaster about particular puzzles and problems. The first volume surveys ancient and medieval puzzles, then considers old puzzles as sources of mathematical questions not envisioned by their creators. The volume ends with open problems in the history of recreational mathematics (e.g., does the liar paradox occur in Arabic or Indian works?). The second volume begins with an apologia for recreational mathematics (why is it so useful?), followed by essays concentrating on recently-posed problems (some by Singmaster himself), including a chapter of back-of-the-envelope calculations. Not everything is (or could possibly be) here—for example, the Archimedes cattle problem is not mentioned (though I can't be sure—most regrettably, neither volume has an index). One wishes that the many wonderful illustrations could have been reproduced in color (in the ebook, many are).

Bewersdorff, Jörg, *Luck, Logic, and White Lies*, 2nd ed., CRC Press, 2021; xx + 548 pp, $133, $54.95(P), $49.45(ebook). ISBN 978-0-367-55296-1, -54841-4, 978-1-00309287-2.

This book, successor to the first edition (2005) and translated from the 7th German edition, treats games of chance ("luck"), combinatorial games ("logic"), and games of strategy (bluff, or "white lies"). The first part develops succinctly the needed theory of probability and investigates the nature of randomness. The second part explores minimax optimization, Grundy values, Conway's theory of games, and complexity theory. The third part is based on the fact that in a symmetric two-person zero-sum game, the players are guaranteed optimal mixed strategies; for some games, finding such strategies can be done by linear programming. This edition adds a fourth part that investigates measuring the proportion of skill in a game, with particular application to poker. The reader needs to be comfortable with algebra and summation signs, and infinite series make appearances; end-of-chapter notes and footnotes contribute further mathematical depth.

Schneider, Jonathan, Toward a combinatorial theory of SET and related card games, https://arxiv.org/pdf/2010.01882.pdf.

The game of SET has proven to be a rich source for mathematical investigation. Author Schneider visualizes the SET deck as a hypercube and defines a notion of isomorphism of hands. The isomorphism leads to classifying and counting the types of n-card hands for small values of n, and the author generalizes the problem to that of enumerating k-ary codes of length d with n words. He introduces a new SET game variation STUN, and he notes that the game EvenQuads from the Association for Women in Mathematics is similar to SET and hence susceptible to analogous analysis.

Math. Mag. **95** (2022) 167–168. doi:10.1080/0025570X.2022.2026094 © Mathematical Association of America

Capaldi, Mindy (ed.), *Teaching Mathematics Through Games*, MAA Press, 2021; xii + 159 pp, $60(P) ($54 MAA member, $48 AMS member). ISBN 978-1-4704-6284-0.

Games share with mathematics people's desires to detect and investigate patterns and conjectures, and learning mathematics through games can bring joy to the classroom. This book's 17 chapters suggest games appropriate for courses from general education to abstract algebra, with most designed for a single class period; the largest number involve combinatorics and probability. Games include take-offs and variations on familiar ones (tic-tac-toe, Farkle, keno, roulette, the Arkham Horror card game, Risk, Apples to Apples, peg solitaire, SET, Settlers of Cataan, and Carcassonne), as well as original games (numeration systems sudoku, Function Battleship, and Bandits on a Wall). Supplementary materials for some chapters are available at the AMS site for the book.

Eastaway, Rob, and John Haigh, *The Hidden Mathematics of Sport*, Portico, 2021; viii + 200 pp, $14.29. ISBN 978-1-911622-28-4.

This is a "dip-in" book chock full of delightful mathematical tidbits about various sports: tactics for a soccer penalty kick, what constitutes a forward pass, sphericity of sports balls, breaking records, where to aim in darts, points systems, judging in skating, home team advantage (in the English Premier League, 63% of penalties are awarded to the home team!), and more. The authors are British, so the lean is toward sports popular in Britain. Underemphasized is that the book is a reissue with unspecified updating of *How to Take a Penalty* (Robson, 2005).

Karaali, Gizem, and Lily S. Khadjavi, *Mathematics for Social Justice: Focusing on Quantitative Reasoning and Statistics*, MAA Press, 2021; vii + 287 pp, $60(P) ($45 MAA or AMS member). ISBN 978-1-4704-4927-8.

G.H. Hardy noted that a science (including presumably mathematics) is "said to be useful if [it] tends to accentuate the existing inequalities in the distribution of wealth." The editors of this book set out a different criterion for utility, as they did in their previous *Mathematics for Social Justice: Resources for the College Classroom* (MAA, 2019). This book focuses specifically on quantitative reasoning and elementary statistics courses, which at many institutions are taught by mathematics faculty. Different authors have contributed 17 modules with in-class activities, research projects, problem sets, and other materials for engaging students with statistics and numerical reasoning through consideration of social ramifications, such as gerrymandering, gentrification, drug testing, racial profiling, income inequality, and human rights. As you might expect from the book's focus, much of the material could be used at various educational levels.

Carlson, Joshua, et al., Parking functions: Choose your own adventure, *College Mathematics Journal* 52 (4) (2021) 254–264.

The authors consider a wide variety of protocols for the parking of cars along a one-way street. Each car has a preferred spot on the street and will park in it if it is available; if it is already occupied, various strategies are considered. Cars may move only forward, or alternatively may back up only a fixed number of spots. Further, cars may have different lengths, or they may park at random (but get lucky in getting their preferred spots), or they may park instead along a graph. And we may want to assess the overall success of the parking in some way. The authors enumerate many such problems and indicate which have been solved (with references) and which are still open. This article is a fun example of how to generate new mathematics by abstracting from real situations.

Davies, Alex, et al., Advancing mathematics by guiding human intuition with AI, *Nature* 600 (1 December 2021) 70–74. https://www.nature.com/articles/s41586-021-04086-x.pdf.

A too-brief summation of mathematical activity is that it consists of investigating examples, observing patterns, imagining conjectures, and constructing proofs (or counterexamples). Mathematicians have long used aids in the first step, including calculators and computers. The authors provide examples of new mathematical results in knot theory invariants and in the structure of irreducible representations in Fourier analysis, "discovered with the assistance of machine learning," including using the latter to "guide intuition and propose conjectures." Nevetheless, the question remains: Does artificial intelligence (AI) offer a new paradigm, or is it just expanded computing?