QUANTUM

NOVEMBER/DECEMBER 1998

US $7.50/CAN $10.50

Springer

GALLERY Q

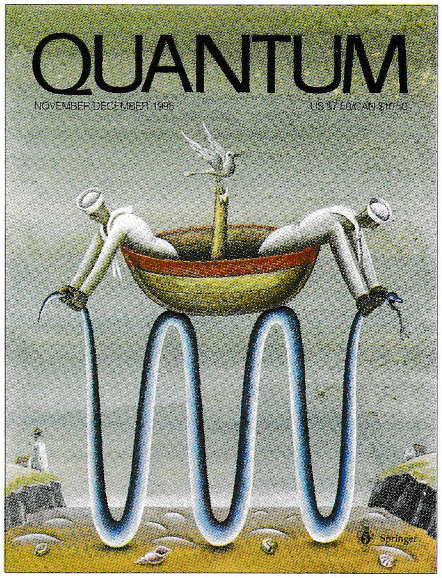
Oil on panel, 12 × 13 inches, Collection of the Salvador Dali Museum, St. Petersburg, Florida

Morphological Echo (1936) by Salvador Dali

UNDERSTANDING THE RELATIONSHIP BETWEEN members of a group is often difficult. The objects in Dali's matrix share certain characteristics in their horizontal and vertical groupings. Animal, mineral, or vegetable; standing, sitting, or prone—how else are these items interrelated? Do they all share a common characteristic? Perhaps Dali's color palette ties them together. Or, the common light source may be the tie that binds. Searching for connections between members of a group of elements can be a very stimulating exercise. When dealing with a group of functions, composing your thoughts can be very valuable. See how group dynamics are exposed when functions are tabled by turning to page 14.

QUANTUM

NOVEMBER/DECEMBER 1998 VOLUME 9, NUMBER 2



Cover art by Jose Garcia

The two sailors on this month's cover are quite fed up with the unruly seas that cost them their mast and are taking matters into their own hands. If you've ever experienced a similar frustration when dealing with wave equations, turn to page 20 for the primer "Sea Waves: Troughs and crests from head to toe." It should provide you with a seaworthy compass for navigating the ups and downs of wave physics.

Indexed in *Magazine Article Summaries, Academic Abstracts, Academic Search, Vocational Search, MasterFILE*, and *General Science Source*. Available in microform, electronic, or paper format from University Microfilms International.

FEATURES

DEPARTMENTS

Faraday's legacy: The joys of methodology

Around the magnet Faraday
Was sure that Volta's lightnings play:
But how to draw them from the wire?
He drew a lesson from the heart:
'Tis when we meet, 'tis when we part,
Breaks forth the electric fire.
—by Herbert Mayo[1]

MICHAEL FARADAY'S PRESence once graced the halls of England's Royal Institution and the world at large. His approach to nature displayed simplicity and curiosity, and was driven by a passion to understand. A discoverer of great laws, such as magnetoelectric induction, Faraday succeeded (despite his scant use of mathematics) because of his keen analytical mind coupled with a remarkable imagination for creating picturesque models to formulate concept-rich inquiries.

His public lectures were almost entirely *qualitative*. They conveyed how the search for causal connections is used both as a matter of course essential to the scientific method and as a source of delight for his audience. As one means of propagating the scientific method beyond the lecture theater, Faraday would, on occasion, give out samples of materials like the ones used in his demonstrations or tell his audience how such materials could be easily obtained and invite those in attendance to try some of the experiments on their own.

Faraday's popular lectures were aimed at children and at the "child" in the attending adults. His series of enthusiastic talks "The Forces of Matter" and "The Chemical History of a Candle" are models of scientific clarity bolstered by a range of captivating demonstrations. He would demonstrate that combustion releases water through the combination of hydrogen from the candle's paraffin with oxygen from the air, and dramatize this by exploding a mixture of hydrogen and oxygen to obtain water. He would show (the counterintuitive) spontaneous burning of finely powdered lead broken out of its vacuum container to unite with the air's oxygen. His dramatic demonstrations were offered not as a succession of incomprehensible miracles, but as theater through which intelligible causal laws play themselves out to delight the understanding of inquiring minds.

Following Faraday, we can subject even our simplest experiences to scientific inquiry: Why does someone hold a pan of water *above* the cup when pouring? Although one is tempted to answer "experience," that is not a sufficient explanation of the event. For the scientist it is not an explanation to claim that because water poured has fallen into the cup in the past then it will fall into the cup in the future. It is, of course, "gravity" that causes the water to fall. But why believe in gravity? Is it because every time you see an object fall you simply tag the experience "gravity"? No, not according to our knowledge of physics. It is because we have reasoned our way through to the belief that there are underlying principles in nature, referred to as Newton's law of gravitation and his second law of motion. It is these that enable us to understand *why* the water falls. Moreover, from a little knowledge of the water's present state, we can also use those principles to predict *how* it will fall. And we can extend those principles out into space to understand the motion of our communication satellites, of planets, of binary stars, and of Halley's comet (predicting when it should be returning to our Solar System and where it will be when it returns).

Why does a candle flame continue to burn? What is combustion? Directing a beam of light through the flame, one sees a darker shadow from the inner region of the flame, and a lighter shadow from its outer region. Faraday demonstrates *how* to conclude that the inner region contains candle wax vapor producing particles of carbon that are heated to incandescence in the outer region. What is the substance? How do you know? Those are issues Faraday wants us to appreciate.

"What is the cause?" This question has been integral to scientific activity. Even today we have seen how such inquiry has led to our belief in the existence of molecules and atoms—from Einstein's expla-

[1]As communicated through Sir Charles Wheatstone to J. H. Gladstone.

nation of Brownian motion to its verifications by Perrin, to the observation of atomic dislocations in crystals by that quantum device called the scanning tunneling electron microscope.

Faraday conveyed his delight in science to members of the audience because of his joyfully active curiosity, his obvious love of scientific investigation, and his flair for drama. And he did this through his implicit belief that *the world is intelligible*. He knew that, in large measure, the way to understanding lies in the joy of the pursuit. Not the discovery but the inquiry. Not the goal but the journey.

The scientific mind conquers nature, not people. And it does so through understanding—through the testing of hypotheses, however passionately formed, and through the lucidity of logic, however coolly applied. This may seem to be a rather romantic view of science. It is. It is also, I believe, a realistic view of science. Science without passion is sterile. And passion without logic is not science.

As we head toward the last year of the twentieth century we are witness to the fruits of Faraday's legacy—a century of scientific achievement without precedent in its number of discoveries and based, in large part, on pillars of this century's science—the theories of relativity, quantum theory, and molecular biology. The experimental results and technological applications of those theories have been amply demonstrated—in our discoveries of the nature of matter (such as the quark structure of the proton), in our discoveries of quasars (exotic astronomical objects near the edge of the observable Universe), in our use of laser "tweezers" to manipulate microscopic life forms, and in our revelation of some of the major secrets of life itself through the discovery of DNA. Now, how can we convey Faraday's legacy to our students?

Laurence I. Gould *is Professor of Physics, and has also taught interdisciplinary courses for non-science majors, at the University of Hartford, West Hartford, Connecticut.*

QUANTUM

THE MAGAZINE OF MATH AND SCIENCE

A publication of the National Science Teachers Association (NSTA)
& Quantum Bureau of the Russian Academy of Sciences
in conjunction with
the American Association of Physics Teachers (AAPT)
& the National Council of Teachers of Mathematics (NCTM)

The mission of the National Science Teachers Association is to promote excellence and innovation in science teaching and learning for all.

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 9 (6 issues) will be published in 1998–1999. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by Quantum Bureau (Moscow, Russia), as well as original material in English. **Editorial offices:** NSTA, 1840 Wilson Boulevard, Arlington VA 22201-3000; telephone: (703) 243-7100; electronic mail: quantum@nsta.org. **Production offices:** Springer-Verlag New York, Inc., 175 Fifth Avenue, New York NY 10010-7858.

Periodicals postage paid at New York, NY, and additional mailing offices. **Postmaster:** send address changes to: *Quantum*, Springer-Verlag New York, Inc., Journal Fulfillment Services Department, P. O. Box 2485, Secaucus NJ 07096-2485.

Printed in U.S.A.

Subscription Information:

North America: Student rate: $18; Personal rate (nonstudent): $25. This rate is available to individual subscribers for personal use only from Springer-Verlag New York, Inc., when paid by personal check or charge. Subscriptions are entered with prepayment only. Institutional rate: $47. Single Issue Price: $7.50. Rates include postage and handling. (Canadian customers please add 7% GST to subscription price. Springer-Verlag GST registration number is 123394918.) Subscriptions begin with next published issue (backstarts may be requested). Bulk rates for students are available. Mail order and payment to: Springer-Verlag New York, Inc., Journal Fulfillment Services Department, PO Box 2485, Secaucus, NJ 07094-2485, USA. Telephone: 1(800) SPRINGER; fax: (201) 348-4505; e-mail: custserv@springer-ny.com.

Outside North America: Personal rate: Please contact Springer-Verlag Berlin at subscriptions@springer.de. Institutional rate is US$57; airmail delivery is US$18 additional (all rates calculated in DM at the exchange rate current at the time of purchase). SAL (Surface Airmail Listed) is mandatory for Japan, India, Australia, and New Zealand. Customers should ask for the appropriate price list. Orders may be placed through your bookseller or directly through Springer-Verlag, Postfach 31 13 40, D-10643 Berlin, Germany.

Advertising:

Advertising Representatives: (Washington) Paul Kuntzler (703) 243-7100; (New York) Jay Feinman (212) 460-1682; and G. Probst, Springer-Verlag GmbH & Co. KG, D-14191 Berlin, Germany, telephone 49 (0) 30-827 87-0, telex 185 411.

Printed on acid-free paper.

Lattices and Brillouin zones

Can you master these domains?

by A. B. Goncharov

THE GEOMETRY OF LATTICES can be quite picturesque. Indeed, the most enjoyable part of this article will be the pictures. If you understand them well enough, then all the basic ideas and constructions will be quite clear to you.

Before proceeding we should note that the problems discussed here are not mere abstractions. On the contrary, they arise directly from the physics of crystals. Toward the end of the article we will treat the physical point of view in greater detail.

Let's start by marking all the points with integer coordinates on the plane. They will be the nodes of a square lattice. Let's choose an origin O from among them. Now, for any other node P of the lattice, we can draw the line l such that the nodes O and P are symmetric with respect to this line. In other words, line l will be the perpendicular bisector of segment OP. These lines will divide the plane into a set of small cells (triangles and convex polygons). Let's assign an integer, called its *rank*, to each cell, according to the following rule: The cell containing point O (it is a square) will be given a rank of 1; cells adjacent to this one will get a rank of 2; cells adjoining to them and different from those, we've already considered, will get a rank of 3, and so on.

Let's paint the cells near the origin in different colors, so that the colors of cells with equal ranks coincide (fig. 1). It turns out that the areas of the domains painted in different colors are equal.

We can take a lattice consisting of regular triangles or a lattice made of regular hexagons and perform the same operations with them (figs. 2 and 3). We will find that the same

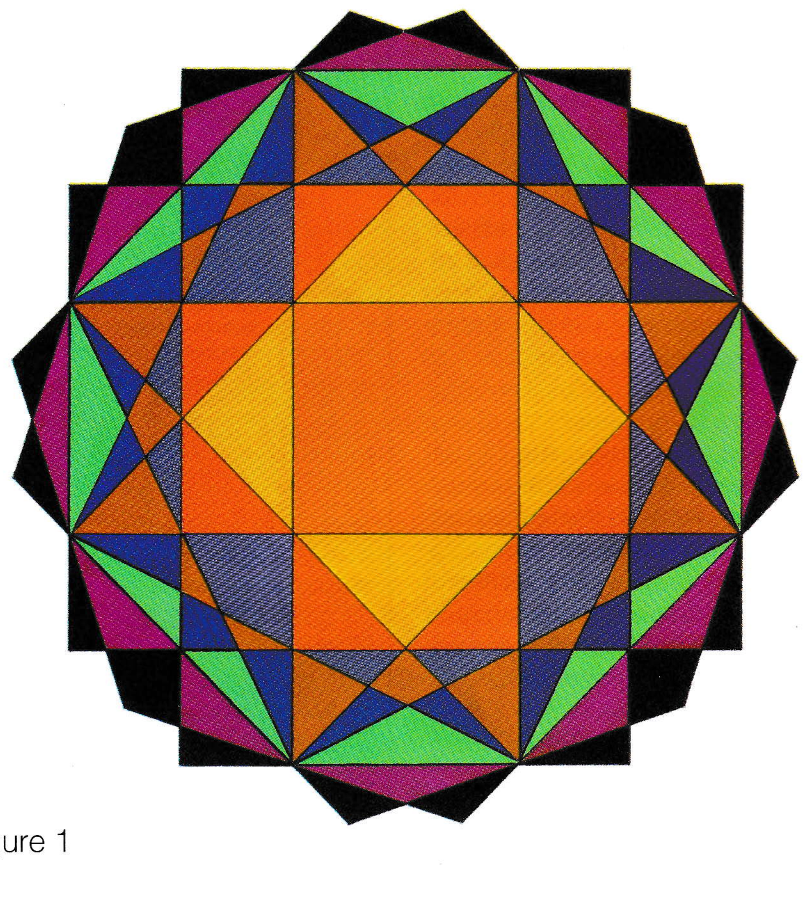


Figure 1

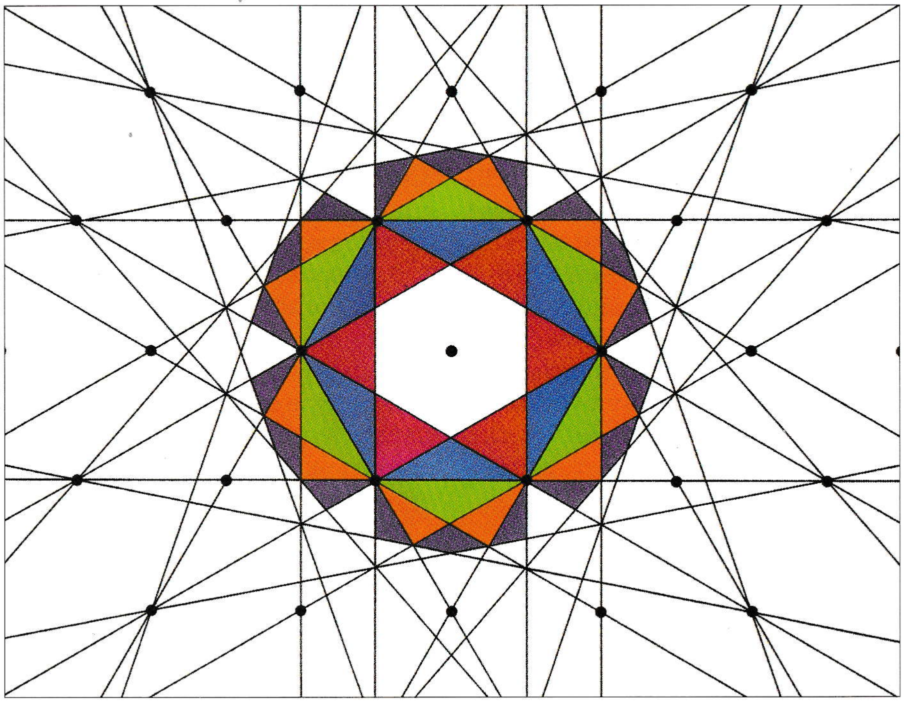


Figure 2

regularity holds: zones painted in different colors have equal areas.

Let's denote the union of all cells of rank r (r = 1, 2, 3, ...) by $D_r(O)$, where O is the "central" node of the lattice. We will show that for all possible lattices the area of $D_r(O)$ does not depend on r.

Having scrutinized figures 2 and 3 for some time, we can make two important observations that suggest two different ways to prove this statement. So, let's start with these observations.

Observation 1. Consider the simplest lattice, with square cells. The blue domain in fig. 4 is $D_6(O)$. The thick lines in the figure break the plane into equal squares, so that any node Q of our lattice is in the center of one of these squares. Call this square $D(Q)$. Then each $D(Q)$ is the image of the central square $D(O) = D_1(O)$ under a parallel translation by the vector **OQ**. If we cut the plane along the thick lines and move all the squares so they coincide with the central square ($D(O)$), then the blue pieces of $D_6(O)$ will fill the whole square, without overlapping. (The same thing happens for all r = 2, 3, ... We suggest that you check it for as many different r as necessary to convince you that this statement is true). This means, of course, that the area of $D_6(O)$—and of $D_r(O)$ for all r—is equal to that of $D_1(O)$.

Observation 2. We will illustrate this observation, partly for the sake of variety, using a lattice consisting of the vertices of a regular hexagonal tiling of the plane. Figure 2 represents the domains $D_r(O)$ for r = 1, 2, ... ,6.

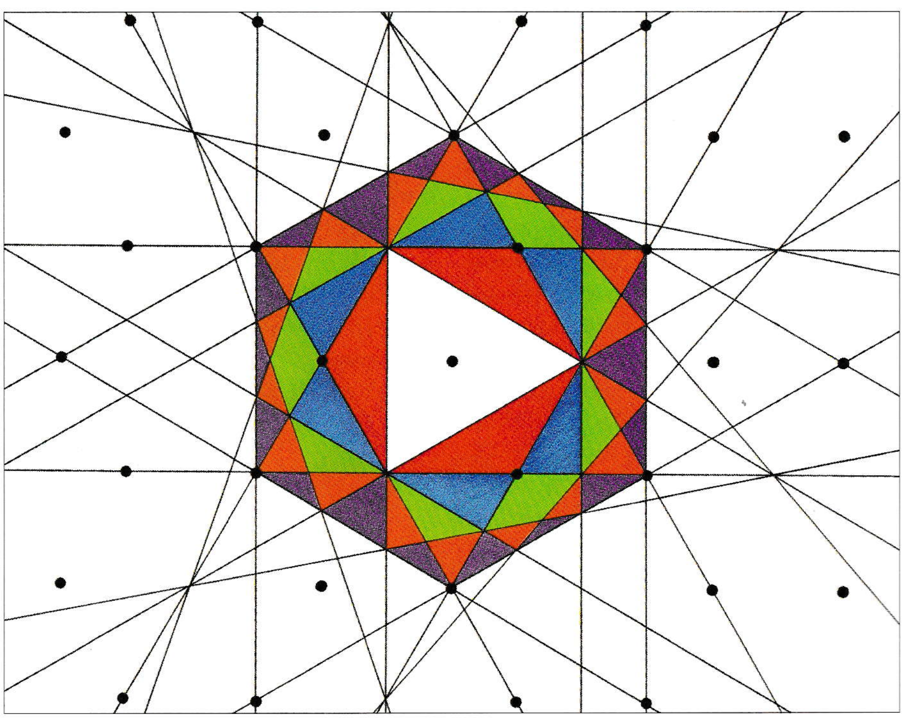


Figure 3

The construction of observation 1 starts with a choice of one lattice point for the origin. But in fact we can choose any lattice point for the origin and perform the same construction. For example, if we choose an arbitrary lattice point Q as the origin, we can construct D_4Q, which is a figure congruent to the six blue triangles of figure 2, but centered at the point Q. Figure 5 shows this construction for several choices of Q. Each Q has a different color, and the corresponding D_4Q is shown in the same color. We now see that the set of domains $D_4(Q)$ for different Q covers the plane without overlapping. (In fact, the same is true for all $D_r(Q)$, r = 1, 2, ...) Thus, the area of $D_4(Q)$ (and, more generally, of $D_r(Q)$ for all r) is equal to the "mean" area that falls on one node. This will be explained further in what follows. Now, let's formulate a lemma that explains both observations.

The main lemma. The domain $D_r(O)$ consists of all the points A of the plane such that the distance from A to the node O is in the r^{th} place in the sequence of distances from A to the nodes of the lattice.

Let's start with the case r = 1. Let P be any lattice point. Note

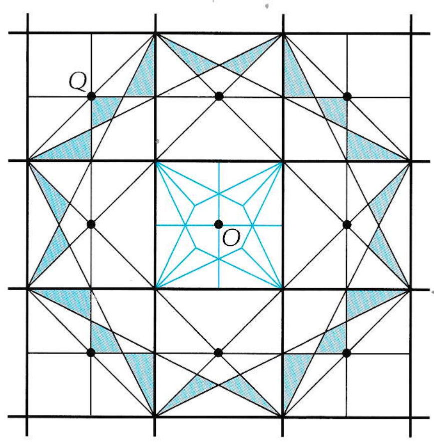


Figure 4

that the perpendicular bisector of segment PO divides the plane into two half-planes such that the points A lying in the half-plane that contains O are nearer to O than to P, and vice-versa. The domain $D(O) = D_1(O)$ is the intersection of all such half-planes for all the nodes P different from O. Therefore $D(O)$ consists of the points A for which O is the nearest node of the lattice. ($D(O)$ is called the Dirichlet domain of the node O.)

Now let B be a point of rank $r > 1$ (that is, let it belong to $D_r(O)$). According to the definition of the rank given above, there must exist a path from B to O that intersects exactly $r - 1$ perpendicular bisectors drawn to a set of segments $OP_1, OP_2, \ldots, OP_{r-1}$. We can construct this path as follows: At first we travel along a segment of a line from point B to any point on the border of domains $D_r(O)$ and $D_{r-1}(O)$. Then, along another line, we come to a point on the border of $D_{r-1}(O)$ and $D_{r-2}(O)$, and so on. And when we come to the domain $D_1(O) = D(O)$, we go straight to point O. This means that there are $r - 1$ nodes $P_1, P_2, \ldots, P_{r-1}$ of the lattice that lie nearer to point B than to O (because we have crossed $r - 1$ perpendicular bisectors in moving from B to O). On the other hand, if there were more than $r - 1$ nodes with this property, we would inevitably have to cross more than $r - 1$ perpendicular bisectors, and thus the lemma is proved.

Now we use this lemma to analyze our observations.

1. Suppose that the lattice maps into itself when translated by the vector **OQ**, where O and Q are two arbitrary nodes. (This condition holds for the square lattice, the "triangular" lattice in figure 1, and generally speaking, any lattice whose nodes are the endpoints of the vectors $m \cdot \mathbf{OA} + n \cdot \mathbf{OB}$, where OAB is a fixed triangle and m and n are arbitrary integers.) Below we call "translations" only the translations by the vectors **OQ**. Let's prove that for every internal point M in the Dirichlet domain $D(O)$ there exists a translation T such that $T(M) \in D_r(O)$. Moreover, we will show that if $T(M)$ lies inside domain $D_r(O)$, then there exists only one such translation. In other words, the domain $D(O)$ breaks up into a set of pieces, which, when translated, assemble into domain $D_r(O)$.

To prove this, let C be a point inside $D(O)$ such that Q is the r^{th} node of the lattice if we order them in accordance with the distances from them to C. According to our main lemma, $C \in D(O) \cap D_r(Q)$, and thus, translating it by the vector **OQ**, we see that $T(C) \in D_r(O)$. The translation T is determined in a unique way unless there exists a node X in the lattice such that $|CQ| = |CX|$. It's clear from this that the translation is unique for all the points in $D(O)$ except for those lying on a finite set of segments. These segments are just the lines along which we must cut the domain $D(O)$ to put the pieces into $D_r(O)$.

2. Suppose that for any two nodes P and Q of the lattice we can find a way to move the lattice onto itself in such a way that P maps onto Q. In the previous case, we spoke about translations that mapped the lattice onto itself. Here, we allow any sort of transformation.

According to our main lemma, for each point C in the plane there is only one node Q such that $C \in D_r(Q)$. It is the r^{th} node in the list where the nodes are ordered according to their distances to C. Now it's clear that the domains $D_r(Q)$ with different Q intersect only at their edges. Our initial assumption about the lattice guarantees that all the domains $D_r(Q)$ are congruent: The motion of the lattice that sends Q into O maps $D_r(Q)$ in $D_r(O)$.

Now we can explain why the areas of $D_r(O)$ are equal for all r. For any lattice we will call its "density" the following limit:

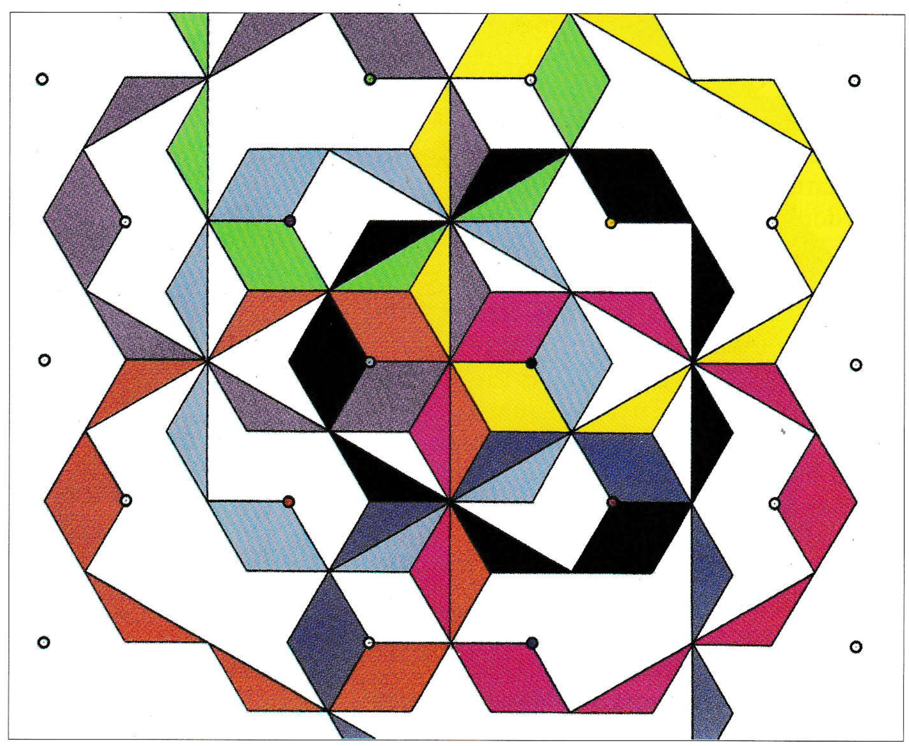


Figure 5

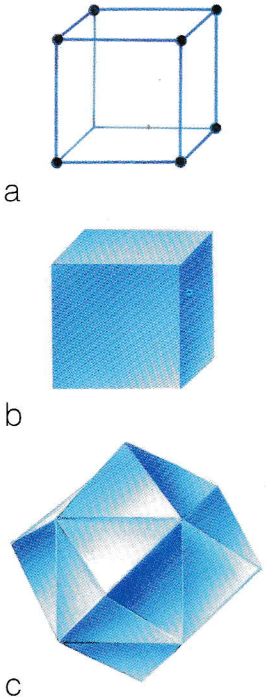


Figure 6

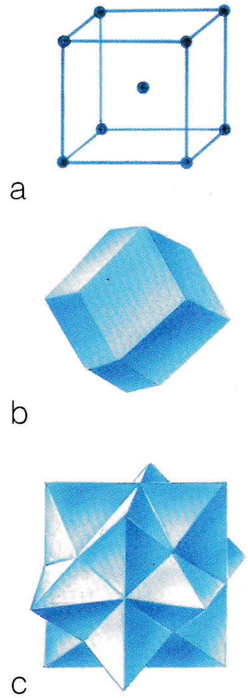


Figure 7

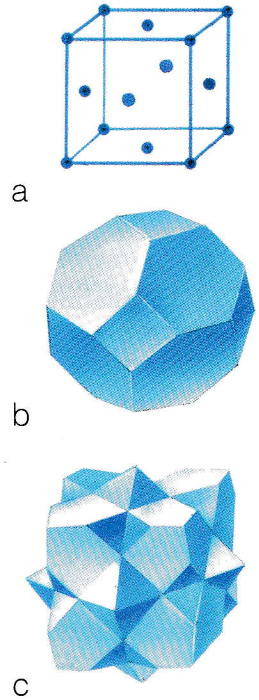


Figure 8

$$\alpha = \lim \frac{K(N)}{N^2},$$

where $K(N)$ is the number of nodes in the $N \times N$ square whose center is O. We'll prove that for each r, $\alpha = 1/S_r$ (here S_r is the area of $D_r(O)$). Let L_r be the greatest distance from point O to a point in domain $D_r(O)$, and let $N > L_r$. Then the union of the domains $D_r(Q)$ taken for all $K(N)$ nodes Q that lie inside the $N \times N$ square covers the $(N-L_r)\times(N-L_r)$ square and ends in the $(N+L_r)\times(N+L_r)$ square. Therefore,

$$\left(1-\frac{L_r}{N}\right)^2 = \frac{(N-L_r)^2}{N^2} \le \frac{S_r \cdot K(N)}{N^2} \le \frac{(N+L_r)^2}{N^2} = \left(1+\frac{L_r}{N}\right)^2.$$

When $N \to \infty$, the right and left parts of this inequality tend to 1. Thus, we have

$$\lim \frac{S_r \cdot K(N)}{N^2} = 1,$$

and therefore, $\alpha = 1/S_r$. So, $1/\alpha$ is the "mean area per node," in particular S_1, the area of domain $D(O)$, is also equal to $1/\alpha$.

Note. These same ideas appear in a famous lemma of Minkowski about a convex body, which has numerous applications in number theory.

Minkowski's lemma. Let C be a convex centrally-symmetric figure such that the node O of the lattice coincides with its center. Then, if the ratio of the areas of C and $D_1(O)$ is greater than 4, then there must exist a node inside C different from O.

Hint for the proof. First, we show that there exist two different nodes P and Q of the lattice such that the images of the domain ½ C under the translations at vectors **OP** and **OQ** overlap. (Here ½ C denotes the figure consisting of the ends of the vectors ½ **OM**, where $M \in C$.) This part of the reasoning is similar to what we have done above, and it is based on the fact that the area of ½C is equal to $\frac{1}{4}S_C > S_{D(O)}$. Take a point X in the aforementioned intersection. Let Y and Z denote the points where X falls when translated by the vectors **PO** and **QO**, respectively. These points lie in the domain ½C. Furthermore, **OY** – **OZ** = **OP** – **OQ**, and thus point A, the endpoint of the vector **OY** – **OZ**, is a node of the lattice ($A \ne O$, since $P \ne Q$). Now we note that the endpoint of the vector

$$\mathbf{OY} - \mathbf{OZ} = \tfrac{1}{2}(\mathbf{OY} + 2(-\mathbf{OZ}))$$

lies inside figure C, since the figure is convex and O is its center of symmetry.

We can apply the reasoning for Observations 1 and 2 to lattices in three- (and more) dimensional space. In this context we must speak of the "middle (hyper-) planes" of the pairs of the nodes (O, P) and of the volumes of the corresponding "cells" $D_1(O)$ and $D_r(O)$. Figures 6a, 7a, and 8a show the four most common examples of solid lattices: a simple cubical lattice, a volume-centered cubical lattice, and a facets-centered cubical lattice. The corresponding domains $D(O)$ and $D_1(O)$ are shown in figures 6b–8b and 6c–8c, respectively. The polyhedrons in fig. 6c and 7b bear the name of *rhombo-dodecahedrons*, and the polyhedron in fig. 8b is called a *truncated octahedron*.

Crystalline constructions

The geometric constructions that we have focused on play an important role in the physics of solids. The domains $D_r(O)$ that we've considered appear naturally in the study of crystal structure. There they bear the name *Brillouin zones*, after the famous French scientist Léon Brillouin, who in the early 1930s thoroughly investigated electron

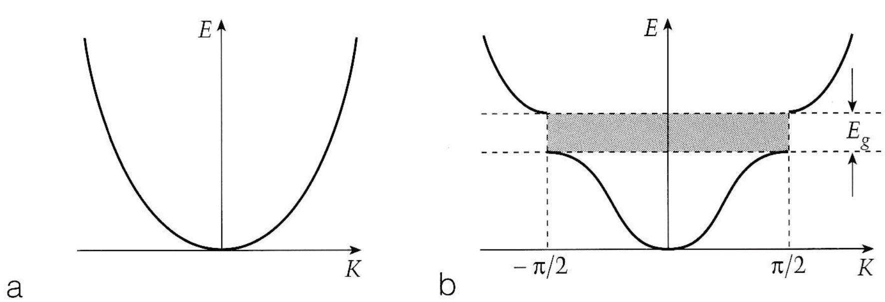


Figure 9

flow in crystals from the quantum point of view.

A crystal's properties of electric conductivity depend primarily on the existence of "energy gaps," which are intervals in the scale of energy that comprise values that the electrons cannot take on. Figure 9a shows how the energy of a free electron (that is, of an electron that does not interact with the crystal) depends on its impulse. Its graph is a parabola (the kinetic energy is proportional to the square of the velocity). If the electron interacts with the ions of the crystal lattice, then breaks can appear in this graph for some values of impulse.

The corresponding graph for a "one-dimensional" crystal is shown in figure 9b. For "two-dimensional" and real, three-dimensional crystals, the impulse **p** is a vector. If we agree to measure all these vectors from the same point O, we'll obtain a space (two- and three-dimensional, respectively) whose points correspond to all possible values of the impulse, the so-called *impulse space*. To each lattice A consisting of atoms in the usual space, one can find a corresponding "dual" lattice P in the impulse space (physicists usually call it the *inverse lattice*). For a properly chosen scale in the impulse space, this lattice is defined by the following rule: Vector **p** from the impulse space belongs to P if and only if the scalar product $\mathbf{a} \cdot \mathbf{p}$ is an integer for all vectors **p** connecting two nodes of lattice A.

It turns out that the breaks in the energy (here we consider it as a function of the impulse space) appear right on the planes, which are perpendicular bisectors drawn to the segments OP, where P is a node of the dual lattice. The parallel translation of the cells that constitute the Brillouin zone $D_r(O)$ of rank r to the main zone $D_1(O)$ also has a physical interpretation.

Exercises

1. Why do the ends of such vectors $\mathbf{p} = \mathbf{OP}$ form a lattice?

2. Draw the dual lattices P for the lattices A shown in figures 1, 2, and 3.

I Can See Clearly Now

by David Arns

A fascinating concept has been shaped in recent years:
An excellent idea that, no doubt, elicits cheers
From optical astronomers who, down here on this Earth,
Greet the news with celebrations, merriment, and mirth.

The problem first was seen more than three centuries ago,
When Christiaan Huygens saw that every heav'nly body's glow,
When looked at through a telescope, was always seen to quiver,
Like seeing a reflection in the ripples of a river.

And later, Isaac Newton noticed similar effects—
"A problem with the optics?" No, this thought he soon rejects;
The problem was the atmosphere: The light's distorting *there*.
"The remedy," he quipped, "is most serene and quiet air."

And so, for years—no, *centuries*—the best that we could do
Was this infernal "twinkling" of the air we must look through.
And photographic plates containing images of stars
Had fuzzy, blobby, blurs that looked five times the size of Mars.

Then came a project: SDI (or "Star Wars," as it's called),
The ground-based super heat-ray had our laser boys enthralled:
"Why, we could shoot down missiles, both the main one and the spare,
If light did not diverge and scatter, going through the air."

They saw that this distortion could be compensated for
By analyzing twinkle (quite a computational chore)
And mirrors that could change their shapes to counteract the blur:
"We'll whip the beam back into shape, and that we know for sure!"

Well, SDI was cancelled, but the knowledge that they'd learned
Was what the doctor ordered, the astronomers discerned:
The ability to counteract the twinkle caused by air—!
The potential of what might be seen was more than they could bear.

So, several of the larger 'scopes—ten or twelve or more—
Will get adaptive optics, which will open up the door
To sharper, clearer images of objects out in space,
From galaxies to quasars to the nebulas' fine lace.

And what will clearer vision do for people, in the main?
Will our thoughts extrapolate like links within a chain?
It would be nice, when looking at a galaxy or star,
To see the cosmos' vastness, and how miniscule *we* are.

David Arns *is a graphics software documentation engineer for Hewlett-Packard in Fort Collins, Colorado, and also operates a small business designing and creating web sites. In his spare time he dabbles in poetry on scientific themes.*

Wave on a car tire

Does the engine limit the maximum speed of a car?

by L. Grodko

ONE OF THE MANY INTRIGUING facets of automobile engineering is the problem of the critical rolling speed of the tire. What is critical speed? What factors affect its value? How can this important parameter be improved? Let's start with a problem that at first glance is not directly related to these questions.

Art by Sergey Ivanov

A wave running on a belt

Consider an endless heavy belt running on two rotating drums with speed V. The segment AB of this belt is threaded through a firmly fixed curved tube (fig. 1). The axis of this tube is a certain curve lying in the page's plane. Find the forces that the tube exerts on the belt. Gravity can be neglected. Assume that there is no friction between the belt and the tube. The tension along the belt is constant with magnitude T. The mass of the belt per unit length is ρ.

Consider a small element of the

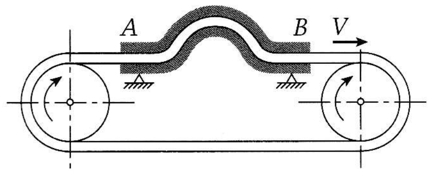


Figure 1

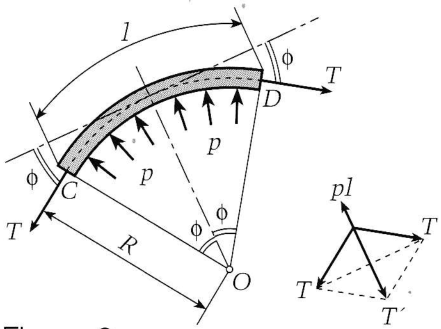


Figure 2

belt CD with length l, which at the given moment is located inside the tube. The segment of the tube that contains this small element can be closely approximated by a circular arc of radius R (fig. 2). The element CD moves with constant speed V along the circle of radius R. Therefore, its centripetal acceleration is V^2/R directed toward point O, which is the center of curvature of the arc CD.[1]

Let's see which forces impart this acceleration to element CD. This element is affected by the tension T due to the neighboring elements of the belt, which are applied to points C and D tangential to arc CD (fig. 2). In addition, this belt is affected by the normal pressure due to the wall of the tube. Let the normal pressure per unit length of the belt be p. Then the force affecting element CD is pl. Here we took into account that $l \ll R$, so the curved element CD is replaced by a straight line segment. According to Newton's second law, the net force F is directed toward the center of curvature O, and it imparts the following centripetal acceleration to element CD:

$$\rho l \frac{V^2}{R} = F, \qquad (1)$$

where ρl is the mass of element CD. Figure 2 shows that $F = T' - pl$, where T' is the equivalent force of two tension forces applied to element CD from adjacent parts of the belt. Since angle ϕ is small, $\sin\phi \cong \phi$, $T' = 2T\phi$, and $l = 2R\phi$. Thus,

$$F = 2T\phi - 2pR\phi.$$

Plugging the formulas for F and l into (1) yields

$$2\rho R\phi \frac{V^2}{R} = 2T\phi - 2pR\phi,$$

from which we get

$$p = \frac{1}{R}\left(T - \rho V^2\right). \qquad (2)$$

We have thus obtained the normal pressure acting on the belt. More specifically, we have found only the fraction of this force that acts on a unit length of the belt in the section where the radius of curvature is R. However, this value is sufficient to make further progress in solving this problem.

Let's analyze formula (2). The expression in parentheses doesn't vary along the tube, so for the given parameters T, ρ, and V, its value is determined by the inverse radius of curvature ($p \propto 1/R$). When $V = 0$ (that is, if the belt doesn't move), $p = T/R$, which means that the pressure is determined by the belt's tension and the tube's curvature at a given point. If the speed V increases, p decreases, and when the speed reaches the critical value

$$V = \sqrt{\frac{T}{\rho}}, \qquad (3)$$

the force cancels along the entire length of the tube. This speed is called the critical speed V_{cr}.

What does the equation $p = 0$ mean? At $V = V_{cr}$ the belt doesn't interact with the tube! At this speed we can take the tube away—the belt will not "notice" and will retain the bend at segment AB.

Relative to an observer who moves to the right at the speed V_{cr} (so the observer doesn't move relative to the upper part of the belt), this bend moves along the belt to the left at the speed V_{cr}. This is a traveling wave.

Now imagine that $V < V_{cr}$ and we take the tube away but apply pressure to the belt on the segment AB such that the pressure is distributed in the same way as it was with the tube. We can see that the bend will be the same as if the tube was in its place. At every point the belt's curvature $1/R$ will relate to the applied forces according to formula (2). If the pressures p are kept constant when the speed is increased, formula (2) predicts that the degree of the belt's bending (characterized by its curvature $1/R$) will also increase. In this process the distribution of $1/R$ along segment AB will be determined by the distribution of pressure along this segment.

When the pressure distribution on segment AB does not vary but the speed increases to the critical value, large bending (curvature) of the belt in segment AB can occur. This inference can be formulated another way. If a steady load (that is, pressure distributed in some way along segment AB) "runs" along the belt with increasing speed (relative to the belt), the curvature of the belt in fragment AB will also increase, and the deformation of the belt caused by the load increases accordingly. When the speed approaches V_{cr}, the curvature of the belt increases infinitely. Therefore, the belt doesn't resist a load at a speed near the critical value.

This conclusion doesn't depend on the character of the load distribution in segment AB just as the formula for V_{cr} doesn't depend on the shape of the curved tube.

Deformation on the roads

The concept of the critical speed of the traveling load can be extended with more qualified reasoning to describe such phenomena as, say, the load of a train car running along the rails. The rail is also bending under the pressure, but the considerations are more complicated, because unlike a belt, a rail has a rigidity, and in addition, lies on an elastic base (the ground). However, the critical speed does exist in this case as well. Calculations show that its value is very high (about 1000 km/h), and it

[1]The *center of curvature* is the center of the circle whose arc is the curved element CD (arc CD). The inverse radius of curvature is called the *curvature*.

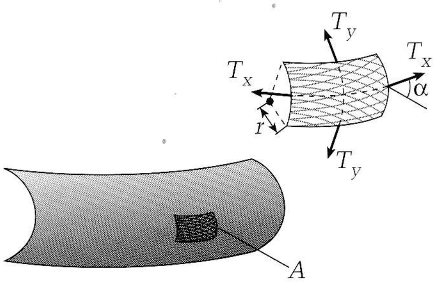


Figure 3

should be taken into account only in designing trains capable of attaining speeds that presently are far from reality.

However, the limitations imposed by the existence of a critical speed have great practical importance for modern cars. The critical speed dictates the maximum speed a car can develop. Curiously, this limitation is imposed not by the engines, but by the tires.

A tire touches the road at what is called the *contact site*. The forces affecting the tire are applied to this contact site, and they constitute the load circulating along the tire. An automobile pneumatic tire is a shell inflated and stretched by the pressure of internal air. The shape of a tire is almost a torus. Let's consider a small element A of the tire's shell (fig. 3). It is affected by neighboring elements with tension directed tangentially along the meridians and parallels. Let the value of these forces per unit length (along the parallels) and unit width (along the meridians) of element A be equal to T_x and T_y, respectively. Analysis of the dependence of these forces on the tire's speed yields the formula for the critical speed of the tire. It looks similar to formula (3) that we obtained earlier for a heavy belt:

$$V_{\text{cr}} = \sqrt{\frac{T_x}{\rho'}}. \qquad (4)$$

Note that in this formula ρ' is the mass per unit area of the shell. Formula (4) shows that the larger the critical speed, the more the tire is stretched in the direction of the parallels.

To enhance the strength of tires, manufacturers reinforce them with a very strong net of cords. Sometimes the tires of big trucks are produced with steel wire reinforcement, and these tires can withstand very large load pressures. Without this wire net a tire would collapse at an extra pressure 20–30 times less than that at which modern reinforced tires would collapse. For passenger cars this pressure is 1.5–2 atm, and for trucks it is 4–6 atm.

The direction of the cords is characterized by the angle α, which affects the value of the critical speed:

$$|V_{\text{cr}}| = K \cot\alpha.$$

Here K is a coefficient determined by the air pressure p inside a tire of radius r in its cross-section and by mass per unit area ρ':

$$K = \sqrt{\frac{pr}{\rho'}}.$$

Thus,

$$V_{\text{cr}} = \sqrt{\frac{pr}{\rho'}} \cot\alpha.$$

This is called Tarner's formula. You can deduce it on your own, based on what you've learned in high school physics.[2]

The value of critical speed calculated according to formula (5) is rather close to the experimental data. For passenger car tires this value is V_{cr} = 150–220 km/h. This figure is surely too small for race cars. Tarner's formula shows how to enhance the critical speed for a tire. One should produce the lightest tires (that is, with smallest value of ρ'), make them thicker (to increase r), pressurize them more, and reinforce them with fibers set at the smallest possible angle α to each other.

A tire's critical speed is a very important parameter, so all tires are subjected to special tests to measure it. In such tests the axis of the wheel is stationary, while the tire rolls on a rotating drum. In the region where the tire emerges from the contact site, one can see waves. The deformation of the tire is very large in such regions. A further increase in speed results in increased deformation (the amplitude of the waves becomes larger). The speed cannot be made precisely equal to the critical value V_{cr}, since the tire collapses before this value is achieved.

The problem of critical rolling speed of tires is the bottleneck in the theory and practice of car design. It is no wonder that it attracts many researchers. This problem has been considered in several hundred theoretical and experimental papers. Still it is far from being solved, and much work is left to be done by *Quantum*. ◘

Quantum articles about deformations and strength:

"Holding up under pressure," A. Borovoy, January, 1990, pp. 30–32.

"All bent out of shape," Kaleidoscope, September/October, 1995, pp. 32–33.

[2]Assume that equilibrium conditions of element A suppose zero curvature in the direction of the parallels: In this way you obtain the formula $T_y = pr$. The condition for no relative motion of the cords (invariance of angle α) results in a relationship between T_y and T_x: $T_x = T_y \cot^2\alpha = pr\cot^2\alpha$.

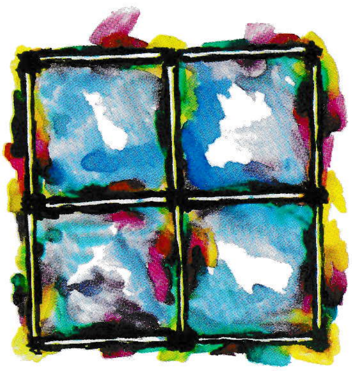
Just for the fun of it!

B246

Twelve days of rest. Each of three successive months has exactly four Sundays. Prove that one of these months is February.

B247

Symbolic gesture. Insert mathematical symbols among the three numbers 4 4 4 so that the resulting expression equals 16.

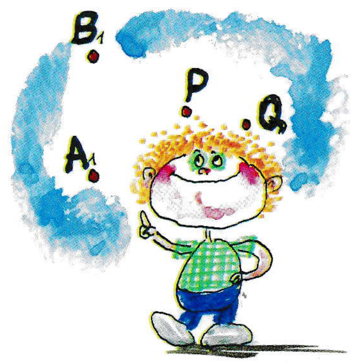


B248

Minimal polygon. Construct a polygon with the minimum possible number of sides, whose sides intersect the 12 segments shown in the figure.

B249

Reconstructive geometry. A student drew a parallelepiped $ABCDA_1B_1C_1D_1$ on the blackboard and labeled the centers of the faces $A_1B_1C_1D_1$ and CDD_1C_1 as the points P and Q. Then he erased the drawing, leaving only four points: A, B, P, and Q. Restore the parallelepiped if you know that segments AA_1, BB_1, CC_1, and DD_1 are the edges.



Art by Pavel Chernusky

B250

Look before you leap. Once upon a time the great Baron Munchausen told the following story. He ran and leapt, intending to jump over a swamp. While in the air he realized that he couldn't reach the opposite bank. Then, still in flight, he decided to turn around so he could safely return to the bank from which he jumped. Is this a true story?

ANSWERS, HINTS & SOLUTIONS ON PAGE 48

Functional equations and groups

A formal introduction

by Y. S. Brodsky and A. K. Slipenko

YOU ARE PROBABLY FAMILIAR WITH FUNCtional equations, although you may not have heard the term. Functional equations are used to define such properties of functions as evenness, oddness, or periodicity.

In general, a functional equation is an equation describing an unspecified funtion. Often it can be "solved" for the unknown function. Some examples are

$$f(x+1)+f(x)=x,$$
$$2f(1-x)+1=xf(x),$$
$$xf(x)+f\left(\frac{4}{2-x}\right)=x.$$

Mathematicians began to study functional equations over 200 years ago, when they appeared in some mechanics problems. Augustin-Louis Cauchy (1789–1857) contributed significantly to the study of such equations. Indeed, there is even an equation named after Cauchy: $f(x+y)=f(x)+f(y)$. In this paper we focus on a method for solving functional equations that uses one of the most important concepts of modern mathematics—the concept of group.

Composition of functions

The number of basic functions studied in school is rather small. Among them are linear and power functions, exponential functions, and trigonometric functions. Other functions are obtained from these basic ones through composition and algebraic operations.

For example, the function $f(x)=\sin(2x+1)$ is the composition of the linear function $g(x)=2x+1$ and the trigonometric function $h(x)=\sin(x)$; that is, $f(x)=(h\circ g)(x)$.

The function $f(x)=\log_{10}\arcsin(x)$ is the composition of the functions $g(x)=\arcsin(x)$ and $h(x)=\log_{10}x$. Note that the domain of the composition $h\circ g$ includes all values x from $D(g)$ for which $g(x)\in D(h)$. In the last example, $D(g)=[-1;1]$, $D(h)=]0;\infty[$. Since $\arcsin(x)>0$ for $x\in\,]0;1]$, we have $D(f)=]0;1]$.

The composition of the same functions taken in a different order—$f(x)=\arcsin(\log_{10}x)$—has another domain: $D(f)=[1/10;10]$. The composition of the fractional linear functions

$$g(x)=\frac{-2x+1}{3x+2}$$

and

$$h(x)=\frac{3x-2}{-x+4}$$

yields the function

$$f(x)=h(g(x))=\frac{3\frac{-2x+1}{3x+2}-2}{\frac{-(2x+1)}{3x+2}+4}=\frac{-12x-1}{14x+7},\ x\neq-\frac{2}{3}.$$

Here

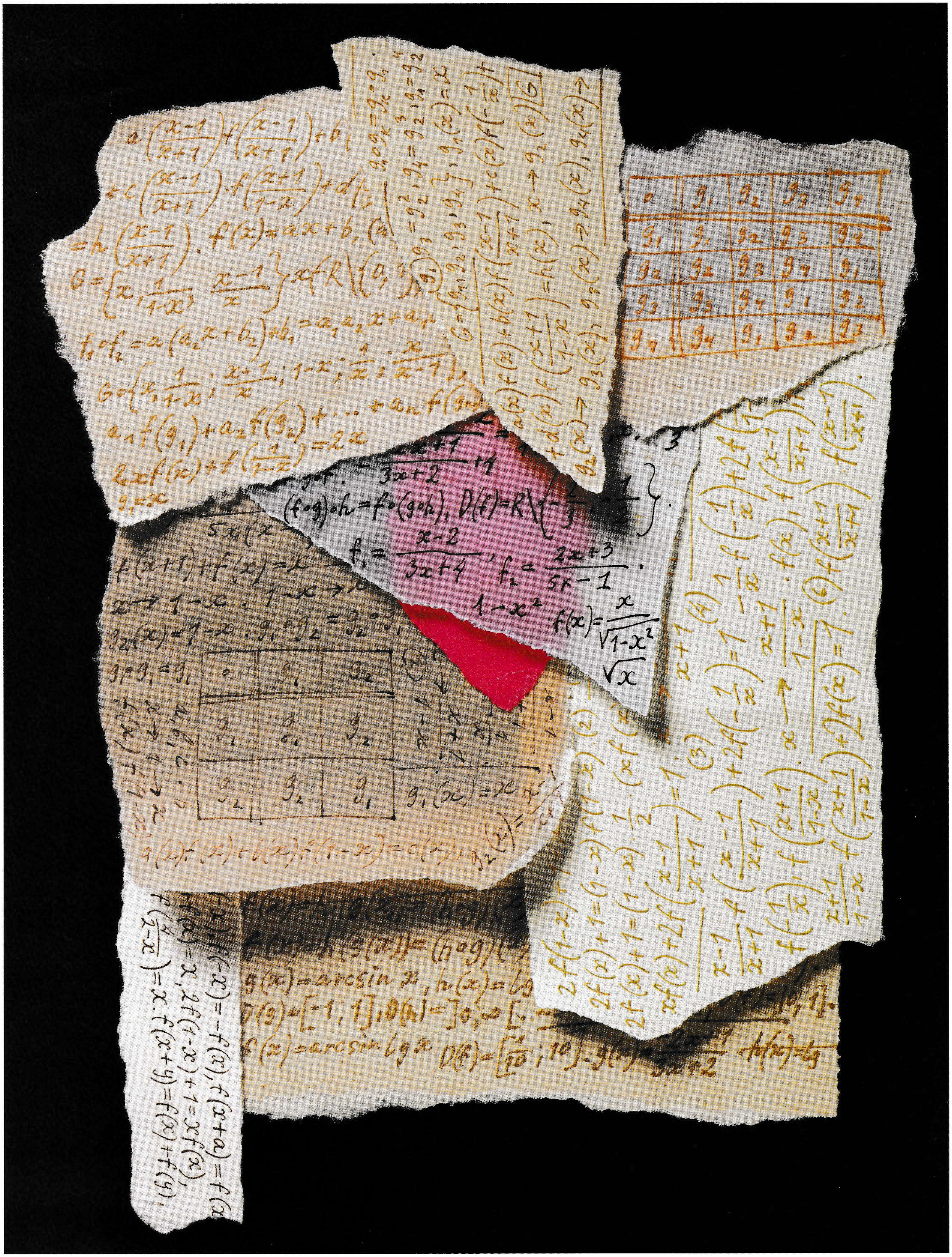
$$D(f)=R\setminus\left\{-\frac{2}{3},-\frac{1}{2}\right\}.$$

As a rule, $f\circ g\neq g\circ f$. At the same time,

$$(f\circ g)\circ h=f\circ(g\circ h),$$

which immediately follows from the definition of the composition.

Art by Vera Khlebnikova



Exercises

1. Find the compositions $f_1 \circ f_2$ and $f_2 \circ f_1$ of the following functions:

$$f_1 = \frac{x-2}{3x+4},\ f_2 = \frac{2x+3}{5x-1}.$$

2. Find the domain of the composition of the functions $1 - x^2$ and $\sqrt{x}$.

3. Let

$$f(x) = \frac{x}{\sqrt{1-x^2}}.$$

Find

$$\underbrace{f \circ f \circ f \circ \ldots \circ f}_{n}.$$

Functional equations

Let us solve the following problem.

Problem 1. Find all functions $y = f(x)$ such that

$$2f(1-x) + 1 = xf(x). \qquad (1)$$

Solution. Suppose that a function that satisfies this equation does exist. Substituting $1 - x$ for x, we obtain

$$2f(x) + 1 = (1-x)f(1-x). \qquad (2)$$

Equation (1) yields

$$f(1-x) = \frac{1}{2}\left(xf(x) - 1\right).$$

Substituting this value of $f(1-x)$ into equation (2), we obtain

$$2f(x) + 1 = (1-x)\cdot\frac{1}{2}\cdot\left(xf(x)-1\right),$$

from which we get

$$f(x) = \frac{x-3}{x^2-x+4}.$$

Direct verification shows that the function obtained satisfies equation (1).

In this equation the functions $f_1 = x$ and $f_2 = 1 - x$ serve as arguments of the unknown function. The substitution of $1 - x$ for x transforms the functions f_1 and f_2 into each other. The substitution $x \to 1 - x$ gave one more equation that involves $f(x)$ and $f(1 - x)$. Thus, we reduced the solution of the functional equation to the solution of a system of two linear equations in two unknowns.

Now, consider a more difficult problem.

Problem 2. Solve the following equation:

$$xf(x) + 2f\left(\frac{x-1}{x+1}\right) = 1. \qquad (3)$$

Solution. We proceed in the same way as in the previous case. Making the substitution

$$x \to \frac{x-1}{x+1}$$

yields

$$\frac{x-1}{x+1}f\left(\frac{x-1}{x+1}\right) + 2f\left(-\frac{1}{x}\right) = 1. \qquad (4)$$

Along with the expressions $f(x)$ and

$$f\left(\frac{x-1}{x+1}\right),$$

we have the new "unknown"

$$-f\left(-\frac{1}{x}\right)$$

here. So we make another substitution in equation (3):

$$x \to -\frac{1}{x}.$$

This yields

$$-\frac{1}{x}f\left(-\frac{1}{x}\right) + 2f\left(\frac{x+1}{1-x}\right) = 1. \qquad (5)$$

Now, in addition to

$$f\left(-\frac{1}{x}\right),$$

we have one more undesirable expression

$$f\left(\frac{x+1}{x-1}\right),$$

in our equation. Try another substitution:

$$x \to \frac{x+1}{x-1}.$$

Finally, we obtain an equation that does not involve new unknowns:

$$\frac{x+1}{1-x}f\left(\frac{x+1}{1-x}\right) + 2f(x) = 1. \qquad (6)$$

Thus, we have constructed a system of four linear equations (3)–(6) in four unknowns

$$f(x),\ f\left(\frac{x+1}{x-1}\right),\ f\left(-\frac{1}{x}\right),\ \text{and}\ f\left(\frac{x+1}{x-1}\right),$$

Successively eliminating the unknowns

$$f\left(\frac{x+1}{x-1}\right),\ f\left(-\frac{1}{x}\right),\ \text{and}\ f\left(\frac{x+1}{x-1}\right),$$

we find that

$$f(x)=\frac{4x^2-x+1}{5x(x-1)}$$

(where $x \neq -1$, $x \neq 0$, $x \neq 1$). As in the solution of equation (1), we assumed that a solution to equation (3) exists. It can easily be verified that f does satisfy equation (3).

How the groups emerge

Let's try to understand how we managed to solve the equations in the preceding section. Consider one more equation

$$f(x+1)+f(x)=x.$$

It does not look more difficult than equation (3). However, all attempts to solve it by the same method are in vain. If we make the substitution $x \to x+1$, the new variable $f(x+2)$ occurs, and so on. The chain does not close, and we never obtain a linear system.

Recall that when we were solving the first equation, we made the substitution $x \to 1-x$. With this substitution, $1-x \to x$. That is, the two functions $g_1(x)=x$ and $g_2(x)=1-x$ behave like $g_1 \circ g_2 = g_2 \circ g_1 = g_2$, $g_2 \circ g_2 = g_2$, and $g_1 \circ g_1 = g_1$ with respect to composition.

Consider the "multiplication" table in Table 1 (in which $g_i \circ g_j$ is at the intersection of row i and column j).

$\circ$	g_1	g_2
g_1	g_1	g_2
g_2	g_2	g_1

Table 1

Each row and column of this table include both g_1 and g_2.

Suppose that we must solve the equation

$$a(x)f(x)+b(x)f(1-x)=c(x), \qquad (*)$$

where a, b, and c are some functions. We can see that the substitution $x \to 1-x$ yields

$$a(1-x)f(1-x)+b(1-x)f(x)=c(1-x), \qquad (**)$$

which, together with equation (*), gives a linear system in unknowns $f(x)$ and $f(1-x)$. Then, the solution continues as for equation (1).

In problem 2, we made the following substitutions:

$$x \to \frac{x-1}{x+1}, x \to -\frac{1}{x}, x \to \frac{x+1}{1-x},$$

That is, we dealt with the functions

$$g_1(x)=x,\ g_2(x)=\frac{x-1}{x+1},$$
$$g_3(x)=-\frac{1}{x},\ g_4(x)=\frac{x+1}{1-x}.$$

Let's see how the functions g_1, g_2, g_3, and g_4 behave with respect to composition. We form Table 2 similar to Table 1 (by writing $g_i \circ g_j$ at the intersection of row i with column k).

$\circ$	g_1	g_2	g_3	g_4
g_1	g_1	g_2	g_3	g_4
g_2	g_2	g_3	g_4	g_1
g_3	g_3	g_4	g_1	g_2
g_4	g_4	g_1	g_2	g_3

Table 2

This table is symmetric about its diagonal. That is, $g_i \circ g_k = g_k \circ g_i$ for any k and j. In addition, all functions g_i occur exactly once in every row and every column. Finally, it is easy to see that $g_3 = g_2^2$, $g_4 = g_2^3$, and $g_1 = g_2^4$. (Here

$$g_2^i = \underbrace{g_2 \circ g_2 \circ \ldots \circ g_2}_{i \text{ times}}.)$$

Thus, the system of functions $G = \{g_1, g_2, g_3, g_4\}$ has the following properties: (a) It is closed with respect to composition; (b) there is the identity mapping $g_1(x)=x$ among these functions; (c) each g_i has an inverse g_i^{-1}: $g_1^{-1}=g_1$, $g_2^{-1}=g_4$, $g_3^{-1}=g_3$, and $g_4^{-1}=g_2$.

The same properties are characteristic of the system of functions $G=\{g_1, g_2\}$ from problem 1.

Now, if we needed to solve any functional equation of the form

$$a(x)f(x)+b(x)f\left(\frac{x-1}{x+1}\right)+c(x)f\left(-\frac{1}{x}\right)+d(x)f\left(\frac{x+1}{1-x}\right)=h(x), \qquad (***)$$

we could do it by making the substitutions $x \to g_2(x)$, $x \to g_3(x)$, and $x \to g_4(x)$, which would give us a linear system of equations. For example, we write the result of the substitution $x \to g_2(x)$. With this substitution, $g_2(x) \to g_3(x)$, $g_3(x) \to g_4(x)$, and $g_4(x) \to g_1(x)$, which gives the equation

$$a\left(\frac{x-1}{x+1}\right)f\left(\frac{x-1}{x+1}\right)+b\left(\frac{x-1}{x+1}\right)f\left(-\frac{1}{x}\right)+c\left(\frac{x-1}{x+1}\right)f\left(\frac{x+1}{1-x}\right)+d\left(\frac{x-1}{x+1}\right)f(x)=h\left(\frac{x-1}{x+1}\right).$$

We make the following definition.

Definition. An arbitrary set of functions G defined on a set M is called a *group* with respect to the operation $\circ$ if it possesses the same properties as the system $\{g_1, g_2, g_3, g_4\}$:

1. For any two functions $f \in G$ and $g \in G$, their composition $f \circ g$ also belongs to G.

2. The function $e(x) = x$ belongs to G.

3. For any function $f \in G$, an inverse function f^{-1} exists, which also belongs to G.

This definition is a particular case of the general definition of the concept of group, which is one of the most important concepts of modern mathematics.

We have already considered two examples of groups. Let us give some more examples.

(a) The set G of linear functions $f(x) = ax + b$, where $a \neq 0$, $b \in R$;

(b) $G = \left\{x, \dfrac{1}{1-x}, \dfrac{x-1}{x}\right\} x \in R \setminus \{0,1\}$;

(c) The set G of functions $f(x) = x + a$.

For example, let's prove that the set of linear functions forms a group. These functions are defined on the set of real numbers R. Let $f_1 = a_1x + b_1$ and $f_2 = a_2x + b_2$. Then

$$f_1 \circ f_2 = a_1(a_2x + b_2) + b_1 = a_1a_2x + a_1b_2 + b_1$$

is a linear function. The function $e(x) = x$ is also linear. If $f(x) = ax + b$, then the linear function

$$f^{-1} = \frac{x}{a} - \frac{b}{a}$$

is inverse to f.

Exercises

4. Prove that the sets of functions (b) and (d) form groups.

5. Does the set of functions

$$G = \left\{x, \frac{1}{1-x}, \frac{x-1}{x}, 1-x, \frac{1}{x}, \frac{x}{x-1}\right\}$$

where $x \in R \setminus \{0, 1\}$ form a group with respect to composition?

Summing up

Now we can present a general method for solving certain functional equations. This method is based on the concept of a group of functions.

In the functional equation

$$a_1f(g_1) + a_2f(g_2) + \ldots + a_nf(g_n) = b \qquad (7)$$

let the arguments of the unknown function $f(x)$ be elements of a group G consisting of n functions $g_1(x) = x$, $g_2(x)$, ..., $g_n(x)$ and let the coefficients a_1, a_2, ..., a_n, and b in (7) also be some functions of x. Assume that equation (7) has a solution. Make the substitution $x \to g_2(x)$. As a result, the sequence of functions $g_1, g_2, \ldots, g_n$ turns into the sequence $g_1 \circ g_2, g_2 \circ g_2, \ldots, g_n \circ g_2$. It turns out that such a sequence contains exactly one copy of each group element. (This result is important, but easy to prove. The proof is left to the reader.) So the new sequence consists of all the elements of the group, but in a different order.

Therefore, the "unknowns" $f(g_1)$, $f(g_2)$, ..., $f(g_n)$ are reordered and we obtain a new linear equation of the same form as (7). Then, we make the substitutions $x \to g_3(x)$, $x \to g_4(x)$, ..., $x \to g_n(x)$ in (7) to obtain a system of n linear equations. If it has any solutions, we must check that they satisfy equation (7).

Consider, by way of example, the following equation:

$$2xf(x) + f\left(\frac{1}{1-x}\right) = 2x. \qquad (8)$$

The set of functions

$$g_1 = x,$$
$$g_2 = \frac{1}{1-x},$$
$$g_3 = \frac{x-1}{x}$$

forms a group with Table 3.

$\circ$	g_1	g_2	g_3
g_1	g_1	g_2	g_3
g_2	g_2	g_3	g_1
g_3	g_3	g_1	g_2

Table 3

Substituting x for

$$\frac{1}{1-x}$$

and for

$$\frac{x-1}{x},$$

in (8) we obtain the following system of equations:

$$\begin{cases} 2xf_1 + f_2 = 2x, \\ \dfrac{2}{1-x}f_2 + f_3 = \dfrac{2}{x-1}, \\ \dfrac{2(x-1)}{x}f_3 + f_1 = \dfrac{2(x-1)}{x}, \end{cases}$$

where

$$f_1 = f(x),$$
$$f_2 = f(g_2(x)) = f\left(\frac{1}{1-x}\right),$$
$$f_3 = f(g_3(x)) = f\left(\frac{x-1}{x}\right).$$

Solving this system, we obtain

$$f_1 = f(x) = \frac{6x-2}{7x}$$

for $x \neq 0$, $x \neq -1$. Verification shows that this function satisfies (8).

In conclusion, we list several examples of groups of functions that can be used for solving functional equations:

$$G_1 = \{x,\ a - x\},$$

$$G_2 = \{x,\ a/x\} \text{ (here and in what follows } a \neq 0\text{)},$$

$$G_3 = \left\{x, \frac{a}{x}, -x, -\frac{a}{x}\right\},$$

$$G_4 = \left\{x, \frac{1}{x}, -x, -\frac{1}{x}, \frac{x-1}{x+1}, \frac{1-x}{x+1}, \frac{x+1}{x-1}, \frac{x+1}{1-x}\right\},$$

$$G_5 = \left\{x, \frac{a^2}{x}, a - x, \frac{ax}{x-a}, \frac{ax-a^2}{x}, \frac{a^2}{a-x}\right\},$$

$$G_6 = \left\{x, \frac{x\sqrt{3}-1}{x+\sqrt{3}}, \frac{x-\sqrt{3}}{x\sqrt{3}+1}, -\frac{1}{x}, \frac{x+\sqrt{3}}{1-x\sqrt{3}}, \frac{x\sqrt{3}+1}{\sqrt{3}-x}\right\}.$$

Exercises

6. Solve the following functional equations:

(a) $xf(x) + 2f\left(-\frac{1}{x}\right) = 3,$

(b) $f\left(\frac{x}{x-1}\right) + xf\left(\frac{1}{x}\right) = 2,$

(c) $f(x) + f\left(\frac{x-1}{x}\right) = 1 - x.$

7. Find $f(x)$ if

$$af(x^n) + f(-x^n) = bx,$$

where $a \neq 1$ and n is an odd number.

8. Find a function that is defined for $x \neq 0$ and satisfies the equation

$$(x-2)f(x) + f\left(-\frac{2}{x}\right) - xf(2) = 5.$$

9. Find at least one function that satisfies the equation $f(f(f(x))) = -1/x$, but does not satisfy the equation $f(f(x)) = -x$.

10. Prove that if $G = \{g_1 = x, g_2, \ldots, g_n\}$ is a finite group of functions with the composition operation, and $\phi = \phi(x)$ is an arbitrary invertible function, then the set

$$G_\phi = \left\{\phi^{-1} \circ g_1 \circ \phi, \phi^{-2} \circ g_2 \circ \phi, \ldots \phi^{-1} \circ g_n \circ \phi\right\}$$

is also a group with the composition operation (the usual note on the domain of definition of each element holds).

This article discussed only one of the many methods that have been developed for solving functional equations. Many such equations will not fall to this method, and will require such concepts as limits and continuity. However, it does show how a simple "trick" that works with certain functional equations generalizes to a powerful method.

Sea waves

Troughs and crests from head to toe

by L. A. Ostrovsky

A VAST NUMBER OF BEAUTIful poems and songs are devoted to the oceans with their eternally moving surfaces. The sea is a symbol of formidable elements, of inconsistency, danger, and inconceivable variety. From ancient times ocean waves have induced human beings to ponder and invent new scientific theories. Wave physics itself was born in the ocean, and it seems that most of our knowledge of oscillatory behavior and of all types of waves originated in the first attempts to understand the nature of ocean waves.

The running sinusoid

Many people can readily explain the motion of solid bodies, which is studied in elementary physics courses. However, to understand wave motion is quite another matter. Waves begin with some disturbance such as a moving ship, a pebble dropped in water, or a gust of wind. The motion of the water particles induced in one place is imparted to neighboring particles, and in due time the initial disturbance spreads over a large distance.

While a wave travels a long distance, the displacement of the water particles is relatively small. The energy is transferred from one particle to another, and so on, like a baton in a relay race.

To understand the properties of wave motion, it is convenient to

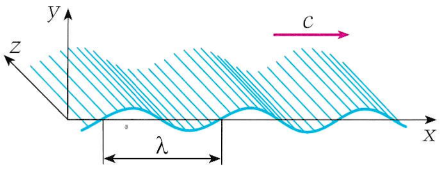


Figure 1. *The sinusoidal surface of water looks like a moving corrugated metal sheet.*

simplify it. In most cases this is done by depicting the motion as sinusoidal. Imagine on the water surface a sinusoidal corrugation that moves along the x-axis at a constant speed (fig. 1). If this corrugation is cut by a vertical plane that is parallel to the x-axis, the resulting cross-section will be a sinusoid.

The distance between two neighboring crests or troughs (or between any two neighboring points that have the same phase) is the same: It is the wavelength λ. As the wave moves along the x-axis at a constant speed, any point of the medium repeats its displacement after a certain time T, which is called the period of oscillation. During this period, the wave travels the distance λ along the x-axis. The speed of this process $c = \lambda/T$ is called the *phase speed*.

There are no infinite sinusoids in nature. Any wave motion is initiated at some place and dies away at another. Often a wave remains almost sinusoidal for a long time, but this "almost" is essential.

To grasp the point, let's consider two sine waves with different, but very similar periods T_1 and T_2 (and, therefore, with almost identical wavelengths λ_1 and λ_2). If the crests (and troughs) of these waves arrive simultaneously at the same point, the resulting trough-to-crest height of the oscillations (twice the amplitude) increases markedly. However, as the waves move away from this point, their wavelength difference results in the accumulation of phase difference, so the crest of one wave will eventually coincide with the trough of another. In this case the waves cancel each other. At still larger distances the cancellation is replaced by an increase in the amplitude. These increasing and decreasing effects are periodically repeated along the x-axis. You can see it for yourself by superimposing two sinusoids with equal amplitudes but just slightly different periods (fig. 2*a*). Figure 2*b* shows the sum of these sinusoids, which looks like a sinusoid with a period close to both T_1 and T_2 but with a regularly varying amplitude. This process is known as "beating," and the resulting wave is called "modulated."

The change in amplitude of the beats is described by a wave that

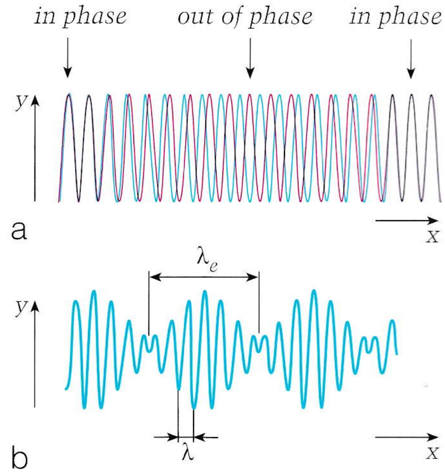


Figure 2. *The sum of two almost identical sinusoids results in an unexpectedly complicated "beating" effect.*

"envelopes" the basic (or carrier) waves. We can easily find the wavelength λ_e of this envelope and its period T_e. The crest of one wave coincides with the trough of another if the number of cycles of the constituent sinusoids between, say, two adjacent maximums of the envelope, differs by one. For the first wave, the number of such cycles is λ_e/λ_1, and for the second it is λ_e/λ_2. Therefore,

$$\frac{\lambda_e}{\lambda_1} - \frac{\lambda_e}{\lambda_2} = 1,$$

from which we get

Art by Jose Garcia

$$\lambda_e = \frac{\lambda_1\lambda_2}{\lambda_2 - \lambda_1} \cong \frac{\lambda^2}{\Delta\lambda}. \qquad (1)$$

Here $\Delta\lambda$ is the difference $\lambda_2 - \lambda_1$. We replaced the square of the geometrical mean of λ_1 and λ_2 by the product $\lambda_1\lambda_2$,

$$\lambda^2 = \left(\sqrt{\lambda_1\lambda_2}\right)^2,$$

which is an approximation of both λ_1 and λ_2 and of the carrier wavelength. (The precise calculation yields the carrier wavelength

$$\lambda_c = \frac{2\lambda_1\lambda_2}{\lambda_1 + \lambda_2}.)$$

In a similar way, we can obtain the formula for the period of the beats:

$$T_e \propto \frac{T^2}{\Delta T}.$$

Now we can find the speed of the envelope, which can also be described as the speed at which the beating propagates:

$$v = \frac{\lambda_e}{T_e} = \frac{\lambda^2}{T^2}\frac{\Delta T}{\Delta\lambda} = c^2\frac{\Delta T}{\Delta\lambda}. \qquad (2)$$

Here c is the phase speed of the carrier wave. If the speeds of both sine waves that compose the carrier wave are exactly equal ($\lambda_1 = cT_1$ and $\lambda_2 = cT_2$), then $\Delta\lambda/\Delta T = c$, so $v = c$. In this case the entire wave travels with the speed c. When $c_1 \neq c_2$, the speed v will differ from c, which means that the carrier wave propagates at one (phase) speed while the envelope wave travels at another speed v. The latter speed has a special name: the *group speed*. Energy is transferred via this group speed.

Like a sinusoid, the envelope is infinite both in time and space. However, if we take not two, but many sine waves (strictly speaking, infinitely many) that have similar wavelengths and periods, we can obtain a so-called "wave packet," or oscillations that occupy only a limited space (fig. 3). As in the previous case, such a packet travels at the group speed while the carrier wave propagates at the phase speed. However, this motion cannot go on forever. If the phase speeds of various sinusoids are different, then according to equation (2), the group speed corresponding to different pairs of sinusoids are not the same. For this reason, at large distances the wave packet will change its shape and finally will be deformed, flattened, and "smeared" in space. This phenomenon is known as dispersion or scattering. Thus, to describe wave propagation in a medium with dispersion we need not one, but at least two different velocities!

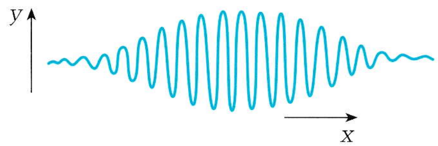


Figure 3. *This wave "packet" is limited in space, but it is the sum of an infinite number of elementary sinusoids, which extend to infinity.*

Waves on water

Let's return to waves on water. To describe them precisely, we need the equations of motion for the water particles. To solve these equations is another problem. However, many features of water waves can be understood without equations if we use dimensional analysis.[1]

What physical values can affect the speed of propagation of sinusoidal surface waves on water? Neglecting wind and assuming the body of water to be infinitely deep, we have only the water density ρ, the acceleration due to gravity g, and some parameter of the wave itself (wavelength λ or period T) at our disposal.[2]

The dimensions of these values are

$$[\rho] = \text{kg/m}^3,\ [g] = \text{m/s}^2,\ [T] = s.$$

Which combination of these parameters yields the dimensions of speed? Since $[c]$ = m/s and the dimension of mass appears only in density $[\rho]$, the density cannot enter the formula that we are constructing. Therefore, only two values are at our disposal, and there is just one combination that yields the dimensions of speed: $[c] = [gT]$. This approximate formula describes both the phase and group velocities.

Now we can say that c and v are proportional to gT. The coefficient of proportionality cannot be found using dimensional analysis. The rigorous theory says that for the phase speed this coefficient is ½ π:

$$c = \frac{gT}{2\pi}. \qquad (3)$$

By definition $c = \lambda/T$, so equation (3) yields the relationship between the wavelength and the period:

$$\lambda = \frac{gT^2}{2\pi}.$$

This formula is known as the dispersion equation.

There is another way to write the formula for c:

$$c = \sqrt{\frac{g\lambda}{2\pi}}.$$

And what about the group speed? According to equation (2), we need the ratio $\Delta T/\Delta\lambda$, which can be obtained from the dispersion equation

$$\Delta\lambda = \frac{g}{\pi}T \cdot \Delta T \Rightarrow \frac{\Delta T}{\Delta\lambda} = \frac{\pi}{gT}.$$

Inserting this into equation (2) and taking into account equation (3), we get

$$v = c^2\frac{\Delta T}{\Delta\lambda} = \frac{gT}{4\pi} = \frac{1}{2}c. \qquad (4)$$

Thus, for waves in deep water, the group speed is half of the phase speed, and both values depend on the wave period. That is, the wave is characterized by a dispersion.

Let's consider the formulas relating the wavelength of a deep water

[1] *Quantum* considered dimensional analysis in "The Power of Dimensional Thinking," Y. Brook and A. Stasenko, May/June 1992, pp. 34–39.

[2] There is still another factor that can affect the wave velocity—the surface tension σ. However, calculations show that it is essential only for very short waves (about 1 cm in length), which are beyond the scope of this article.

T(s)	0.5	1	5	15	20	50
λ(m)	0.39	1.56	39	156	620	3900
v(m/s)	0.39	0.78	3.9	7.8	15.1	39

Table 1

wave and its group speed with the period:

$$\lambda = 1.56T^2,$$
$$v = \tfrac{1}{2}c \cong 0.78T.$$

(Here T is measured in seconds, λ in meters, and the velocities in meters per second). Table 1 shows the actual wave parameters.

It is interesting that comparatively short waves propagate at a pedestrian's speed, yet long waves can outrun a car. If we take a large period—say, $T \cong 1$ h, which corresponds to $\lambda \cong 20{,}000$ km—we get the fantastic speed $v \cong 10{,}000$ km/h. However, the oceans are not deep enough to provide a place for such formidable waves—they would reach the ocean floor even in its deepest locations, so our formulas would not be true any more.

The formulas also cannot be used for very short waves with a wavelength of less than a few centimeters, because the leading factor in this range is surface tension. Still, these formulas correctly describe waves in a very broad range of wavelengths—from dozens of centimeters to dozens (if not hundreds) of kilometers.

Paradoxically, a wave is formed by the motion of water particles, but these very particles are never carried far by the wave! Look at a float, for example. It is not moved horizontally by the waves; it moves up and down in small-amplitude oscillation in the same spot.

To see how water particles move in the wave process, it is convenient to substitute the standard frame of reference for one that travels at the phase speed of the wave. In this frame of reference, the corrugated water surface will appear still, but the entire water mass will flow in the opposite direction at speed c. The water particles slide on the wavy surface just as any ball rolls on a rigid corrugated surface. At the top of the wavy profile, such a ball (or water particle) will have a somewhat smaller speed than the average value c, but having rolled to the bottom of a trough, it compensates for the loss.

In this frame of reference, we have not only the uniform motion of the entire water mass but also the periodic movement of the water particles in the horizontal and vertical directions. Thus, if we return to the initial "ground-based" system, the uniform motion of the water mass will be stopped, but the horizontal and vertical oscillation will continue, because the accelerated (and decelerated) part of the motion is the same in all inertial systems. The sum of the vertical and horizontal oscillation of water particles results in circular motion with a radius equal to the wave amplitude (fig. 4).

The particles that lie under the surface of the water also describe circles, but circles of smaller radii.[3]

The amplitude of this oscillation decreases exponentially with depth. That is, it varies in proportion to $e^{-2\pi z/\lambda}$, where z is the depth. This means that when the depth equals the wavelength, the wave amplitude decreases by $e^{2\pi} \cong 535$ times in comparison with the surface oscillation. Therefore, even strong storms have virtually no effect in oceans at depths of more than 100 m or so. In a pond, when the value of λ is less then $2\pi d$ (where d is the depth of the pond), the pond can be considered infinitely deep, and the wave will not disturb its bottom.

Surely, in the real conditions of the ocean, waves are not sinusoidal, but are composed of many different elementary sinusoids. Therefore, the motion of the water particles is also much more complicated: It consists of revolutions with various velocities along circles of different radii. When waves are weak, such a motion is limited and approximately symmetrical in the vertical direction. However, a stronger wave, even a strictly periodical wave, will

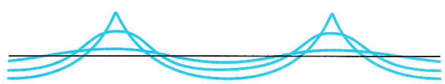


Figure 5

not be sinusoidal: Its crests are sharper than its troughs (fig. 5). With a further increase of the wave amplitude, a sharp bending appears on the wave's crest with an angle of about 120°, and this top disintegrates and falls down. This is how whitecaps, or white horses, form on the crests of waves. When wind is strong, lakes, rivers, seas, and oceans are covered with whitecaps.

Wind waves

At the end of the nineteenth century the famous English physicist Sir John William Rayleigh noted that "the basic law of sea undulation is the absence of any law." Much water has flowed under the bridge

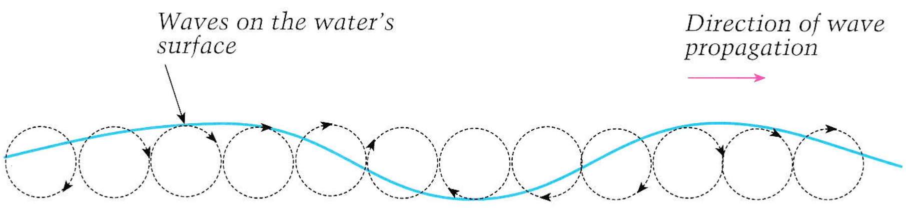


Figure 4

[3]*Quantum* wrote about the damping of waves with depth in "The Bounding Main," I. Vorobyov, May/June 1994, pp. 20–25.

since Sir John's times, and now a descriptive classification of wind and the sea surface is available for seamen. The wind force is characterized by the 12-point Beaufort scale, and the wave height by a 9-point scale of surface waviness.

These scales are mostly based on verbal descriptions of the effects of the wind, such as "sea like a mirror" and "fairly frequent white horses." This classification reminds us of the famous Pushkin fairy tale "The Fisher and the Little Golden Fish." At the beginning, the old fisher saw that the sea was "just troubled." Then the Blue Sea became turbid and black. And finally a "black storm" broke out—"the angry wave ran, howled, and wailed terribly." It seems that we could assign Beaufort numbers to the waviness with which the Golden Fish replied to each subsequent request of the poor old man!

Recently, the methods of measuring sea waves have radically improved. In addition to wave recorders that register the oscillation of the sea surface near a ship, devices that record waves from a distance become more and more important. Even a simple photo of the sea surface made from a ship or a plane provides a large amount of data. Marine radiolocation is also efficient. To crown these methods, "space oceanography" is rapidly developing, drawing on the measurements made by satellites.

As a result of many years of study, a vast amount of information has accumulated on sea waves. However, even nowadays it is not an easy matter to get answers about how wind generates waves in water. Though the basic mechanisms have been deciphered, there is still much to be discerned.

Imagine a smooth water surface. The wind begins to blow, and waves appear. Why? The wind is not a regular uniform flow of air along the water surface. It always contains random fluctuations in pressure, which disturb the water and make its surface curved. When these fluctuations act haphazardly, they will probably raise no waves. However, the case is quite different when the wind's speed is close to the phase speed of a wave that it causes. In this case a resonance takes place. The water's surface begins to oscillate in time with the air pulsation, so the moving wave increases continuously. As the wave amplitude increases, the wave begins to affect the fluctuations in the air flow, amplifying them and increasing the water wave in return. This positive feedback occurs more rapidly as a wave gets bigger.

Such an increase in amplitude can go on for a long time, after which a wave can be rather tall. However, this process will necessarily stop, even when the wind keeps blowing. First, the crests of the waves will be sharpened (as in fig. 5), and finally they will collapse into whitecaps. Second, waves do not live independently. Rather, they interact with each other, transferring energy, which is thereby redistributed among the waves in a very complicated way.

As a result, a realistic and quite complicated physical picture of waves caused by wind is formed, which includes waves of various amplitudes, wavelengths, and even directions of propagation. Nevertheless, there are waves that prevail in this picture. The phase speed of them is close to the wind's speed u, so the respective wavelength is about $2\pi u^2/g$. Thus, the stronger the wind, the larger the wavelength—and the wave's height. In general, the height of smooth waves is no larger than 1/7 of their wavelengths, since larger crests fall down and produce whitecaps. A weak wind can cause only small waves—though a very weak wind produces no waves at all.

As we mentioned, when short waves are considered, surface tension should be taken into account. It's effect results in a particular dependence of the phase speed on wavelength. This speed doesn't tend to zero, but it reaches some minimal value at the wavelength of 1.73 cm, and then it increases with further decreases of λ. For pure water without surface films, the minimum phase speed is 23 cm/s. This value is the threshold speed for wind to raise water waves. The corresponding waves are rather short, and their wavelengths are no more than a few centimeters. We can easily observe such waves in any pool as the ripples that appear during a short gust of wind.

It should be noted that sometimes rather large sine waves do appear on the sea's surface even in mild weather. These swells are the far cry of a distant storm. In the center of a storm there is a very complicated wave pattern, but only the long waves have a chance to reach distant lands, because they have the largest group speed and, more critically, they fade very slowly. These swells sometimes travel thousands of kilometers, and they cause ships to rock and pitch unpleasantly.

A sea of waves

The physics of ocean waves poses more questions than it provides answers. In this article we did not plunge into the depths of the water wave machinery. We "swam" on the surface, because the waves caused by winds disturb only a thin upper layer of the ocean, with thickness of dozens of meters. These waves are very short compared to the depth of the ocean. However, there is another type of wave, the tsunami, whose length is larger than the depth of most deep places in the oceans. These waves embrace the entire depths of the ocean's waters and propagate with the largest speed corresponding to the ocean's depth d: $c = v = \sqrt{gd}$. And this is not the end of the story. In deep waters, there are the low-speed, so-called "internal waves" with velocities of no more than 1–2 m/s, but which have huge amplitudes of tens and sometimes hundreds of meters. Well, the speed of one meter per second is not the shortest value for the ocean waves. Indeed, Rossby waves travel at a speed of 1–2 cm/s and have periods of a number of months! Such waves depend upon the rotation of Earth. These waves are of great importance for physicists, oceanographers, and meteorologists, but they are the makings of a separate article.

Challenges

Math

M246

Divisor champ. Find the three-digit number that has the greatest number of different divisors.

M247

Irrational values. Does a nonlinear function exist that is defined and differentiable for all real numbers and takes rational values at rational points and irrational values at irrational points?

M248

Pyramidal angles. A triangular pyramid $ABCD$ has the equilateral triangle ABC as its base, and $AD = BC$. If the three plane angles at vertex D are equal, what values can these angles take on?

M249

Integration cogitation. Calculate

$$\int_0^2 \left(\sqrt{1+x^3}+\sqrt[3]{x^2+2x}\right)dx.$$

M250

Diametric proportion. The perimeter of $\triangle ABC$ is k times greater than side BC, and $AB < AC$. A diameter of the inscribed circle is drawn perpendicular to BC. In what proportion does the median to BC divide this diameter?

Physics

P246

Sliding wedges. A wedge of mass M with a 45° base angle is placed on a flat horizontal table. A wedge of the same mass and base angle is set on the first such that its upper face is horizontal (fig. 1). A brick of mass m is placed on this wedge, and the system is held motionless. What velocity will the brick acquire in the time τ after the system is set free? Neglect friction and assume the character of the motion doesn't vary. (A. Zilberman)

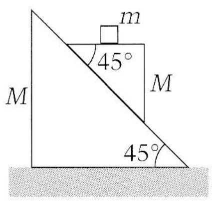


Figure 1

P247

Heat in a tree. Thermal conductivity is two times greater along the fibers of a tree than across them. Two long, thin cylinders of the same size are made of this tree. The axis of the first goes along the fibers, and the axis of the second makes a 30° angle with them. The sides of the cylinders are thermally insulated. Equal temperature differences are applied between the bases of the cylinders. How much does the flow of thermal energy differ in these cylinders? (S. Varlamov)

P248

Dual coils. A capacitor C and two inductors L_1 and L_2 are connected in parallel (fig. 2). Initially, C is not charged and there is no current in L_2, but the current I_0 flows in L_1. Find the max. charge on the capacitor and the max. current at point A. (A. Zilberman)

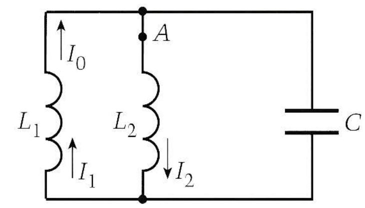


Figure 2

P249

Magnetic fall. An initial horizontal velocity is imparted to a square wire frame with perimeter $4a$ and mass m. The frame moves in a vertical plane under the influence of a magnetic field directed perpendicular to this plane (fig. 3). The field's magnetic induction varies according to the formula $B(z) = B(0) + kz$, where k = const. The resistance of the frame is R. The frame's velocity eventually assumes a constant value v. Find the frame's initial velocity. The acceleration due to gravity is g. (V. Mozhayev)

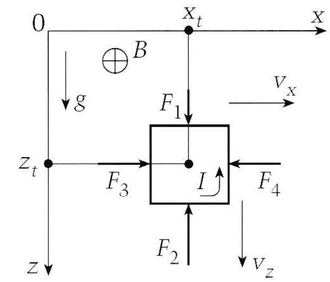


Figure 3

P250

Sunlit grains. Each square meter of a body's surface heated to the temperature T radiates $L = 5.67 \cdot 10^{-8}\, T^4$ W per unit time. At what distance R from the Sun will iron filings melt if the density of solar radiation (energy incident on unit area per unit time) at Earth's orbit is $L_0 = 1400$ W/m^2? The melting point of iron is $T_0 = 1535$ K, and $R_0 = 1.5 \cdot 10^{11}$ m is the distance between the Sun and Earth. (A. Stasenko)

ANSWERS, HINTS & SOLUTIONS ON PAGE 45

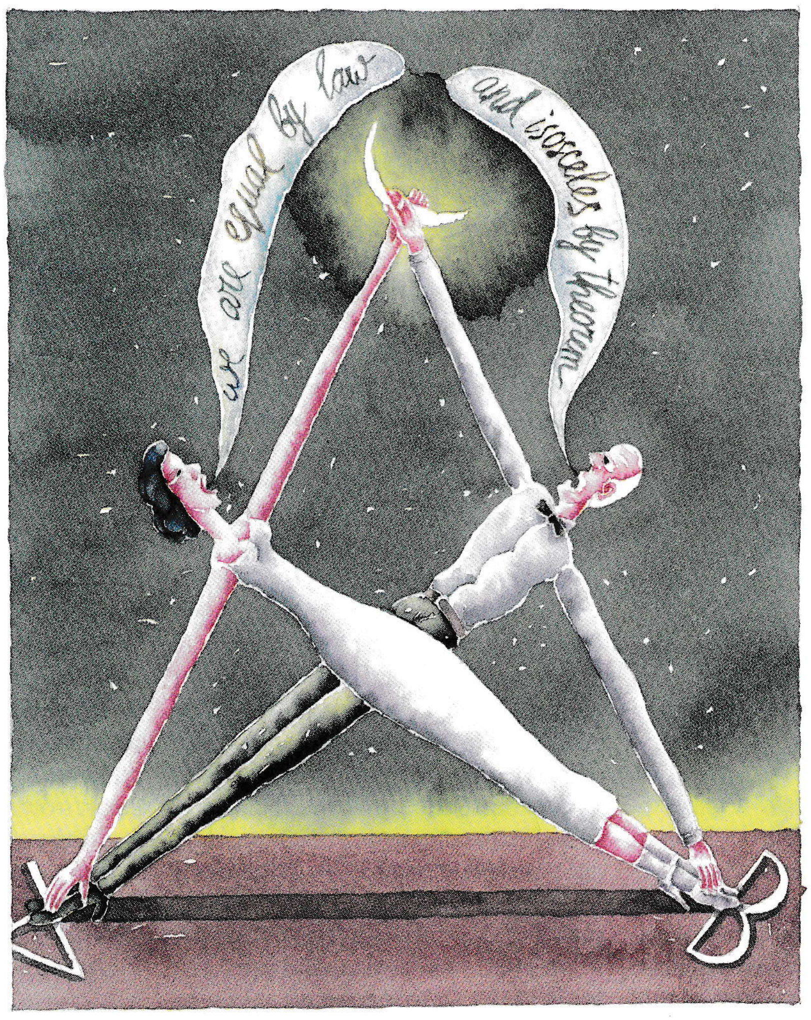
AT THE BLACKBOARD I

The Steiner–Lehmus theorem

by I. F. Sharygin

VERY EARLY IN OUR STUDY of geometry, we get to know three important line segments related to any triangle: its median, its altitude, and its angle bisectors. Of these, angle bisectors are the most troublesome, and thus we will take a closer look at them here. In what follows, we will refer to them simply as *bisectors*. Even the formula for the length of a bisector is rather difficult to derive—much more difficult than the corresponding formulas for the median and altitude.

The problem of constructing a triangle from its medians is not very difficult. It is a little more difficult to construct a triangle from its altitudes. However, it turns out to be impossible to construct a triangle from its three bisectors (using only a compass and straightedge), even though this triangle is uniquely determined (a proof of this fact involves deep properties of polynomial equations). Many facts similar to those that can easily be proved for medians and altitudes are difficult to prove or do not hold when applied to bisectors. The Steiner–Lehmus theorem is an impressive and well-known example. Though this theorem has been known for a long time, it still attracts attention.



Art by Sergey Ivanov

Steiner–Lehmus theorem: If two bisectors of a triangle are equal, then the triangle is isosceles.

Many proofs of this theorem are known. Competitions have even been held among geometry buffs for the most interesting and fresh proof. Here we give a proof that is not very elegant, but reveals the geometric essence of this theorem and makes it possible to obtain more general facts. This proof is based on the following criterion for congruence of triangles: Two triangles are congruent if the following pairs of corresponding parts are equal: a side, the angle opposite this side, and the bisector of this angle.

Let's prove this. Arrange the two triangles under consideration such that their equal sides coincide (this common side is denoted by BC) and the opposite vertices lie on the same side of BC and on the same side of the perpendicular bisector of BC. Suppose that these vertices do not coincide and call them A_1 and A_2 (fig. 1).

Then we draw the circumcircle of triangle A_1BC. Be-

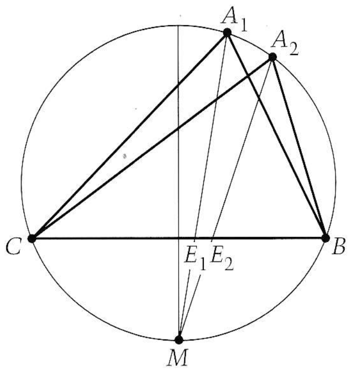


Figure 1

cause $\angle BA_1C = \angle BA_2C$, it follows that point A_2 is also on this circumcircle. (In fact, it is generally true that two triangles that agree in a side and the angle opposite this side will have equal circumcircles.)

Next, draw bisectors A_1E_1 and A_2E_2 in triangles A_1BC and A_2BC. By assumption, $A_1E_1 = A_2E_2$. Extend these bisectors to the point M at which they meet the common circumscribed circle of the triangles (both bisectors meet the circle at the same point M, the midpoint of arc BC). Draw the diameter of the circle through point M—it is perpendicular to BC. In the situation shown in figure 1, we have $MA_2 < MA_1$, because MA_2 is farther from the center of the circle than MA_1. At the same time, $ME_2 > ME_1$, because the projection of ME_2 onto BC is greater than the projection of ME_1. Subtracting the second inequality from the first, we discover that $A_2E_2 < A_1E_1$, which contradicts the assumption. Thus, the congruency test is proved.

Now we can prove the Steiner–Lehmus theorem. Let bisectors AA_1 and BB_1 of a triangle ABC be congruent (fig. 2). Call the point of their intersection I. Triangles ACA_1 and BCB_1 are congruent by the test just proved (they have the common bisector CI of the angle C). Therefore, $AC = BC$. The theorem is proved.

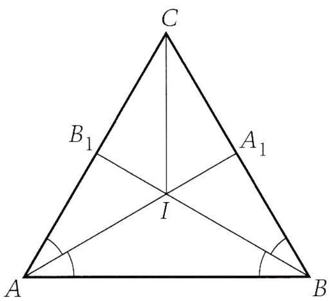


Figure 2

The proved congruency test implies a more general statement: If in a triangle ABC points B_1 and A_1 lie on sides AC and BC, respectively, and segments AA_1 and BB_1 are congruent and meet on the bisector of angle C, then $AC = BC$ (fig. 3).

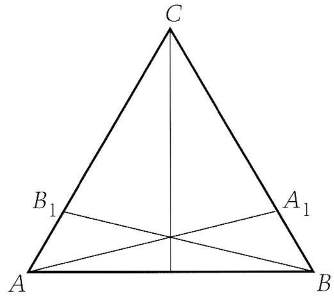


Figure 3

As we know, a triangle has bisectors of its exterior angles as well. If the triangle is not isosceles, we can consider the three segments that are its exterior angle bisectors. (The bisector of the exterior angle opposite to the base of the isosceles triangle is parallel to the base and does not meet it).

Suppose two exterior angle bisectors of a triangle are congruent. Must this triangle be isosceles?

It turns out that if the bisectors in question are drawn from the endpoints of the shortest or longest side of the triangle, the statement is true (figs. 4a, b). However, this is not so in the general case. The well-known Bottema triangle is an example: This is the triangle with the angles 12°, 132°, and 36°. In this triangle the exterior bisectors of the angles at the endpoints of the middle-size side are equal to each other and equal to this middle-size side (fig. 5).

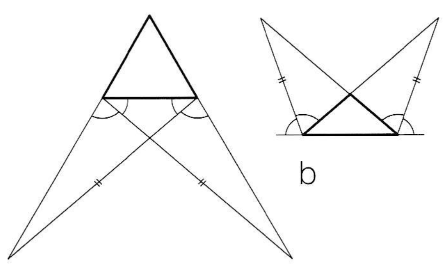


Figure 4

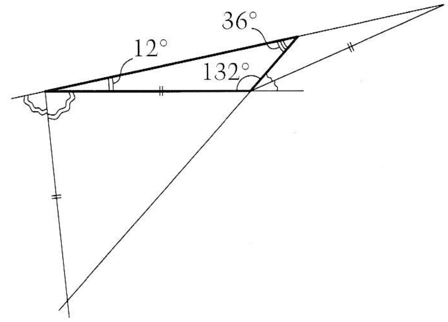


Figure 5

Here are several problems that are variations on the topic considered. We will not solve them here, but we will give the answers.

Problem 1. Bisectors AA_1 and BB_1 of a triangle ABC meet at a point I. It is given that $IA_1 = IB_1$. Does this imply the equality $CA = CB$?

No, it does not. But it does follow that either $CA = CB$ or angle C is 60°.

Problem 2. Each angle bisector of a triangle meets the opposite side at a point equidistant from two other sides. Does this imply that the triangle is equilateral?

It turns out that besides equilateral triangles, the triangle with the sides 1, 1, and $\sqrt{2} - 1$ (and all similar triangles) possess this property.

The complete answer to the following question is quite unexpected.

Problem 3. Suppose that a triangle with vertices at the feet of another triangle's bisectors is isosceles. Is the other triangle isosceles as well?

In the general case, the answer is negative. Careful analysis yields the following result. Let AA_1, BB_1, and CC_1 be the bisectors of a triangle ABC. Sides AB and AC cannot be equal if angle A is obtuse and its cosine is in the interval

$$\left(-\frac{1}{4}, \frac{\sqrt{17}-5}{4}\right),$$

which corresponds to angles in the range from about 102°40′ to 104°28′. In other cases, the triangle is isosceles. ◘

The quantum nature of light

by D. Sviridov and R. Sviridova

THE QUANTUM THEORY DEveloped in the 1920s needed experimental verification. If light is radiated in the form of quanta, we can try to see them. At first glance, this thought is just an idle fancy. What experiment could prove or disprove the concept of the quantum discreteness of light using such an imperfect device as the human eye? The problem seems insoluble. To believe that the human eye is able to distinguish quanta, one should be a specialist both in optics and the physiology of vision.

The most important step in formulating new scientific problems is to exceed the limits of the current views on certain relationships in nature. Once Werner Heisenberg said, "A naturalist is interested primarily in posing questions and only secondarily in answering them. To formulate a problem seems to him to be much more valuable if it helps to develop the human mind. In most cases the answers have only transient and evanescent value: In the course of time they can be made obsolete by the enlargement of knowledge in physics."

The man who saw quanta

Such a unique problem was put forward by the President of the Russian Academy of Sciences, Sergei Ivanovich Vavilov (1891–1951). In 1920, when Vavilov was Head of the Optical Department at the Biological Physics Research Institute of the Ministry of Health, he addressed the problem of the quantum structure of light. In his monograph "The Microstructure of Light," which generalized all his work on the nature of light, Vavilov wrote, "The properties of light can best be revealed under the limiting conditions: in studying the weak luminous fluxes formed by a small number of light quanta, in investigating the processes that go on during a billionth of a second, and in analyzing molecular interaction at extremely small distances." And so he carried out an experiment of this kind with the hope of observing individual quanta of light.

In 1729 the French scientist Pierre Bouguer (1698–1758) experimentally found the attenuation law for light traversing a medium. Attenuation occurs because all substances absorb light that passes through them. Let J_0 be the intensity of light with wavelength λ that hits a layer of substance with thickness d, and let J be the intensity of the light emerging after traveling through the layer. Using these notations, Bouguer's law says

$$\ln \frac{J}{J_0} = -kd,$$

where k is the absorption coefficient. This law laid the foundation for all subsequent measurements of light absorption in a medium. "In his field of research, Pierre Bouguer is as famous a scientist as Kepler or Newton in mechanics," wrote Vavilov.

Many experiments showed that the absorption coefficient doesn't depend on the intensity of the incident radiation. The quantum theory of light prompted attempts to find such a dependence. The absorption coefficient was found to be constant within a huge range of light intensities, which varied in the experiments by a factor of 10^{20}. However, if light really existed in discrete units, reasoned researchers, then there should be fluctuations in the intensity of the emerging beam when it becomes extremely attenuated. Furthermore, the number of quanta absorbed during short spans of time should not be constant, and Bouguer's law will be violated. Thus, researchers needed to detect these stochastic quantum oscillations.

For many years the "Quantum Hunters," as Vavilov's colleagues were nicknamed, tried to carry out such an experiment. It was not until 1941 that they succeeded.

Can the human eye detect such a microscopic event? After many years of work it was found that the eye can sense several quanta of light that reach it simultaneously (the threshold value is eight quanta per second). How could Vavilov's team

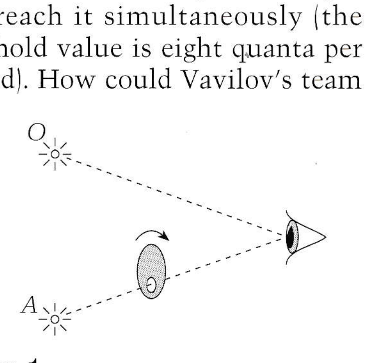
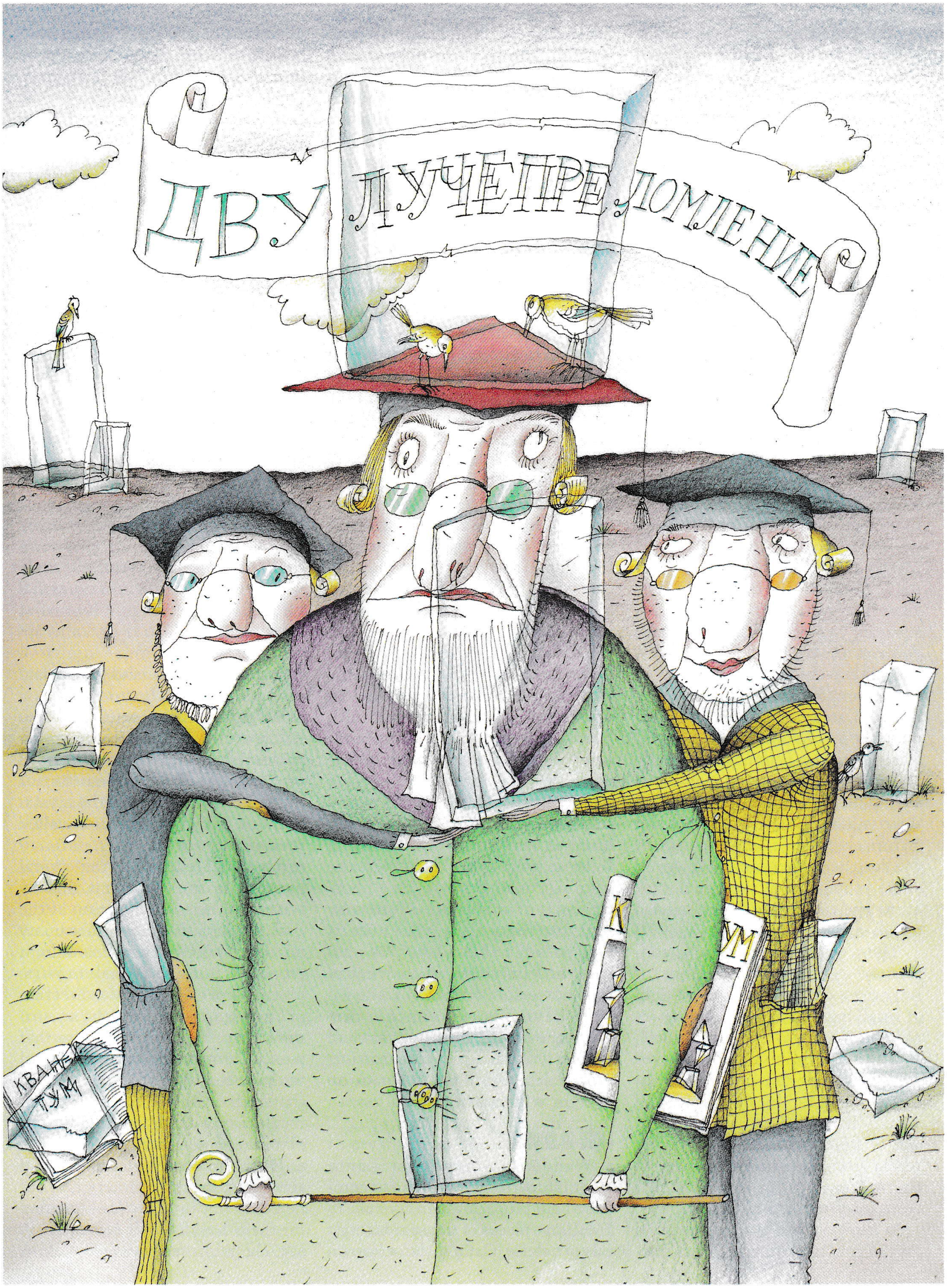


Figure 1

Art by V. Ivanyuk


ДВУЛУЧЕПРЕЛОМЛЕНИЕ
КВАНТУМ

demonstrate this?

In Vavilov's experiment a small faintly luminous spot *A*, whose brightness could be continuously attenuated, is observed by the eye (fig. 1). When the light is very weak, only a few quanta hit the eye per second. Under these conditions oscillation in the brightness of the source should appear. Therefore, an observer should see that a steady source of light begins to twinkle. The experiment seems simple, but a number of factors make it rather difficult.

First of all, one should use a source of very low intensity, because under normal conditions most of the particles of radiation produce a continuous stream of light. Second, there are both "classical" and quantum fluctuations of light. The "classical" fluctuations, caused by the motion and interaction of atoms and molecules, are related to processes inside the light source. One can make a source in which these classical fluctuations are virtually absent. For example, fluorescent molecules dissolved in a very viscous substance, which are not subjected to the damping action of the viscous medium, will radiate light continuously at a constant level. By contrast, quantum fluctuations are always observed when the medium is sufficiently rarefied. Third, these fluctuations should be reliably and significantly detected by the eye in the designed experiment. However, some properties of the eye do not allow the experiment to be carried out in such a simple mode.

The point is that our eye is constantly moving, so variations in brightness can be observed not only at small but also at large intensities of light. To overcome this obstacle, in the experiment the observer stares at some brighter point *O* (usually red in color) located beside the weakly luminous point *A*. The image of point *O* is created at the retina's center, while the image of the source at point *A* is shifted a constant distance from the center.

In addition, the eye has the property of retaining images. This may lead to a merging of the rapid intensity oscillations of the light source, which would be hidden in the averaged luminous background. To prevent this effect, a disk with an orifice is placed between the source and the observer. The disk makes one turn per second, and the eye observes the source only through the orifice (for example, during 1/10 second).

This very simple setup makes it possible to detect a very intricate phenomenon. When the number of quanta is larger than some visual threshold, the observer perceives a burst of light each time the orifice opens. If the number of quanta is decreased to the threshold value, not every opening of the orifice will cause a visible burst. By gradually decreasing the brightness of the source at point *A*, Vavilov's team was able to detect such variations in intensity. The lower the intensity of the source, the larger the number of omissions. The numbers of omissions and bursts determine the mean number of quanta in the burst. Thus it was really possible to see the quantum nature of light with the human eye. Now we have very sensitive devices, such as photomultipliers and quantum counters, but it was the human eye that first saw the quantum of light.

Along with V. Levshin, Vavilov found that Bouguer's law was violated at very large intensities of light as well. They observed the luminescence of crystalline phosphorescent materials excited by high-intensity light. However, it was possible to explain this phenomenon only after the advent of new powerful sources of light: lasers. Vavilov's experiments provided the basis for a new branch of physics, quantum electronics, which deals with the interaction of light with matter.

Light in crystals

Wonderful and at first glance even fantastic phenomena accompany light propagation in crystals. In 1669 large pieces of transparent crystals of calcite ($CaCO_3$), later known as Iceland spar, were delivered from Iceland to Denmark. By studying the optical properties of these crystals, Erasmus Bartholin, a professor at the University of Copenhagen, stumbled upon the wonderful phenomenon of birefringence, or double refraction.

Also in 1669, Bartholin's compatriot Nicolaus Steno (1638–1685) found one of the most important laws of crystallography, which exemplifies the highest harmony in nature—the principle of angular invariance. In his 1669 treatise "Concerning Solids Naturally Contained within Solids," Steno wrote, "In the face of a crystal, the number of sides and their lengths may vary without changing its angles."

Bartholin tested this law by studying the Iceland spar crystals. Outlining the faces of crystals, he compared various drawings. Once he put a crystal on a sketch and saw that the sketch appeared twice. When he took the crystal off, there was only one drawing. His notes and virtually everything he looked at through the strange crystal also appeared in tandem.

If a crystal of Iceland spar is laid on a piece of cardboard that has a small orifice illuminated from below, we find that the ray of light that passes through this orifice is divided into two rays. One of them travels normal to the surface of the crystal without refraction, and it is called the *ordinary ray*. The *extraordinary ray* is deflected within the crystal but after passing through it, this ray travels in the same direction as the first. Properties of the extraordinary ray depend on the direction of light propagation in the crystal. In physics this dependence on the direction of propagation is called "anisotropy." Investigation of the rays that passed through Iceland spar crystal with the help of a polarizer showed that both rays are completely polarized in mutually perpendicular planes.

Bartholin determined the refractive index for the ordinary ray, but he could not find any laws regulating the behavior of the extraordinary ray. He published the results of his studies in Leipzig, Copenhagen, and London. However, Bartholin's discovery was not accepted by the scientific community. The Royal Society of London set up a committee to check Bartholin's work. Although the committee included such celebrities as Newton, Hooke, and Boyle, it labeled the discovery misleading and the corresponding laws false.

The works of Bartholin were forgotten, and it wasn't until 20 years later, in 1691, that the famous Dutch physicist and mathematician Christiaan Huygens (1629–1695) confirmed the correctness of Bartholin's discoveries and observed double refraction in quartz crystals. In his "Treatise on Light," he explained the phenomenon of double refraction in Iceland spar on the basis of his wave theory of light.

In 1801 the French crystallographer and mineralogist René Just Hauey (1743—1822) supplied in his "Course of Mineralogy" a list of birefringent crystals. He determined the property of double refraction by looking at a thin needle through the faces of crystals. When birefringence was pronounced, the needle was doubled. Hauey was the first to group crystals according to whether they were single or double refracting. He pointed out that single-refracting crystals were substances whose "molecules are characterized by a high degree of symmetry," which means crystals in the form of a cube, octahedron, and so on.

Newton tried to attribute the wonderful property of double refraction in Iceland spar to the particular arrangement of particles in this crystal. He wrote, "The corpuscles of Iceland crystal act on the rays all in the same direction, thereby producing paradoxical refraction. Therefore, can it not be supposed that during formation of this crystal, its constituting corpuscles not only arranged themselves in lines and rows but also turned their identical sides in the same direction due to some polar ability?"

Huygens also related birefringence to the regular structure of the crystals: "It seems that the regularity found in these natural masterpieces, the crystals, resulted from the arrangement of the smallest invisible and identical constituent corpuscles. Iceland spar consists of small round bodies that are spheroid but not spherical due to some flatness."

The theory of light propagation in crystals was elaborated by the French physicist Augustin-Jean Fresnel (1788–1827). He showed that in general two waves traverse the crystals, which are polarized in the mutually perpendicular planes. He proposed crystal classification based on the type of optical surfaces and considered the problems of elliptic and circular polarization and rotation of the polarization plane. He also pointed out the possibility of the existence of conical refraction and found the quantitative laws of refraction and reflection of light, which make it possible to determine the intensity and polarization of light after reflection and refraction.

What is the modern view on the nature of double refraction? Electromagnetic waves (light) shift the electron shells relative to the atomic nuclei in the crystal. In addition, the ions are displaced relative to each other in the ionic lattices. However, this shift occurs only at low frequencies (in the infrared region of the spectrum), because the ions are rather heavy and cannot follow the high-frequency fields. This displacement of charged particles in crystals is called electric polarization. In turn, electric polarization generates an electromagnetic field that interferes with the field of the original wave. If the polarization of a crystal depends on the direction of the electric field of the wave, it leads to anisotropy of the dielectric permittivity and of refractive index. This anisotropy of the refractive index causes double refraction.

Let's consider double refraction in crystals composed of elongated nonspherical molecules that are longer than they are wide. Suppose these molecules are arranged such that their major axes are parallel. If an electromagnetic wave travels through such a crystal, its molecular structure favors the oscillation of electrons along the molecular axis and not in the transverse direction. The electric field directed along the axis of the molecules will produce one effect, and the electric field directed normally to the molecular axis would produce quite another effect. Thus, these two waves travel with different velocities and have different refractive indexes. Therefore, the phenomenon of double refraction arises. ◙

Quantum articles about the discreteness of light and double refraction:

"The nature of light," A. Eisenkraft and L. D. Kirkpatrick, November/December, 1996, pp. 30–31.

"A polarizer in the shadows," A. Andreev, January/February, 1994, pp. 44–48.

Warp speed

Because we think, we think the universe is about us.
But does it think, the universe?
Then what about?
About us?

—May Swanson

by Larry D. Kirkpatrick and Arthur Eisenkraft

"TAKE US TO MAXIMUM warp," Captain Jean-Luc Picard orders, and the starship *Enterprise* begins to travel faster than the speed of light to avoid trouble. Warp 9.6 is the highest normal rated speed for the *Enterprise* and corresponds to a speed 1909 times the speed of light. After reaching safe haven, Captain Picard uses a subspace signal to set up a videoconference with Earth, even though the *Enterprise* is thousands of light-years from Earth.

Although such faster-than-light travel is commonplace in science fiction such as *Star Trek*, ordinary matter in our ordinary world must obey the laws of physics. The speed of light is the speed limit in the Universe. Only massless particles such as photons can travel at the speed of light; massive particles—such as those making up the starship *Enterprise*—must sluggishly travel at slower speeds.

This makes any observation of something appearing to travel faster than the speed of light rather astonishing. Such observations have been made in astronomy and will be the focus of our contest problem.

But first we make a digression to talk about high-speed photography. Let's assume that we have a camera with a very fast shutter speed, say, a very small fraction of a nanosecond. (Such cameras are only available in physics stores along with massless pulleys and frictionless surfaces.) We want to take a photograph of a thin rod as it passes by at a relativistic speed.

Let's assume that the camera is located at the origin and is pointing along the $+y$-direction. The thin rod is 3 m long with its ends at $y = 3$ m and $y = 6$ m as shown in figure 1.

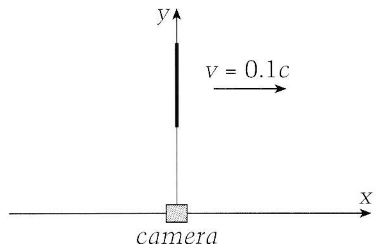


Figure 1

The rod is traveling in the $+x$-direction at a speed of $0.1c$. When the rod passes the x-axis, we consider light reflected from each end. The light reflected from the close end takes 10 ns to reach the camera and expose the film. However, the light reflected from the far end takes 20 ns to reach the camera and bumps into the closed shutter. The light from the far end that enters the camera must have been reflected earlier to allow for the extra distance it has to travel. That is, the light from the far end that arrives at the camera simultaneously with the light reflected from the near end must have been emitted a little more than10 ns earlier.

This means that the camera records the near end as being at $x = 0$, but the far end as being located at $x = -0.3$ m. The camera does not "see" the rod lying along the x-axis, but rotated through an angle of 5.7°. Who says cameras don't lie?

Our contest problem is based on one of the three theoretical problems used in the International Physics Olympiad held in Reykjavik, Iceland, on 2–10 July 1998. One of us (LDK) had the privilege of working with the Icelandic exam committee and found the problems to be very interesting and challenging.

A. In 1994, GRS1915+105 was observed to emit ejecta in opposite directions. As reported by I. F. Mirabel and L. F. Rodriguez in *Nature* (vol. 371, p. 46), the ejecta were probably produced by a neutron star or a black hole similar to the process occurring in quasars, but on a smaller scale. They call the object a microquasar.

Art by Tomas Bunk


TOM BUNK - OCT '98

Fits to the observations over a period of 34 days showed that the ejecta left the microquasar with angular speeds of ω_1 = 17.6 milliarcseconds/day and ω_2 = 9.0 milliarcseconds/day. If the microquasar is located at a distance $R = 3.86 \cdot 10^{20}$ m from Earth, what are the components of the velocities of the two ejecta perpendicular to the line of sight (the transverse velocities)?

B. You are probably surprised to discover that one of these velocities has a component larger than the speed of light. To see how this

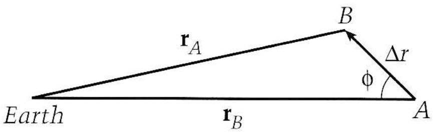


Figure 2

arises, let's do the following calculation. Assume that an object is traveling at a speed $v = \beta c$ at an angle $\phi < 90°$ relative to the line of sight as shown in figure 2. Denote its original position by $\mathbf{r_A}$ and its final position by $\mathbf{r_B}$. In a time interval Δt the object travels a distance Δr. What is the time interval Δt_0 between the arrival of the signal from position A to the arrival of the signal from position B as observed on Earth?

C. What is the transverse velocity observed for this motion in terms of β, R, and ϕ?

D. What is the minimum value of β for which we can observe a transverse velocity greater than the speed of light for some angle ϕ? What angle corresponds to this minimum β?

E. Draw a graph of β versus ϕ showing the region where we can observe apparent superluminal transverse velocities.

Please send your solutions to *Quantum*, 1840 Wilson Boulevard, Arlington, VA 22201-3000 within a month of receipt of this issue. The best solutions will be noted in this space.

Depth of knowledge

What a wonderful showing for our Depth of Knowledge contest problem in the May/June 1998 issue. Correct solutions were submitted by Zach Frazier, a senior at Ferris High School in Spokane, Wash.; professor F. Y. Wu of Northeastern University, Mass.; John Parmon of Narbeth, Penn.; and Scott Wiley, a physics teacher from Westcaco, Tex. The first part of the problem required readers to calculate the depth of a well if a dropped stone is heard hitting the bottom after 3 s.

The total time to hear the sound is the sum of the time for the stone to fall plus the time for the sound to rise.

$$T = t_1 + t_2,$$

where

$$t_1 = \sqrt{\frac{2h}{g}}$$

and

$$t_2 = \frac{h}{v}.$$

Knowing the total time T = 3 s, the acceleration due to gravity g = 9.8 m/s^2, and the speed of sound v = 340 m/s, we can solve for the height of the well h by using the solver or the poly function on a TI-85 calculator or the quadratic equation. The depth of the well that satisfies these equations is 40.65 m.

The second part of the problem assumes that there is air resistance where the resistive force is proportional to the velocity of the stone:

$$mg - kv = m\frac{dv}{dt},$$

$$\int \frac{dt}{m} = \int \frac{dv}{mg - kv}.$$

Let $u = mg - kv$. We have

$$\int \frac{-k}{m} dt = \int \frac{du}{u},$$

$$\frac{-k}{m}t = \ln \frac{u}{u_0},$$

$$e^{\frac{-kt}{m}} = \frac{mg - kv}{mg},$$

$$v = \frac{mg}{k}\left(1 - e^{\frac{-kt}{m}}\right).$$

To determine the depth of the well, with air resistance, we must derive an equation for the distance traveled by the stone:

$$v = \frac{dx}{dt} = \frac{mg}{k}\left(1 - e^{\frac{-kt}{m}}\right),$$

$$\int dx = \int \frac{mg}{k}\left(1 - e^{\frac{-kt}{m}}\right)dt,$$

$$x = \frac{mgt}{k} - \frac{m^2 g}{k^2} e^{\frac{-kt}{m}} - C.$$

Since $x = 0$ when $t = 0$,

$$C = \frac{m^2 g}{k^2}$$

and

$$x(t) = \frac{mgt}{k} - \frac{m^2 g}{k^2} e^{\frac{-kt}{m}} - \frac{m^2 g}{k^2}.$$

Using the given information that k = 0.01 kg/s, m = 0.05 kg, and that the total time T to hear the splash is still 3.0 s, we can now determine the depth of the well.

The total time is the time for the stone to fall t_f and the time or the sound to rise:

$$T = t_f + \frac{x(t_f)}{v}.$$

Readers solved the equation graphically, numerically, with *Mathematica*, and with a spreadsheet. The depth of the well is now 34.26 m.

Completing the square

by Mark Saul and Titu Andreescu

LET'S TALK FIRST ABOUT SOLVING QUADRATIC equations. Here is a pedagogical trap that often catches beginning teachers. We can learn something from the error.

Problem: Find two numbers whose sum is 13 and whose product is 30. Only an algebraic solution will be accepted.

Many people can see right away that the answer is 10 and 3. But our hapless beginner will insist on an algebraic solution and start like this: Let x and y be the required numbers. Then $x + y = 13$ and $xy = 30$. From the first equation, $y = 13 - x$, and substituting this value into the second equation, we find that $x(13 - x) = 30$, which simplifies to $x^2 - 13x + 30 = 0$. We must now solve this quadratic equation.

We choose (says our unlucky novice) the technique of factoring. How do we factor such a trinomial? We need to represent it as $(x - \alpha)(x - \beta)$; that is, we need to find two numbers α and β whose sum is 13 and whose product is 30. That is, we have come back to our original problem!

We first learn to solve quadratic equations by factoring, and we often delight in our mastery of the factorization of trinomials. But factoring is just guessing, and it is of no use if the roots are irrational, complex, or even very large integers. This is why two more standard techniques have been developed for solving quadratic equations: completing the square and using the formula.

Completing the square is a powerful technique, which can be used to solve any quadratic equation with real coefficients. It is algorithmic: We always know what to do next, and the method always works after a finite number of steps. Indeed, it can even be encapsulated in a formula (the famous "quadratic formula") that even a computer can understand.

The algorithm also generalizes in several directions. The following problems are typical.

Problem 1. If x is a real number, find the smallest possible value of $x^2 - 8x + 21$.

Solution. Let's ignore the constant term and complete the square. We must add 16 to $x^2 - 8x$. So we write

$$x^2 - 8x + 21 = x^2 - 8x + 16 + 5 = (x - 4)^2 + 5.$$

Since the square of a real number cannot be negative, the minimal value of the expression is 5, and it is achieved when $x = 4$.

Problem 2. If x and y are real numbers, find the smallest possible value of

$$x^2 + y^2 - 8x + 6y + 17.$$

Note that a solution using calculus is beyond the usual first-year course in this subject.

Problem 3. Find the center of the ellipse whose equation is

$$4x^2 + 6y^2 - 24x + 12y = 20.$$

Problem 4. (This problem is a bit ahead of its time!) Find the smallest real number r such that

$$x^2 + y^2 + 19x + 99y + r \geq 0$$

for all real numbers x, y.

Problem 5. Are there one-to-one functions f: $\mathbf{R} \to \mathbf{R}$ such that $f(x^2) - f^2(x) \geq 1/4$ for all (real) x? (Here we write $f^2(x)$ for $[f(x)]^2$.)

Problem 6. Let a, b, c be three real numbers. Prove that at least one of the numbers $a - b^2$, $b - c^2$, $c - a^2$ does not exceed 1/4.

Sometimes we find that the squares in our problem are already complete, and we just have to recognize them.

Problem 7. Find all triples (x, y, z) of real numbers such that

$$x^2 = 4(y - 1)$$
$$y^2 = 4(z - 1)$$
$$z^2 = 4(x - 1)$$

Solution. Adding and simplifying, we find that

$$x^2 - 4x + 4 + y^2 - 4y + 4 + z^2 - 4z + 4 = 0.$$

If we try to complete the square, we quickly find that they are already complete(!), and we have

$$(x - 2)^2 + (y - 2)^2 + (z - 2)^2 = 0.$$

Thus the only possible solution is (2, 2, 2).

Sometimes we can "complete the square" by adding and subtracting a middle term, rather than the constant term, of an expression.

Problem 8. Factor $x^4 + 4$ over the integers.

Solution. At first glance this may seem impossible: The given polynomial is the sum of two squares, which usually cannot be factored. But the presence of a fourth power in the first term lets us "complete the square" thus:

$$x^4 + 4 = x^4 + 4x^2 + 4 - 4x^2 = (x^2 + 2)^2 - (2x)^2.$$

Now the expression factors as the difference of two squares. Answer:

$$(x^2 + 2x + 2)(x^2 - 2x + 2).$$

The reader is invited to check that these factors multiply to the given expression.

Problem 9. Factor $4a^4 + b^4$ (over the integers).

Problem 10. Show that the number $4^n + n^4$ is composite for all integers $n > 1$.

Problem 11. For any two real numbers a and b, show that

$$a^2 + b^2 \geq (1/2)(a + b)^2.$$

Problem 12. Show that

$$a^2 + b^2 + c^2 \geq ab + ac + bc.$$

Solution. It would be nice if our inequality involved $a^2 + b^2 + c^2$ and also $2ab + 2bc + 2ac$. Then we could proceed as in problem 11. So let's try multiplying the whole inequality by 2. We get

$$2a^2 + 2b^2 + 2c^2 - 2ab - 2ac - 2bc \geq 0,$$

which groups as

$$(a - b)^2 + (b - c)^2 + (c - a)^2 \geq 0,$$

which is certainly true.

Problem 13. Show that

$$a^2 + b^2 + c^2 - ab - bc - ca \geq (3/4)(a - b)^2.$$

Problem 14. Show that

$$a^2 + b^2 + c^2 \geq (1/3)(a + b + c)^2.$$

Problem 15. For any n real numbers $a_1, a_2, ..., a_n$, show that

$$a_1^2 + a_2^2 + \ldots + a_n^2 \geq \frac{1}{n}(a_1 + a_2 + \ldots + a_n)^2.$$

Problem 16. Find the minimum value of

$$x^4 + y^4 + z^4 - 4xyz,$$

where x, y, z are real numbers. Where is this minimum achieved?

Solution I. The arithmetic-geometric mean inequality for four variables tells us that

$$a + b + c + d \geq 4\sqrt[4]{abcd}.$$

Letting $a = x^4$, $b = y^4$, $c = z^4$, $d = 1$, we have

$$x^4 + y^4 + z^4 + 1 \geq 4xyz,$$

or

$$x^4 + y^4 + z^4 - 4xyz \geq -1,$$

with equality when $x^4 = y^4 = z^4 = 1$ and $xyz \geq 0$, which implies $(x, y, z) = (1, 1, 1), (1, -1, -1), (-1, 1, -1)$, or $(-1, -1, 1)$.

Solution II. Can you get the same result by completing the square?

Problem 17. Let $a_1, a_2, ..., a_n$ be real numbers. Prove that

$$\frac{1}{2}(a_1 + a_2 + \ldots + a_n) \geq \sqrt{a_1 - 1^2} + 2\sqrt{a_2 - 2^2} + \ldots + n\sqrt{a_n - n^2}.$$

ANSWERS, HINTS & SOLUTIONS ON PAGE 49

Factoring

Compliments of Richard Askey of the University of Wisconsin–Madison, here are alternative solutions to problems 15 and 16 in "Symmetry in Algebra, Part III," page 42 in the July/August 1998 issue.

Problem 15. Let m and n be two odd integers. Show that

$$\frac{1}{a^m + b^m + c^m} = \frac{1}{a^m} + \frac{1}{b^m} + \frac{1}{c^m}$$

if and only if this equality holds when m is replaced by n.

Proof. Clearing off fractions, this becomes

$$c^{2m}(a^m + b^m) + c^m(a^m + b^m)^2 + a^m b^m(a^m + b^m) = 0$$

or

$$(c^m + a^m)(c^m + b^m)(a^m + b^m) = 0.$$

Thus, it holds if and only if two of the numbers a, b, c are negatives of each other.

Problem 16. Factor

$$x^3 + y^3 + z^3 - 3xyz.$$

If

$$f(x, y, z) = x^3 + y^3 + z^3 - 3xyz,$$

then $f(x, x, x) = 0$. This suggests "perturbing" the value x by considering

$$f(x, x+r, x+s) = x^3 + x^3 + 3x^2r + 3xr^2 + r^3$$
$$+ x^3 + 3x^2s + 3xs^2 + s^3 - 3x\left(x^2 + rx + sx + rs\right)$$
$$= 3x\left(r^2 + s^2 - rs\right) + r^3 + s^3 = (3x + r + s)\left(r^2 + s^2 - rs\right)$$
$$= (x + y + z)\left(x^2 + y^2 + z^2 - xy - xz - yz\right).$$

This can be factored into linear factors by using the quadratic formula. The result is

$$f(x, y, z) = (x + y + z)(x + wy + w^2z)(x + w^2y + wz),$$

where $w^3 = 1$ and $w \neq 1$. A general cubic equation can be written as

$$t^3 + 3bt^2 + ct + d = 0.$$

If $x = t + b$, then x satisfies

$$x^3 + px + q = 0$$

for constants p and q. If

$$p = -3yz,$$
$$q = y^3 + z^3,$$

then y can be obtained by solving a quadratic equation in y^3, so the factorization of $f(x, y, z)$ into linear factors gives the solution of a general cubic equation.

AT THE BLACKBOARD II

Errors in geometrical proofs

by S. L. Tabachnikov

MANY STUDENTS COME across a "proof" that $1 = -1$. Here is an example of such a proof.

Let $a + b = c$ and $a = b = 1$. Multiply both sides of the equality $a + b = c$ by $a + b$ to obtain $a^2 + 2ab + b^2 = c(a + b)$. We can rearrange this equation to obtain

$$a^2 + ab - ac = -ab - b^2 + bc$$

or

$$a(a + b - c) = -b(a + b - c).$$

Dividing both sides by $a + b - c$, we obtain the absurd equality $a = -b$, or $1 = -1$.

After a moment of embarrassment, we see the error in this derivation at once: We cannot divide by $a + b - c$ because this expression equals zero. Many other similar proofs employ the same idea of disguised division by zero, and it is easy to find the error in these proofs.

Erroneous proofs of false propositions in geometry are less well known. The search for errors in geometrical proofs is often a difficult yet instructive task. The eminent book by Y. S. Dubnov *Errors in Geometrical Proofs* (Moscow: Nauka, 1969) is dedicated to such problems.[1] This book, which has been out of print for many years, contains 15 geometrical "proofs" with complete analysis. In this article, we discuss one of them. We invite you to follow the reasoning carefully and try to find the error.

"Theorem": All triangles are isosceles.

Proof. Consider a triangle ABC

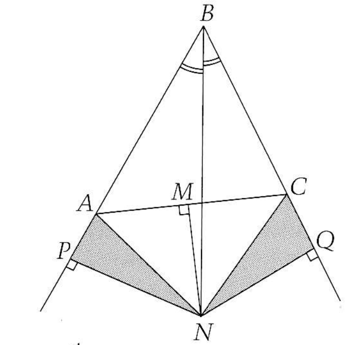


Figure 1

(fig. 1). Draw the bisector of angle B and the perpendicular bisector to the base AC. In what follows, these segments are simply called the *bisector* and *perpendicular bisector*. Call the point of their intersection N. Drop the perpendiculars NP and NQ onto the lines AB and BC.

Since N lies on the bisector, it is equidistant from the lines AB and BC. Therefore, $|PN| = |QN|$. Since N lies on the perpendicular bisector, it is equidistant from points A and C. Thus, $|AN| = |CN|$.

The right triangles ANP and CNQ are congruent by hypotenuse-leg. Thus, $\angle NAP = \angle NCQ$. In addition, triangle ANC is isosceles. Therefore, $\angle NAM = \angle NCM$. Adding equal angles, we conclude that $\angle PAM = \angle QCM$. Thus, $\angle BAM = \angle BCM$, and triangle ABC is isosceles.

The proof is finished. However, you may have some objections. The following are some likely objections and attempts to answer them.

Objection 1. How can we be sure that the bisector and perpendicular bisector meet? They might be parallel.

Answer. If the bisector is parallel to the perpendicular bisector, it must be perpendicular to the base of the triangle. Therefore, it is also the altitude of triangle ABC. Thus, this triangle is isosceles. As we can see, the conclusion that the triangle is isosceles retains its validity!

Objection 2. The bisector can coincide with the perpendicular bisector.

Answer. I have two answers to this objection. First, in this case, the bisector also coincides with the altitude, and the triangle is, again, isosceles. Second, if the bisector coincides with the perpendicular bisector, the point N can be chosen arbitrarily on the bisector. Then the proof can proceed as before.

Objection 3. What if the bisector meets the perpendicular bisector on the base AC—that is, if $N = M$?

Answer. In this case, the bisector coincides with the median BM.

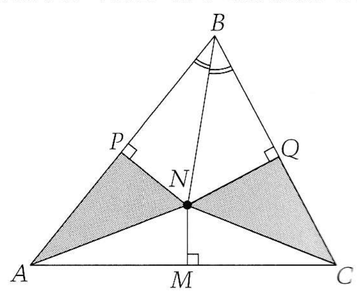


Figure 2

[1]Yakov Semenovich Dubnov (1887–1957) was a well-known geometer and teacher who taught at Moscow University for many years. He also gave lectures for high school students that provided the basis for this book.

This, again, implies that triangle ABC is isosceles.

Objection 4. We have not considered the case when the bisector meets the perpendicular bisector inside triangle ABC.

Answer. This case is illustrated in figure 2. Reasoning as in the basic proof (fig. 1), we conclude that $|AN| = |NC|$ and $|NP| = |NQ|$. Therefore, $\triangle APN = \triangle CQN$. Thus, $\angle NAP = \angle NCQ$. In addition, triangle ANC is isosceles, and therefore $\angle NAM = \angle NCM$. Summing up the equal angles, we find that $\angle BAC = \angle BCA$. Thus, triangle ABC is isosceles.

However, I can present another counterargument: The case shown in figure 2 is impossible because an angle bisector in any triangle always lies between the corresponding median and altitude.

Thus, these four objections are invalid. I invite the reader to stop here and try to find a decisive objection against the "proof" given. If you need help, see Answers, Hints & Solutions on page 49.

Sketchy proofs?

Now you will doubtless consider drawings with more suspicion. Some readers might even want to go further and write rigorous mathematical proofs that eliminate the need for drawings, thus excluding the possibility of making an error. Professional mathematicians usually satisfy themselves with the potential possibility of making their proofs formal. However, they often use various graphical illustrations. This can be explained by the fact that formal derivations are much longer than common proofs, and they are much more difficult to understand and devise.[2] Thus, it is unreasonable to discard graphical illustrations when proving geometric

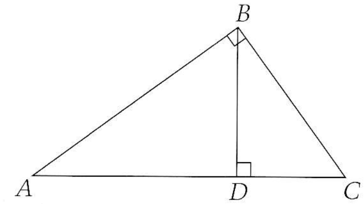


Figure 3

theorems. We must, however, know how to use them correctly.

We invite the reader to find errors in the following "proofs" from Dubnov's book.

Problem 1. As is well known, the sum of the angles of any triangle equals 180°. The proof of this fact is based on the parallel axiom. The following proof does not use this axiom.

Decompose an arbitrary triangle ABC into two triangles (fig. 3). Let x be the yet unknown sum of the angles of the triangle. Then, $\angle 1 + \angle 2 + \angle 6 = x$ and $\angle 3 + \angle 4 + \angle 5 = x$. Combine these two equations to obtain $\angle 1 + \angle 2 + \angle 3 + \angle 4 + \angle 5 + \angle 6 = 2x$. Since $\angle 5 + \angle 6 = 180°$ and $\angle 1 + \angle 2 + \angle 3 + \angle 4$ is the sum of the angles of ABC (that is, x), we obtain the equation $x + 180° = 2x$. Therefore, $x = 180°$.

Problem 2. Let us prove the "theorem" that states that a rectangle inscribed in a square is also a square.

Consider a square $ABCD$, and let $KLMN$ be an inscribed rectangle (fig. 4). Drop the perpendiculars $KP \perp AD$ and $NQ \perp CD$. These perpendiculars are equal to the sides of the square, and thus are equal to each other. Segments KM and NL are also equal since they are the diagonals of rectangle $KLMN$. Therefore, the right triangles KPM and NQL are congruent by hypotenuse-leg. Consequently, $\angle KMP = \angle NLQ$. Consider

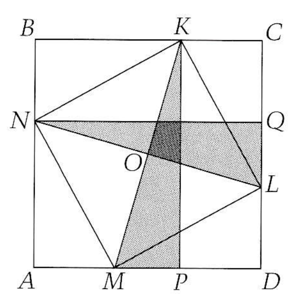


Figure 4

the quadrilateral $OLDM$. Since $\angle OMP = \angle OLQ$, the sum of the angles OLD and OMD equals 180°. Thus, the sum of the other two opposite angles of the quadrilateral $OLDM$ also equals 180°—that is, $\angle MOL + \angle MDL = 180°$. However, $\angle MDL = 90°$, and thus, $\angle MOL$ is a right angle. This implies that $KLMN$ is a rectangle with perpendicular diagonals, which must be a square.

Problem 3. Let us prove the "theorem" that states that a perpendicular and a slanting line to the same line do not meet.

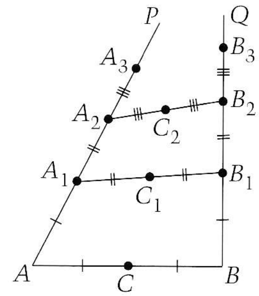


Figure 5

Consider a slanting line AP and a perpendicular BQ to the segment AB (fig. 5). Let C be the center of AB. Lay off the segments AA_1 and BB_1 equal to $AC = BC$. We claim that the rays AP and BQ do not meet at a point interior to segments AA_1 and BB_1. Indeed, if such an intersection point existed, the following inequalities would be satisfied for triangle AKB: $|AK| \le |AA_1|$ and $|BK| \le |BB_1|$, which implies

$$|AK| + |KB| \le |AA_1| + |BB_1| = |AB|.$$

However, the last inequality contradicts the triangle inequality.

Now, connect points A_1 and B_1 and repeat the preceding construction to obtain the points A_2 and B_2. Here $|A_1A_2| = |A_1C_1| = |C_1B_1| = |B_1B_2|$. As before, the rays AP and BQ do not meet at an interior point of segments A_1A_2 and B_1B_2. In particular, points A_2 and B_2 are different, and

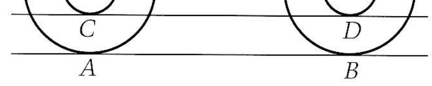


Figure 6

[2]For example, it is rather difficult to write the number 1 in one of the formal languages—the language of the set theory by N. Bourbaki. "The full notation ... would have contained several tens of thousands of symbols—rather a lot for the number 1," said Y. I. Manin on the Soviet radio program "Provable and Unprovable" in 1979.

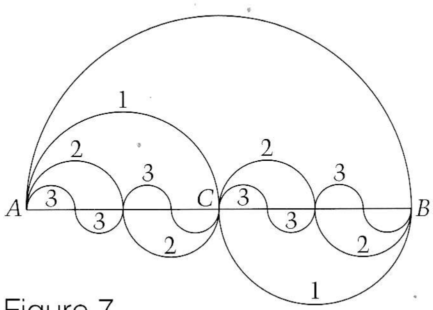


Figure 7

the construction process can be continued infinitely. Therefore, the rays AP and BQ do not meet.

Problem 4. Let us prove that all circles have equal circumferences.

The larger circle in figure 6 makes a full revolution moving from point A to point B. Thus, the distance AB equals the circumference of the larger circle. The small circle inside the larger one also makes a full revolution moving from point C to point D. Therefore, $|CD|$ equals the circumference of the smaller circle. Since the lengths of the segments AB and CD are clearly equal, both circles have equal circumferences.

Problem 5. It is well known that the circumference of the circle of radius R equals $2\pi R$. However, we will "prove" that the circumference of a circle is twice its radius.

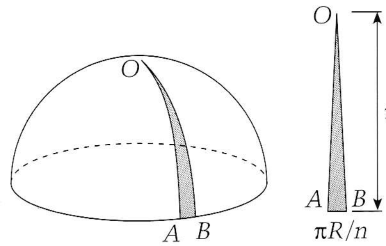


Figure 8

Draw a semicircle with diameter AB (fig. 7). Divide the segment AB in half by the point C and construct semicircles on segments AC and CB, placing them on different sides of the line AB. Each of these semicircles has a diameter equal to half the diameter of the initial semicircle. Thus, their circumference is half that of the initial semicircle. Therefore, the length of the wavelike curve labeled "1" in figure 7 is equal to the circumference of the initial semicircle.

Now, divide each of the segments AC and CB in half and construct the wavelike curve labeled "2." Again, its length equals the length of the initial semicircle. Continuing this process, we obtain a wavelike curve of the same length at each step. The distance of the points of this curve from line AB does not exceed the radius of the semicircles that constitute the curve. Therefore, it tends to zero. Consequently, the sequence of wavelike curves tends to the segment AB. Since the lengths of these curves are equal to the circumference of the initial semicircle, all these lengths must be equal to the length of AB. Thus, the circumference of the circle is twice its radius.

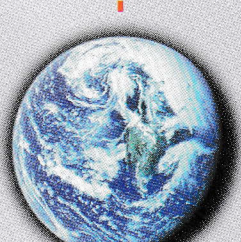
Problem 6. Let us prove that the area of a sphere of radius R is π^2R^2.

Consider the hemisphere with the pole O and divide its equator into n equal parts. The area of the hemisphere is n times larger than the area of each of the small spherical triangles shown in figure 8. Consider one such triangle. Its base equals $2\pi R/n$, and its altitude tends to $\pi R/2$ as $n \to \infty$. Therefore, its area tends to $\pi^2R^2/2n$. Thus, the area of the hemisphere is equal to

$$n\frac{\pi^2R^2}{2n} = \frac{\pi^2R^2}{2},$$

and the area of the sphere is π^2R^2. ◘

ANSWERS, HINTS & SOLUTIONS ON PAGE 49



Why is a burnt match bent?

by V. Mil'man

PLAYING WITH MATCHES can be dangerous. Nevertheless, let's do some simple and instructive experiments with burning matches, taking every precaution. In these tests we'll observe how a wooden match changes shape as it burns.

First we must take safety measures in addition to the standard ones such as tying back long hair, securing loose clothing, and so on: Hold the matches with tweezers over a basin of water set on a metal sheet. Once everything is prepared, we are ready for the experiments.

Test 1. Hold a burning match horizontally. The flame moves along the match, and as it does the charred portion of the match bends upward. Different matches display different degrees of bending. Some of them become twisted after burning. It is noteworthy that it is the cooled (charred) part of a match that bends.

Test 2. Hold a burning match in the flame of a burner (for instance, a gas stove). The charred part barely bends at all.

Test 3. Let's observe the burning of matches of different thickness. Thick matches become more twisted than thin ones, and the splinters of a match twist the least.

Look attentively once more at a burning match held horizontally. The flame slowly travels along it. The wood in the flame doesn't bend. This part of the match is black, which means that its temperature is not very high—no more than 500–600°C. Immediately behind the flame area there is a narrow ($\cong$ 2 mm) red belt. This is a zone of maximum temperature ($\cong$ 700–750°C) where burning is just finished. If we look at this region in profile, we see that its upper part is red-hot, but the lower part is black. Therefore, the upper part begins to cool from a higher temperature, so it is red for a longer period. The cause of this uneven burning is the convective flow in the surrounding air.

Could it be that the temperature difference between the upper and lower parts of a match is the reason for the bending? Note that a match bends such that its convexity is always directed toward the lower temperature in the burning process. This hypothesis is corroborated by the following experiments.

Blow carefully on a burning match from above but do not extinguish the flame. Inspect the charred piece—it is almost straight. When we blow the flame downward, the temperatures of the upper and lower parts of the match are made equal, and the match doesn't bend. By contrast, if the burning match touches a cool metal object (say, a nail) from below (do not extinguish the flame with it!) the match bends more than it normally would.

Thus, our hypothesis about the cause of the bending of the burning match seems to be qualitatively supported by experiments. Now let's make a quantitative estimation using the following model.

We divide a match theoretically into two horizontal parts. During burning, the temperature of the upper part is higher than that of the lower part, but the lengths of the parts are equal. As it cools, the upper (and more heated) part contracts more than the lower part, because the temperature difference between the air and the upper part of the match is larger. Thus, the length of the upper part of the cooled match will be less than that of the lower part, and the match will bend such that the convexity is directed downward, or to the side that is colder during burning.

This model is similar to a bimetallic plate, which is composed of two metals with different coefficients of thermal expansion. Based on this model, let's evaluate the temperature difference between the upper and lower parts in the region of the red-hot belt.

Let l_0 be the match's length and d be its thickness. As an approximation we assume that these values remain nearly constant during burning. However, the lengths of the upper and lower parts of the burnt and bent match are not equal. The difference between the lengths of the parts is

$$l_2 - l_1 = l_0 \alpha \Delta T, \qquad (1)$$

where α is the coefficient of thermal expansion of the wood and ΔT is the difference in the maximum temperatures of the upper and lower parts of the match. Since

$$l_1 = \beta\left(R - \frac{d}{4}\right)$$

Art by Pavel Chernusky

and

$$l_2 = \beta\left(R + \frac{d}{4}\right),$$

we have (fig. 1)

$$l_2 - l_1 = \frac{\beta d}{2} = \frac{l_0}{R}\frac{d}{2}. \qquad (2)$$

Comparing equations (1) and (2) yields a formula for R:

$$R = \frac{d}{2\alpha\Delta T}. \qquad (3)$$

It is not easy to measure R, but we can express it via the easily measurable parameter h (fig. 1):

$$h = R(1 - \cos\beta) = R\left(1 - \cos\frac{l_0}{R}\right).$$

To a close approximation we can consider

$$1 - \cos\beta = \frac{\beta^2}{2}.$$

(Note: The error in this approximation is no more than 1% for $\beta < 38°$ and 10% for $\beta < 60°$.) Therefore,

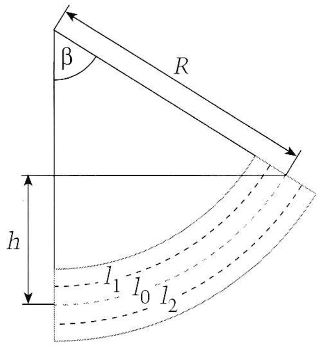


Figure 1

$$h = \frac{l_0^2}{2R}.$$

Plugging these values into equation (3) yields

$$h = \frac{l_0^2 \alpha\Delta T}{d},$$

from which we obtain

$$\Delta T = \frac{h}{l_0^2}\frac{d}{\alpha}. \qquad (4)$$

Thus, to evaluate the value ΔT, we should experimentally find the ratio h/l_0^2. The results of measurements show that to a close approximation we can consider

$$\frac{h}{l_0^2} = \text{const} \cong 10^{-2}\ \text{mm}^{-1}.$$

Assuming $\alpha = (5\text{–}10) \cdot 10^{-5}\ \text{K}^{-1}$ and $d = 1$ mm, we find according to equation (4) that $\Delta T = (100\text{–}200)°\text{C}$.

Measurements of ΔT made using a thermocouple yielded the following data:

$$T_{\text{upper}} = (730 \pm 10)°\text{C},$$
$$T_{\text{lower}} = (650 \pm 10)°\text{C},$$

and

$$\Delta T = (80 \pm 20)°\text{C}.$$

We see that our theoretical estimate agrees with these data rather well. Therefore, we can consider that although this model doesn't take into account the chemical nature of the burning process, it is still correct.

It is quite possible that by observing a burning match you will advance the understanding of why it twists. In this case try to substantiate your hypothesis with experimental data. But remember—while carrying out your experiments, *be careful and take the appropriate safety measures*!

Hurling at the abyss

by A. Stasenko

ONCE UPON A TIME, HAVING donned his heavy armor, a knight set out to perform his regular duty of liberating a kidnapped princess. Suddenly, a narrow pathway ended at an abyss, and the only thing that was left of the bridge was a platform suspended on unstretchable cables over the middle of the abyss (fig. 1). The platform was far enough from the edge that our errant hero could neither reach it nor jump to it. In addition, jumping would be risky since the platform was covered with ice.

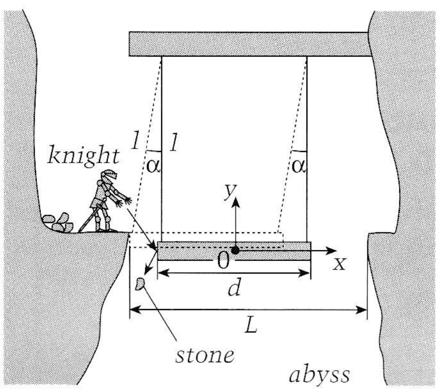


Figure 1

In utter desperation the knight took a stone of mass m and hurled it directly at the bridge. The sound of an absolutely elastic collision echoed many times, and the stone plunged into the abyss. The knight glared at the bridge, and then his jaw dropped in astonishment: The bridge had begun to oscillate! The warrior recalled the principles of applied mechanics he learned at military college and quickly guessed what had happened.

The elastic impact of the stone with the edge of the heavy platform and the following rebound resulted in a change in the stone's momentum of $\Delta p = -2mv_x$ (fig. 2). Therefore, the same size impulse was imparted to the platform:

$$\Delta p = 2mv_x.$$

The canny knight was pressed for time (the princess awaited him!), so he simplified his reasoning and assumed that the mass of the stone was far less than that of the bridge. Those who are not in a hurry may repeat these calculations, taking into account the fact that the stone bounces off a moving platform. Thus, after the first collision, the platform's speed increased by

$$\Delta V = \frac{\Delta p}{M},$$

and the platform began to oscillate with a small swing and almost no damping.

Our resourceful hero used the dusty pathway to plot this process. In rectangular coordinates he drew the dependence of the velocity V on the displacement x. Before the first collision, the platform was at the origin 0 (fig. 3). At the moment of collision its speed increased by ΔV (note the upward arrow drawn to point A in fig. 3a). In other words, some kinetic energy was imparted to the platform, and it began to move in the positive x-direction.

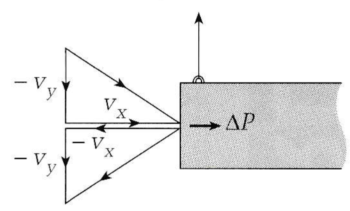


Figure 2

Since the cables do not stretch, the platform's center of mass moves along an arc. Thus, the platform is lifted in the gravitational field. During this motion the kinetic energy is converted into potential energy, and when the speed becomes zero (point B), the displacement from the equilibrium position is a maximum. Having reached this point, the platform changes direction and begins to move toward the knight, gaining maximum speed at point C and stopping for an instant at point D_1. If damping is absent, this process repeats itself infinitely.

Scrutinizing his graph, the knight understood that it was worthwhile to throw another stone at point D_1. After the next swing the platform should come to point D_2 (figure 3b), and so on, until point D_N coincides with $x = (L - d)/2$.

Thus, hesitations behind, the brave knight decided to estimate how many stones he would need to

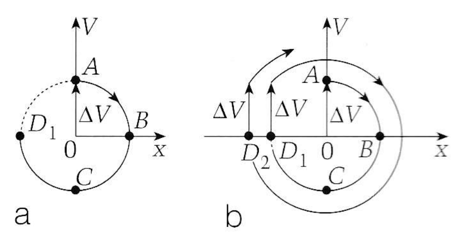


Figure 3

hurl to draw the edge of the slippery platform just to the edge of the precipice so that he could safely step onto the platform.

He laid down his shield and on it wrote the law of conservation of energy:

$$M\frac{V_{\max}^2}{2} = Mgl(1-\cos\alpha_{\max}),$$

where $V_{\max}$ is the maximum speed of the platform (we can see that it is achieved at the lowest point of the oscillation trajectory), and $\alpha_{\max}$ is the maximum angle of deflection, which can be easily found from the right triangle (fig. 1):

$$\cos\alpha_{\max} = \frac{L-d}{2l}.$$

At this angle, the platform's velocity is zero: All its kinetic energy has turned into potential energy. Assuming that every stone hurled at the moment when the platform assumes the position nearest the knight increases the platform's momentum by the same value, the necessary number of stones can be found from the formula:

$$V_{\max} = N\Delta V = N\frac{\Delta P}{M} = N\frac{2mv_x}{M}.$$

By plugging this formula into the conservation of energy formula, the knight obtained

$$\frac{1}{2}\left(\frac{N\cdot 2mv_x}{M}\right)^2 = gl\left(1-\frac{L-d}{2l}\right),$$

from which he got

$$N = \frac{M}{2mv_x}\sqrt{g(2l-L+d)}.$$

Now the smart and brave knight could obtain the numerical estimates assuming the platform's mass to be $M = 10^3$ kg, that of a stone $m = 1$ kg, the horizontal projection of a stone's velocity at the moment of collision $v_x = 10$ m/s, the width of the abyss $L = 50$ m, the platform's length $d = 30$ m, and the length of each cable $l = 50$ m. Calculations gave him the number of stones to be hurled:

$$N = 1.4\cdot 10^3.$$

How long would the chivalrous knight need to hurl stones? The number of oscillations is known, so we need only one thing more—the value of the period of oscillation. Of course, it depends on the length of the pendulum l (m) and the acceleration due to gravity g (m/s²). To obtain the necessary dimension (s) from these two values, we can combine them only in the following way:

$$\sqrt{\frac{l(\text{m})}{g(\text{m/s}^2)}} \cong T(\text{s}).$$

At this fleeting moment of revelation the knight recalled what his grandpa used to say in the fifteenth century: "Remember, child, when oscillations are considered, the number 2π always jumps into the formulas by some miracle."

Therefore, the period of the platform's oscillation is

$$T = 2\pi\sqrt{\frac{l}{g}} = 2\pi\sqrt{\frac{50\text{ m}}{10\text{ m/s}^2}} = 14\text{ s}.$$

Since T is also the period between successive collisions, we can see that our gallant hero should work no less than

$$t = 14\cdot 1.4\cdot 10^3\text{ s} \cong 2\cdot 10^4\text{ s} \cong 5.5\text{ h}.$$

(It was lucky that the knight could disregard damping!) Confronting this daunting task was not easy, but the princess desperately needed the knight's help, so he dutifully began his deed. ◘

Quantum articles about conservation laws and parametric resonance:

A. Chernoutsan, "Swinging Techniques," May/June 1993, pp. 64–65.

A. Eisenkraft and L. Kirkpatrick, "Click, click, click ...," September/October 1990, pp. 41–42.

A. Eisenkraft and L. Kirkpatrick, "Moving Matter," November/December 1996, pp. 32–33.

ANSWERS, HINTS & SOLUTIONS

Math

M246

If a number has the form $p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_k^{\alpha_k}$, where $p_1, p_2, \ldots, p_k$ are prime integers, the number of its different divisors (including 1 and the number itself) is $(\alpha_1 + 1)(\alpha_2 + 1) \ldots (\alpha_k + 1)$. This formula is well known and can be easily proved. (For a proof of this formula, see any book on elementary number theory.)

We see that the number of different primes involved in the prime factor decomposition of our number cannot be greater than four, because $2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 > 1000$. We will look for the smallest of the desired numbers (if more than one such number exists). Therefore, in the prime decomposition of this number, only the primes 2, 3, 5, and 7 can occur, and only to certain powers.

Now we face a simple search problem: 2^9 has 10 divisors, $2^8 \cdot 3$ has 18 divisors, and so on. We also should consider the following numbers: $2^7 \cdot 3$, $2^6 \cdot 3^2$, $2^6 \cdot 3 \cdot 5$, $2^5 \cdot 3^3$, $2^4 \cdot 3^2 \cdot 5$, $2^3 \cdot 3^3$, $2^3 \cdot 3^2 \cdot 5$, $2^3 \cdot 3 \cdot 5 \cdot 7$, and $2^2 \cdot 3^2 \cdot 5^2$. We immediately see that the number $2^3 \cdot 3 \cdot 5 \cdot 7 = 840$ has the maximum number (32) of divisors. There are no other three-digit numbers with this many divisors.

Answer: 840.

M247

For example, the function

$$f(x) = \begin{cases} \dfrac{1}{x}, x \geq 1, \\ 2 - x, x \leq 1 \end{cases}$$

possesses the properties described (fig. 1).

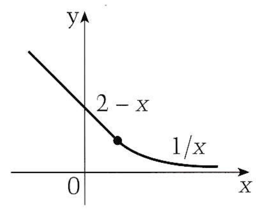


Figure 1

M248

If triangles ABD, BCD, and CAD are congruent, then they are all equilateral and the angles at vertex D are 60°. Now suppose that the three triangles are not all congruent. It follows from the problem statement that the radii of the circles circumscribed about the three triangles are equal. Indeed, they each have a side of the same length, and the angles opposite these sides are equal, and these elements determine the circumradius. Now, triangles ABD and CAD are isosceles (with bases BD and CD, respectively). Therefore, sides BD and CD of triangle BCD are equal. Consider this triangle and circumscribe a circle about it. Construct two triangles A_1BD and CA_2D inscribed in this circle and equal to triangles ABD and CAD, respectively (fig. 2). It follows from our considerations that the pentagon DA_1BCA_2 is regular. Thus the angles we seek are 36° in this case.

Answer: 60° or 36°.

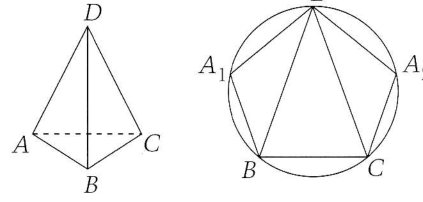


Figure 2

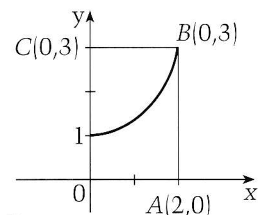


Figure 3

M249

Consider a rectangle with the vertices $O(0, 0)$, $A(2, 0)$, $B(2, 3)$, and $C(0, 3)$ on the coordinate plane. The graph of the function $y = \sqrt{1+x^3}$ passes through the points (0, 1) and (2, 3) and partitions our rectangle into two parts (fig. 3). The area under the graph is

$$\int_0^2 \sqrt{1+x^3}\,dx.$$

We compute the area of the part of the rectangle that is above the graph. The function $y = \sqrt{1+x^3}$ is monotonic on the segment [0, 2]. Thus we can express x in terms of y:

$$x = \sqrt[3]{y^2 - 1}.$$

Therefore, the area we seek is

$$\int_1^3 \sqrt[3]{y^2 - 1}\,dy.$$

Make the following change of variable under the integral sign: $y = t + 1$. This yields the following expression for the area:

$$\int_0^2 \sqrt[3]{t^2 + 2t}\,dt.$$

Now, we see that the given integral is equal to the area of the rectangle

$OABC$, which is 6.

Answer: 6.

M250

We use the traditional notation for the sides of the triangle: $BC = a$, $CA = b$, and $AB = c$ (where $b > c$). Let P, M, and K be the midpoint of BC, the point at which BC is tangent to the inscribed circle, and the foot of the altitude dropped to BC, respectively (fig. 4). Let F be

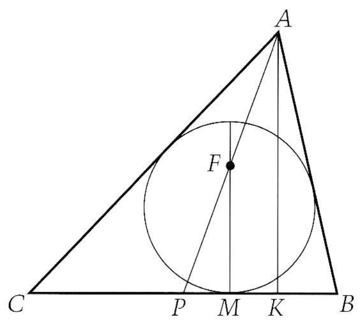


Figure 4

the point of intersection of AP with the diameter of the inscribed circle that passes through M, let r be the radius of the inscribed circle, and let $AK = h_a$. We have: $CP = a/2$, $CM = s - c$ (where s is the half-perimeter of the inscribed circle[1]) and

$$CK = b\cos C = \frac{2ba\cos C}{2a} = \frac{a^2 + b^2 - c^2}{2a}.$$

(This last equality is from the law of cosines.)

These equalities yield

$$PM = CM - CP = \frac{b-c}{2}$$

and

$$PK = CK - CP = \frac{b^2 - c^2}{2a}.$$

From the similar triangles PMF and PKA, we find

$$\frac{MF}{h_a} = \frac{PM}{PK} = \frac{a}{b+c} = \frac{a}{b+c+a-a} = \frac{1}{k-1}. \qquad (1)$$

In addition, the well-known formulas for the area of a triangle yield[2]

$$\frac{2r}{h_a} = \frac{a}{s} = \frac{2a}{a+b+c} = \frac{2}{k}. \qquad (2)$$

Now, equations (1) and (2) yield

$$\frac{2r - MF}{MF} = \frac{2r}{MF} - 1 = \frac{2r}{h_a}\frac{h_a}{MF} - 1 = \frac{2(k-1)}{k} - 1 = \frac{k-2}{k}.$$

Answer: $\frac{k-2}{k}$.

[1]This formula can be obtained as follows. Let N and L be the points of tangency of the inscribed circle with the sides AC and AB, respectively, and $CM = CN = x$. Then, $AL = AN = b - x$, $BM = BL = c - (b - x) = c - b + x$. Now the equation $CM + BM = BC$, or $x + (c - b + x) = a$, yields $x = (a + b - c)/2 = s - c$.

[2]If K denotes the area, then we have $K = ½\, h_a a$ and $K = rs$.

Physics

P246

Let's denote the acceleration of the lower wedge by a, noting that it is horizontal (fig. 5). Relative to this

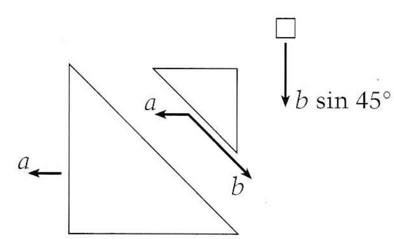


Figure 5

wedge, the acceleration of the upper wedge is directed along the plane of their contact—that is, downward at an angle of 45° to the horizontal.

The total acceleration of the upper wedge can be conveniently decomposed into the sum of two vectors: the horizontal one equal to a, and the second vector with the value b, which is directed downward at the angle of 45°. Thus, the horizontal projection of the upper wedge's acceleration is directed opposite the acceleration of the lower wedge and is equal to $(b \cos 45° - a)$, and the vertical acceleration is $b \sin 45°$. Because no horizontal force acts on the brick, its acceleration is directed vertically and equals the vertical projection of the acceleration of the upper wedge: $b \sin 45°$.

Let's find the relationship between a and b. Recall that in the absence of external horizontal forces, the horizontal acceleration of the center of mass must be zero. Since the brick moves only in the vertical direction,

$$M(b\cos 45° - a) = Ma,$$

$$b = \frac{2a}{\cos 45°} = 2\sqrt{2}a.$$

Therefore, the horizontal component of the total acceleration of the upper wedge relative to the table is a, and the vertical component of this acceleration (and of the brick itself) is $2a$. The value of the total acceleration of this wedge is $\sqrt{5}a$.

To determine the accelerations of the bodies, let's use conservation of energy. In a time τ the loss in potential energy will be equal to the gain in the total kinetic energy of the system. The speed and displacement gained to the time τ can be obtained with the formulas of uniformly accelerated motion. A decrease in the potential energy results from the vertical displacement of the upper wedge and the brick:

$$\Delta U = (M + m)g\frac{2a\tau^2}{2}.$$

The total kinetic energy of the bodies is

$$K = \frac{M(a\tau^2)}{2} + \frac{M(\sqrt{5}a\tau)^2}{2} + \frac{m(2a\tau)^2}{2}.$$

By equating ΔU and K, we obtain the acceleration a:

$$a = g\frac{M+m}{3M+2m}.$$

Therefore, in a time τ after the start of the motion, the speed of the brick will be

$$v = 2a\tau = 2g\tau\frac{M+m}{3M+2m}.$$

P247

The thermal flow Q is proportional to the temperature difference $(T_2 - T_1)$ per unit length L along the direction of thermal flow and to the cross-sectional area S. Denoting the coefficient of proportionality by K (known as the coefficient of thermal conductivity) we obtain the thermal flow in the first case:

$$Q_1 = \frac{KS(T_2 - T_1)}{L}.$$

The second case is far more difficult. Let's assume that the total thermal flow is composed of the component flows that are directed along the fibers (at an angle α to the cylinder's axis) and perpendicular to them. Of course, we could take other directions to decompose the vector, but the chosen ones are most suitable, because we know the coefficients of thermal conductivity along them. In the "longitudinal" direction (along the fibers) the temperature drop per unit length is smaller than in the first case: It is equal to $[(T_2 - T_1)\cos\alpha]/L$. The cross-sectional area in this direction is also smaller: $S\cos\alpha$. Similar formulas are valid for the thermal flow across the fibers, but in that case angle α must be replaced by $(90° - \alpha)$ and the thermal conductivity coefficient must be halved. Thus, in the second case the total thermal flow is

$$Q_2 = \frac{KS\cos^2\alpha\cdot(T_2 - T_1)}{L} + \frac{0.5KS\sin^2\alpha\cdot(T_2 - T_1)}{L}.$$

The ratios of thermal flows will be

$$\frac{Q_1}{Q_2} = \frac{1}{0.75 + 0.5\cdot 0.25} = \frac{8}{7} \cong 1.14.$$

P248

The coils are connected in parallel, so at any moment their self-induced emfs are equal and their changes in magnetic flux are identical. The same conclusion can be inferred by considering the superconducting contour formed by both inductors.

Labeling currents in the coils I_1 and I_2, we have (taking into account the signs corresponding to the chosen directions of positive current):

$$L_1I_0 = L_1I_1 + L_2I_2.$$

The maximum charge on the capacitor occurs when the current to the capacitor is zero, or rather, when $I_1 = I_2$. This charge can be found using conservation of energy:

$$Q_m = I_0\sqrt{\frac{CL_1L_2}{L_1 + L_2}}.$$

The maximum current in the second coil is reached when the charge on the capacitor is zero. At this moment, conservation of energy can be written without the term for the energy stored in the capacitor:

$$L_1I_1^2 + L_2I_2^2 = L_1I_0^2.$$

Combining this equation with the equation for the magnetic fluxes, we obtain

$$L_1I_0^2 = L_1\left(I_0 - \frac{L_2I_2}{L_1}\right)^2 + L_2I_2^2.$$

This equation has two roots for I_2. One of them is zero and corresponds to the minimum value of this current. Note that the conditions for the minimum and maximum are quite similar: In both cases the self-induced emf is zero. The second root is the answer to the problem:

$$I_2 = 2I_0\frac{L_1}{L_1 + L_2}.$$

P249

If there were no magnetic field, the frame would move under the action of gravity with constant velocity $(\mathbf{v}_0)$ in a horizontal direction (x-axis) and with a constant acceleration $\mathbf{g}$ along the z-axis. A homogeneous magnetic field B could not change the motion of the frame. However, in this problem the field varies along the z-axis: $B(z) = B_0 + kz$. This means that the field grows linearly with z, so the magnetic flux Φ threading the frame will change during the fall. Accordingly, an emf will be generated in the frame, and the induced current will flow around this closed contour. The frame will experience a force from the magnetic field. Let's find the value and direction of this force.

At some moment t the frame's center of mass is located at the point (x_t, z_t). The projections of its velocity on the x- and z-axes are v_x and v_z. At this moment the magnetic flux Φ threading the frame is

$$\Phi = \frac{\left(B_0 + k\left(z_t - \frac{a}{2}\right)\right) + \left(B_0 + k\left(z_t + \frac{a}{2}\right)\right)}{2}a^2 = (B_0 + kz_t)a^2.$$

Here $B_0 + k(z_t - a/2)$ and $B_0 + k(z_t + a/2)$ are the values of the magnetic field at the upper and lower sides of the frame. Since $B(z)$ is a linear function, we use the mean value of B to calculate the magnetic flux Φ.

At this moment, the emf in the frame is

$$|\mathscr{E}| = \frac{|\Delta\Phi|}{\Delta t} = ka^2\frac{|\Delta z|}{\Delta t} = ka^2|v_z|,$$

and the induced current is

$$I = \frac{|\mathscr{E}|}{R} = \frac{ka^2}{R}|v_z|.$$

According to Lenz's law, the induced current flows counterclockwise. According to Ampere's law, the upper side of the frame experiences the force

$$|\mathbf{F}_1| = \left(B_0 + k\left(z_t - \frac{a}{2}\right)\right)Ia = \left(B_0 + k\left(z_t - \frac{a}{2}\right)\right)\frac{ka^3}{R}|v_z|.$$

and the lower one the force

$$|\mathbf{F}_2| = \left(B_0 + k\left(z_t + \frac{a}{2}\right)\right)Ia$$
$$= \left(B_0 + k\left(z_t + \frac{a}{2}\right)\right)\frac{ka^3}{R}|v_z|.$$

We can see that the forces $\mathbf{F}_3$ and $\mathbf{F}_4$ acting on the lateral sides of the frame are equal by value and have opposite signs:

$$|\mathbf{F}_3| = |\mathbf{F}_4|$$
$$= \frac{\left(B_0 + k\left(z_t - \frac{a}{2}\right)\right) + \left(B_0 + k\left(z_t + \frac{a}{2}\right)\right)}{2}Ia$$
$$= (B_0 + kz_t)\frac{ka^3}{R}|v_z|,$$

and

$$\mathbf{F}_3 + \mathbf{F}_4 = 0.$$

Thus, v_x = const, so the frame will move along the horizontal axis with a constant velocity equal to the initial velocity v_0.

Therefore, the vertical motion of the frame is determined by the forces $\mathbf{F}_1$, $\mathbf{F}_2$, and the force of gravity $m\mathbf{g}$. Since the velocity $\mathbf{v}$ of the frame is constant, the velocity projection on the z-axis is also constant, and the vertical acceleration $\mathbf{a}_z$ is zero:

$$m|\mathbf{a}_z| = m|\mathbf{g}| + |\mathbf{F}_1| - |\mathbf{F}_2|$$
$$= mg - \frac{k^2a^4}{R}|v_z| = 0.$$

From here we get the projection $v_{st.z}$ of the steady-state frame's velocity on the z-axis:

$$v_{st.z} = \frac{mgR}{k^2a^4}.$$

Thus, the steady-state speed of the frame is $v = \sqrt{v_0^2 + v_{st.z}^2}$ where v_0 is the x-projection of the velocity $\mathbf{v}$, which is equal (as we have shown) to the initial velocity imparted to the frame. Therefore,

$$v_0 = \sqrt{v^2 - v_{st.z}^2} = \sqrt{v^2 - \left(\frac{mgR}{k^2a^4}\right)^2}.$$

There is another way to find $v_{st.z}$, which is based on conservation of energy. In the steady-state motion, the change in the potential energy of the frame moving under the action of Earth's gravity during the period Δt is equal to the heat dissipated by the frame in the same period:

$$mgv_{st.z}\Delta t = I_{st.z}^2 R\Delta t = \left(\frac{ka^2}{R}\right)^2 v_{st.z}^2 R\Delta t.$$

Thus,

$$v_{st.z} = \frac{mgR}{k^2a^4}.$$

P250

A grain of iron is heated to the temperature T at which the energy radiated by this grain equals the energy received from the Sun. The energy radiated per unit time is proportional to the square of the surface area of the grain. To estimate this area, we assume the grains to be spherical and have a mean radius r. The power (energy per unit time) radiated by a grain heated to the temperature T_0 equals

$$P_{rad} = 4\pi r^2 L.$$

The incident solar power is proportional to the area of the largest cross-section of the particle (πr^2):

$$P_{inc} = \pi r^2 L_0.$$

Here L_0 is density of solar radiation at the distance R from the Sun where the grains are located. Since a constant amount of energy is radiated by the Sun into a unit solid angle (fig. 6), then $LS_1 = L_0S_2$ and

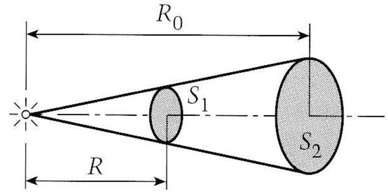


Figure 6

$$L = L_0\frac{S_2}{S_1} = L_0\left(\frac{R_0}{R}\right)^2.$$

Thus,

$$P_{inc} = \pi r^2 E_0\left(\frac{R_0}{R}\right)^2.$$

Equating P_{rad} and P_{inc}, we obtain

$$4\pi r^2 P = \pi r^2 L_0\left(\frac{R_0}{R}\right)^2.$$

Finally,

$$R = \frac{R_0}{2}\sqrt{\frac{L_0}{P}} \cong 5 \cdot 10^6 \text{ km}.$$

Note: Mercury, the nearest planet to the Sun, is $5 \cdot 10^6$ km from it. Why doesn't it melt?

Brainteasers

B246

If none of these months is February, then the total number of days in them cannot be less than $91 = 7 \cdot 13$. Thus, the total number of Sundays would not be less than 13.

B247

One solution is

$$4 \cdot \sqrt{4} \cdot \sqrt{4} = 16.$$

B248

See figure 7. The desired polygon cannot have fewer than four sides, because each of its sides cannot in-

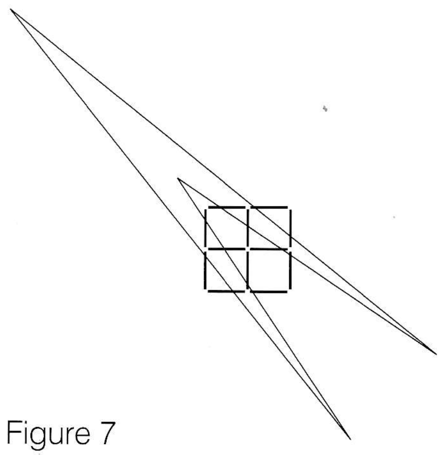


Figure 7

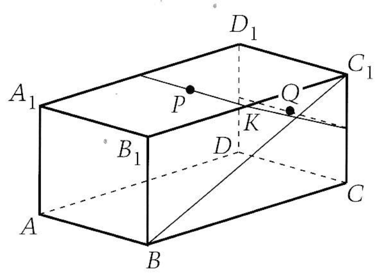


Figure 8

tersect more than two of the eight exterior segments.

B249

See figure 8. Draw line segments through points P and Q parallel to AB such that the segments on both sides of the points are equal to half of AB. The endpoints of these segments are the midpoints of the edges B_1C_1, D_1A_1, CC_1, and DD_1. Label the midpoints of B_1C_1 and CC_1 as K and M. The diagonal BC_1 of the parallelogram BCC_1B_1 passes through the midpoint of the segment KM and is partitioned by the intersection point in the proportion 3:1 (from vertex B). Now it is easy to construct point C_1. For this purpose, we must connect B with the midpoint of KM and extend this segment by 1/3 of its length. The rest of the construction is easy to complete.

B250

Memory again played a trick on the old baron. It is impossible to change the direction of motion without interaction with other bodies.

Errors in geometrical proofs

If your objection is that the bisector and perpendicular bisector meet outside triangle ABC while the points P and Q lie on its sides (fig. 9), you are wrong. As before, $|AN| = |NC|$ and $|NP| = |NQ|$. Therefore, triangles ANP and CNQ are congruent, and $\angle NAP = \angle NCQ$. Besides, triangle ANC is isosceles, and thus, $\angle NAM = \angle NCM$. Subtracting equal angles from equal angles, we find that $\angle BAC = \angle BCA$—that is, triangle ABC is isosceles.

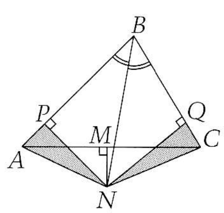


Figure 9

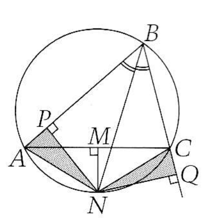


Figure 10

What actually happens in the general case is that the perpendicular bisector and bisector meet at a point N outside the triangle, and the perpendiculars dropped from N onto the sides BA and BC are situated such that one of them falls on the side itself and the other falls on the extension of the other side (fig. 9). In this case, again, $\triangle APN = \triangle CNQ$ and $\angle NAP = \angle NCQ$. Besides, $\angle NAM = \angle NCM$. However, in this case, $\angle BAC = \angle NAP - \angle NAM$ and $\angle BCA = 180° - (\angle NCQ + \angle NCM)$. We can see that this does not imply that angles BAC and BCA are equal.

However, you should not be satisfied by this explanation. We must prove that the elements of the triangle are really situated as shown in figure 10 (or vice versa, $Q \in [BC]$ and $P \notin [AB]$).

Circumscribe a circle about triangle ABC and assume that $|AB| > |BC|$. By a property of inscribed angles, the bisector of angle B meets arc AC at its middle point. The perpendicular bisector of AC passes through the same point. Thus, N is the midpoint of arc AC. Since $|AB| > |BC|$, $\overset{\frown}{AB} > \overset{\frown}{BC}$. Adding the equal arcs $\overset{\frown}{AN}$ and $\overset{\frown}{NC}$ to both sides of this inequality, we obtain the inequality $\overset{\frown}{BAN} > \overset{\frown}{BCN}$. Therefore, $\overset{\frown}{BAN}$ is larger than the semicircle, and $\overset{\frown}{BCN}$ is smaller. Thus, $\angle NCB$ is obtuse, and $\angle NAB$ is acute. Therefore, point Q lies on the extension of the side BC, and P lies on the side AB.

Problem 1. We have groundlessly assumed that the sum of the angles of the triangle is a constant and does not depend on the triangle. In fact, in Lobachevskian geometry, in which the parallel axiom does not hold, the sum of the angles of any triangle is less than 180°, and the difference between 180° and the

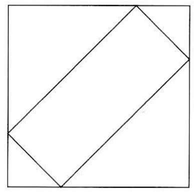


Figure 11

sum of the angles of the triangle is proportional to its area. Thus, the closer the sum of the angles of the triangle is to 180°, the smaller the triangle's area.

Problem 2. Conduct the proof using the drawing in figure 11.

Problem 3. We proved quite correctly that segments A_iA_{i+1} and B_iB_{i+1} do not intersect. However, for $i \neq j$, A_iA_{i+1} and B_jB_{j+1} can intersect.

Problem 4. If the bigger circle rolls without slipping, the smaller circle inevitably slips. Therefore, segment CD is not equal to the circumference of the smaller circle.

Problem 5. The sequence of wavelike curves actually approaches the segment AB. However, this fact does not imply that the sequence of their lengths approaches the length of AB.

Problem 6. We tacitly assumed that the small spherical triangle can be developed (unrolled) onto the plane and used the formula for the area of the plane triangle. Actually, no spherical triangle can be developed onto the plane. In our case, this follows from the fact that our spherical triangle has two right angles, which is impossible in the case of a plane triangle.

Gradus

2. We complete two squares at the same time, proceeding as in problem 1:

$$\begin{aligned} x^2 + y^2 - 8x + 6y + 17 &= x^2 - 8x + 16 + y^2 + 6y + 9 - 8 \\ &= (x-4)^2 + (y+3)^2 - 8. \end{aligned}$$

As before, the minimal value is –8, achieved when $x = 4$ and $y = -3$.

3. We know that the ellipse

$$a(x-h)^2 + b(y-k)^2 = c$$

has its center at the point (h, k). We can put the given equation in this form by completing the square:

$$4x^2 + 6y^2 - 24x + 12y = 4(x^2 - 6x) + 6(y^2 + 2y) = 20,$$

so

$$4(x^2 - 6x + 9) + 6(y^2 + 2y + 1) = 4(x - 3)^2 + 6(y + 1)^2 = 62,$$

and the center of the ellipse is at (3, –1). In more advanced work involving linear algebra, this technique is generalized quite far.

4. Completing the squares, we obtain

$$\left(x+\frac{19}{2}\right)^2+\left(y+\frac{99}{2}\right)^2+r-\frac{5081}{2}\geq 0.$$

So the smallest r with the desired property is 5081/2.

5. There are no such functions. Since the function involves the squares of real numbers (and all the problems in this column involve squares!), we might want to look at the numbers 0 and 1, which behave atypically when squared.

Indeed if there were such a function, we would have $f(0) - f^2(0) \geq 1/4$ and $f(1) - f^2(1) \geq 1/4$. Completing the square, we find that $(f(0) - 1/2)^2 \leq 0$ and $(f(1) - 1/2)^2 \leq 0$. But then $f(0) - f(1) = 0$, so that f is not one-to-one after all.

6. Suppose $a - b^2$, $b - c^2$, $c - a^2$ all exceeded 1/4. Then

$$(a - b^2) + (b - c^2) + (c - a^2) > 1/4 + 1/4 + 1/4 = 3/4,$$

or

$$a^2 - a + b^2 - b + c^2 - c < -3/4.$$

Completing the square, we find that this is equivalent to

$$(a - 1/2)^2 + (b - 1/2)^2 + (c - 1/2)^2 < 0,$$

which is absurd.

9. As in problem 8, we have

$$\begin{aligned}4a^4 + b^4 &= 4a^4 + 4a^2b^2 + b^4 - 4a^2b^2\\ &= (2a^2 + b^2)^2 - (2ab)^2\\ &= (2a^2 + 2ab + b^2)(2a^2 - 2ab + b^2).\end{aligned}$$

10. For even n, the given number is itself even (and greater than 2), so it is composite. If n is odd, we can write

$$4^n+n^4=4\left(2^{\frac{n-1}{2}}\right)^4+n^4.$$

then, in the result of problem 9, we let

$$a=2^{\frac{n-1}{2}}$$

and $b = n$, and we will have our factorization. We need only check that the smaller of these two factors (the second) is not 1. Indeed, this factor is just

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

and for $n \geq 3$, this is certainly more than 1.

11. Multiplying out, we have

$$2a^2 + 2b^2 \geq a^2 + 2ab + b^2,$$

which reduces to

$$a^2 - 2ab + b^2 \geq 0,$$

or

$$(a - b)^2 \geq 0,$$

which is certainly true. As in problem 7, the square is already complete.

13. This result is a bit strange. The left-hand side depends on c, but the right-hand side does not. Let's start with the left side. Proceeding as in problem 12, we have

$$a^2 + b^2 + c^2 - ab - bc - ca = (1/2)[(a - b)^2 + (b - c)^2 + (c - a)^2].$$

But how do we eliminate c from this expression? Noting that

$$(b - c) + (c - a) = (b - a),$$

we can use the result of problem 11 to see that

$$(b - c)^2 + (c - a)^2 \geq (1/2)(b - a)^2,$$

and the conclusion follows:

$$\begin{aligned}a^2 + b^2 + c^2 - ab - bc - ca &= (1/2)[(a - b)^2 + (b - c)^2 + (c - a)^2]\\ &\geq [(a - b)^2 + (b - c)^2 + (c - a)^2]\\ &= (3/4)(a - b)^2.\end{aligned}$$

14. Multiplying out the right-hand side quickly leads us to the statement of problem 12.

15. The reader may have suspected this general result already. The algebra is the same as in problem 10, and it is straightforward.

16. (Solution II.) We have the expression

$$S = x^4 + y^4 + z^4 - 4xyz.$$

We can make a square out of the first two terms by subtracting $2x^2y^2$, and out of z^4 by subtracting $2z^2 - 1$:

$$S = x^4 - 2x^2y^2 + y^4 + z^4 - 2z^2 + 1 - 4xyz + 2x^2y^2 + 2z^2 - 1.$$

Now we have all the squares we need, since

$$\begin{aligned}S &= x^4 - 2x^2y^2 + y^4 + z^4 - 2z^2 + 1\\ &= 2(x^2y^2 - 4xyz + z^2) - 1\\ &= (x^2 - y^2)^2 + (z^2 - 1)^2 + 2(xy - z)^2 - 1,\end{aligned}$$

so again the minimum is –1, achieved for the same values of x, y, and z as before.

17. Our strategy will follow the pattern of problem 12 (and some others in this set): We transform the problem into one where we want to show that the sum of a bunch of terms is nonnegative, then express the sum as a sum of squares. Indeed, the given inequality is equivalent to

$$a_1-2\sqrt{a_1-1^2}+a_2-2\cdot 2\sqrt{a_2-2^2}+\ldots+a_n-2n\sqrt{a_n-n^2}\geq 0. \qquad (1)$$

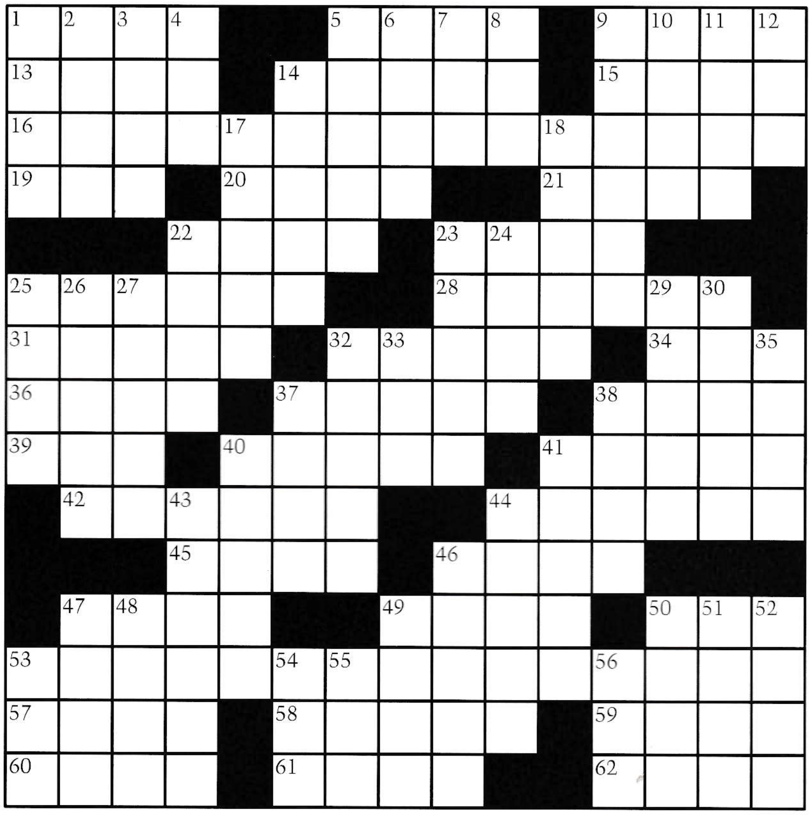
A typical pair of terms on the left side is $a_k-2k\sqrt{a_k-k^2}$. Let's try to make this into a perfect square. If this perfect square is $A^2 - 2AB + B^2$, an inspection of the radical term leads us to conjecture that we should try $A=\sqrt{a_k-k^2}$, and we can write

$$\begin{aligned}a_k-2k\sqrt{a_k-k^2} &= a_k-k^2-2k\sqrt{a_k-k^2}+k^2\\ &=\left(\sqrt{a_k-k^2}-k\right)^2.\end{aligned}$$

This is true for each k, so the expression on the left in equation (1) is a sum of squares, and cannot be less than 0.

Criss × cross science

by David R. Martin



Across

1 Sodium carbonate
5 Unruly child
9 Parasite's provider
13 ___ wax (ozocerite)
14 Magnetic flux density unit
15 Turkish title
16 $\mathbf{F}_{12} = -\mathbf{F}_{21}$
19 Type of cell or ice
20 Ancient German
21 A sand desert
22 Fourth dimension
23 ___ sapiens
25 Not inert
28 Seventh planet
31 Gay ___
32 Norw. playwright Henrik ___ (1828–1906)
34 Catch
36 Certain asteroid
37 Copper and zinc
38 Enclosure
39 Quill
40 Darling
41 Sulked
42 Type of heat
44 ___ of mass (Brit. sp.)
45 Central American oil tree
46 Eight bits
47 Comfortable
49 Swiss river
50 Abscisic acid: abbr.
53 $T^2 = (4\pi^2/GM)a^3$
57 Width times length
58 Aluminum or copper, e.g.
59 Winglike
60 Focusing device
61 Boot accessory
62 Chro-___ bike frame

Down

1 It is chiefly quartz
2 European river
3 Like the morning grass
4 Not science
5 Moisten while roasting
6 "Cell Heredity" author, with 30D
7 Olive family tree
8 Approx. $1.54444 \cdot 10^7$ pascals
9 Strong particle?
10 Stare flirtatiously
11 Floor covering
12 Fancy marble
14 Midget
17 Pointed arch
18 Type of spectroscopy
22 Layer
23 Bold, saucy girl
24 Natural mineral compounds
25 Certain analgesic
26 Desert animal
27 Urao
29 Not likely
30 See 6D
32 Angry
33 Unit of pressure
35 Miner's pick
37 Seismologist ___ Gutenberg (1889–1960)
38 Retina cell
40 1936 chem. Nobelist
41 Unit of length
43 Webers/m^2
44 Sci-fi writer ___ Kornbluth (1923–1958)
46 Iranian poet Mohammad ___ (1885–1951)
47 Bird's nose
48 Uncovered
49 Aleutian island
50 Stable isomer prefix
51 Fertility god
52 Amiss
53 ___ Baisakhi (dusty squall in Bengal)
54 ___ value (avg.)
55 Aug. follower
56 1943 physiol. Nobelist

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE SEPTEMBER/OCTOBER PUZZLE

T	E	R	A			T	A	M	P		F	L	E	D
A	L	U	L	A		E	N	O	L		R	A	T	A
L	I	P	I	N		S	T	O	A		A	S	H	Y
C	A	P	A	C	I	T	O	R	S		U	S	E	S
				O	R	A	N		M	I	N	O	R	
A	R	S	I	N	E			T	A	N	H			
S	E	A	R		D	E	C	I		C	O	M	E	T
I	N	U	R	E		C	U	E		A	F	I	R	E
S	O	L	A	R		C	O	R	I		E	N	I	D
			T	O	T	E			S	E	R	I	E	S
	P	O	I	S	E		P	E	E	R				
H	O	B	O		F	A	H	R	E	N	H	E	I	T
O	W	E	N		L	E	A	R		S	A	L	V	O
M	E	S	A		O	R	S	O		T	U	B	A	S
O	R	E	L		N	O	E	L			L	E	N	S

Contact

by Dr. Mu

WELCOME BACK TO COWCULATIONS, THE column devoted to problems best solved with a computer algorithm. Around farm animals there are various methods to communicate, make contact, or otherwise get understood. Farmer Paul knows all of his registered Holsteins personally and can tell instantly when one of us is having a bad-hair day. Yes, an occasional cow will step out of line and throw her weight around, but all cows ultimately know their place. When 50 cows walk into the barn in the morning for milking, everyone goes immediately to her own stall. And when milking is over, the senior citizens have the privilege of leaving the barn first.

Making contact with other farm animals is another story. As a rule, very little communication exists between us. Recently, I took up the challenge of trying to understand my fellow ostrich, who babbles on constantly. I decided to do some research on bit patterns associated with numbers in hopes of applying my results to deciphering ostrichspeak. My goal was to examine the bit patterns of their speech and record the frequency of the most common patterns. I began my study by examining the bit patterns

Art by Mark Brenneman

for some rather large numbers such as the millionth prime number raised to the 10th power.

```
x = Prime[10^6]^10
```

7931478080938745260256491876836253331064605399963659381034368467337248 49

Next, I transformed this number into its binary representation. The algorithm to transform numbers to other bases is short, sweet, and fast.

Suppose that x is to be expressed in base b. That is, $x = a_0 + a_1 b + a_2 b^2 + \ldots + a_n b^n$. To find the coefficients (the base b representation of x), we do the following.

$$\mathbf{a_0 = Mod[x,b]},$$
$$\mathbf{x = Floor}\left[\frac{\mathbf{x}}{\mathbf{b}}\right] = \mathbf{a_1 + a_2 b^1 + \ldots + a_n b^{n-1}},$$
$$\mathbf{a_1 = Mod[x,b]},$$
$$\mathbf{x = Floor}\left[\frac{\mathbf{x}}{\mathbf{b}}\right] = \mathbf{a_2 + a_3 b + \cdots + a_n b^{n-2}},$$
$$\mathbf{a_2 = Mod[x,b]}.$$

We repeat this process of computing **Mod[x, b]** and reducing x to **Floor[x/b]** until $x = 0$. The base b representation of $x = (a_n, a_{n-1}, \ldots, a_1, a_0)$ is the coefficients in reverse order as they were generated from the algorithm above. Here is the simple and fast ($O(\log_b(x))$) algorithm for converting x to base b.

```
baseExpansion[x_, b_] := Module[{q = x,
    ans = {}},
  While[q ≠ 0, AppendTo[ans,
  Mod[q, b]]; q = Floor[q/b]];
Reverse[ans]]
```

Now we can find the binary representation of our number.

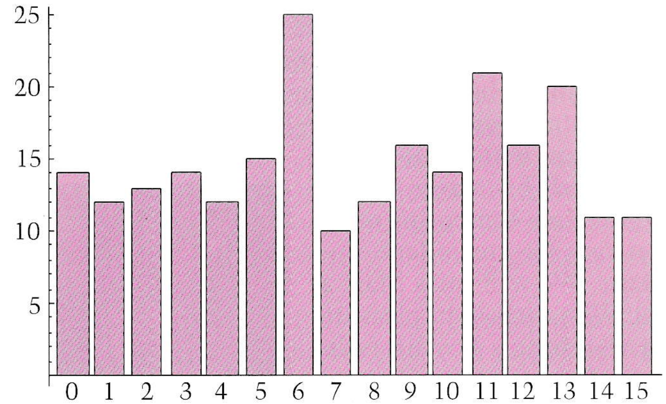
```
baseExpansion[x, 2]
```

```
{1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1,
1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1,
1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1,
1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1,
1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1,
0, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 1,
0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0,
0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1,
0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1,
1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1,
0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0,
1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1,
0, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0,
1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0,
0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1}
```

Next, I searched for all consecutive bit strings of length 4. There are 236 such strings within the bit string shown above. They matched one of 16 possible patterns {0, 0, 0, 0}, {0, 0, 0, 1}, ..., {1, 1, 1, 1}. The following is the frequency count found for these patterns.

14	{0, 0, 0, 0}
12	{0, 0, 0, 1}
13	{0, 0, 1, 0}
14	{0, 0, 1, 1}
12	{0, 1, 0, 0}
15	{0, 1, 0, 1}
25	{0, 1, 1, 0}
10	{0, 1, 1, 1}
12	{1, 0, 0, 0}
16	{1, 0, 0, 1}
14	{1, 0, 1, 0}
21	{1, 0, 1, 1}
16	{1, 1, 0, 0}
20	{1, 1, 0, 1}
11	{1, 1, 1, 0}
11	{1, 1, 1, 1}

Finally, I assigned each of the bit patterns its corresponding decimal number, $\{0, 0, 0, 0\} \to 0$, $\{0, 0, 0, 1\} \to 1$, $\{a, b, c, d\} \to a2^3 + b2^2 + c2^1 + d$, ..., $\{1, 1, 1, 1\} \to 15$, and graphed the results.



The graph represents the frequency count of the bit string patterns $\{a, b, c, d\}$ found in the binary representation of the millionth prime raised to the 10th power. Each bit pattern is assigned the corresponding integer $a2^3 + b2^2 + c2^1 + d$ in the graph. This suggests a problem, which, you guessed it, is your Challenge Outta Wisconsin.

COW 13

Write a program that will transform any number into its corresponding binary representation and produce a frequency count for consecutive bit strings for any specified length found in the binary representation. For a bit pattern of length n, assign the decimal number to each bit string and graph the corresponding frequency distribution defined on $\{0, 1, \ldots, 2^n - 1\}$. Test your program on the number x = **Prime[10^6]^{100}**, and examine the bit strings of length 4.

While listening to the Ostrich sing,
Output the sounds in a long bit string.
Find the patterns that keep the beat,
And count them all as they repeat.
When you're done, present the graph.
You've solved this COW and done the math.
—Dr. Mu

Letter from Portugal

Paul & Paula's Holstein Dairy
Creamery Lane,
Primeville, Wisconsin

September 12, 1998

Dear Paul and Paula,

How time flies when you're having fun in Portugal. Here it is the next-to-last day of the 10th IOI (International Olympiad in Informatics) in Setúbal, about an hour's drive south of Lisbon. We just got back from the awards ceremony at the town theater, where 22 gold, 40 silver, and 59 bronze medals were awarded to approximately half of the 248 participants representing 65 countries.

Our veteran, Matt Craighead, from St. Paul Academy in Minnesota, received a silver medal in his final appearance as a USA team member. In the two previous IOI's he received a bronze and a gold medal, making him our most decorated team member from the United States. Even though he is only 16 and normally would have three more years of eligibility, he just entered MIT as a freshman, which automatically ends his career in this world-class computing competition for pre-college students. We will miss Matt a lot. We have had the pleasure of watching a boy literally grow before our eyes into a fine young man who we'll hear from again. In fact, we'll be calling on him in the year 2003, as well as on other former USA team members, to help us conduct the 15th IOI in the United States.

Adrian Sox, 18, from Upper Dublin High School in Fort Washington, Pennsylvania, brought home a gold medal in his first try at IOI. Adrian had been our top-ranked competitor all year, winning one of the three Internet competitions, ranking number one in the USA National Competition, and placing first in the week-long summer training program held at the University of Wisconsin–Parkside in July. It was reassuring to know that compared to the best computer problem solvers from all over the world, many of them with years of IOI experience, Adrian ranks with the very best. This was Adrian's final year, because he entered Carnegie Mellon University as a freshman in the fall. We not only lost our top competitor, but a very tough ultimate Frisbee player. Expect a call from us in 2003, Adrian.

Alex Wissner-Gross, a junior at Great Neck South High School in Great Neck, New York, and Chuong Do, a sophomore at Garland High School, in Garland, Texas, completed our team of four members. Both missed the cut-off score to medal this year, but both have a chance, since they are still in high school, to come back and try again. Just getting to the IOI is a reward in itself—an all-expense-paid trip to a new part of the world for an exciting and stimulating week of fun and competition. But now, with a bit of experience under their belts, Alex and Chuong have become veterans who know what it takes to crack into the medal range at IOI. We're expecting great things from them in 1999.

You may be wondering if an IOI is all work and no play. Let me tell you a bit about its recreational side. Two days were set aside for excursions. One of the reasons that Portugal hosted the IOI in 1998 was the presence of the 1998 World Exposition. So, naturally, one of the excursions was to Expo 98. Another excursion was to the Palmela castle, which the first king of Portugal took from the Moors in 1147 and later offered to the nuns of Saint James. Today the castle, restored by

Staff and finalists of the 1998 USA Computing Olympiad.

The 1998 USA Computing Team: (from left to right) Matt Craighead, Adrian Sox, Chuong Do, and Alex Wissner-Gross.

the state, is a national monument hotel.

Running an IOI with 250 students is not something a country gets much experience doing. After all, this is a once-in-a-lifetime undertaking. As a result, it is not unusual to have a few bugs creep into the operation. Sure enough, a bug hit at the conclusion of the first day's five-hour competition. The program being used by the judges to grade all 248 students failed. As a result, the grading was delayed approximately eight hours while the bugs were fixed. This caused exhaustion both for the team members, who had been up since 5:00 A.M., and the judges charged with grading all the programs. We made the decision to send our team to bed after 11 P.M. The United States team, which comes near the end of the alphabet, was not graded until 2 A.M. I stayed around to witness the process and just made it for the last ferry ride back to Troia at 3 A.M.

One might think that grading computer programs would be easy. A data set of numbers is input to each program, and the output is checked with a known solution. If the output matches the known solution and it is produced within a given time limit, you get full credit for the data set. A total of five different data sets were tested with each program, and the total score was then computed. A perfect score meant all five data sets produced the expected answer within the time period allotted. What could be easier?

Automated grading systems are wonderful, if they work as expected. They are completely objective and do all the laborious checking of answers. However, if they have a mind of their own or have been instructed to only accept answers of a particular form, then a correct program that produces all the right output can be given a zero mark.

This is not just a hypothetical situation; it happened on the first problem of the first day's competition. Students who produced a correct answer—let's say 123, but output their solution as 123#, where # is a space, received zero points for a problem worth 100 points. Correcting this problem required a meeting of the general assembly, which debated the rules and overwhelmingly voted that this problem should be regraded since the IOI is basically an *algorithm* contest and not a *tricky output* competition.

Unfortunately, the problems that plagued the first round overwhelmed the judges, and they were never able to recover the time to do the regrading. So the results stood, and approximately 20 students lost 100 points each, which had the effect of dropping them down one medal level. Matt was one of those affected. But it didn't bother him as much as it did the rest of us. He knew he had the correct answer, and that's all that mattered. In all fairness to the Portuguese judges, they worked very hard under difficult and stressful conditions. We applaud them for their valiant efforts.

Tomorrow the team and its leaders and I are headed home. I took lots of digital photos that I will put up shortly on our web site at http://usaco.uwp.edu. By the way, Slovakia had the most successful team at this year's IOI and the only one with four gold medals. You don't need to be a big country to have top-flight computer programmers.

Team leaders Don Piele, Rob Kolstad, and Greg Galperin all send their best wishes. Oh yes, I can't forget to mention our wonderful sponsor USENIX, which puts up the money each year in support of the USA Computing Olympiad. What would we do without them?

Looking forward to my return to Wisconsin.

Dr. Mu

p.s. The 1999 IOI will be held in Antalya, Turkey. Can I go along as the team mascot again? ◘

Bulletin Board

Teacher fellowship program

Earthwatch Institute will offer funding for K–12 educators to participate in two-week field research expeditions around the world during spring and summer 1999. Through generous support from program sponsors, Earthwatch Institute aims to develop multidisciplinary science and social studies curriculums in schools nationwide and to rejuvenate teachers and enhance the academic experience of students. The Earthwatch Institute Teacher Fellowship Program is endorsed by NSTA, the National Science Teachers Association.

Earthwatch Institute, an international nonprofit organization founded in 1972, supports the work of renowned scientists in the fields of archaeology, cultural diversity, endangered ecosystems, and biodiversity, among others. While in the field, educators work side by side with researchers on one of more than 60 ongoing research studies. In most cases, no special skills are necessary. Funding ranges from partial to full fellowships with travel stipends offered to approximately 50 percent of participants.

In 1998 more than 200 Teacher Fellows from 35 states took part in projects such as:

- *Roman Fort on Tyne.* Excavate a Roman military site in northern England to determine why the fort was abandoned, and investigate Roman influence on the region.
- *Dancing Birds.* Observe mating rituals of the seemingly altruistic long-tailed manakin in Costa Rica and consider the genetic implications of this unique behavior.
- *Search for Neanderthals.* Follow in the footsteps of our distant relatives and study how they met their basic needs, communicated, and developed culturally 100,000 years ago.
- *End of the Dinosaurs.* Research the theory that changes in Earth's climate triggered a massive extinction among dinosaurs.

Applicants are encouraged to download an application from the Earthwatch *Global Classroom* at www.earthwatch.org. Applications are due **February 15, 1999,** with rolling admissions after that date contingent upon available funding.

To receive an application by mail, please contact Matt Craig, Education Awards Manager, by calling 617-926-8200 ext. 118 or sending e-mail to mcraig@earthwatch.org.

Sunday best

It takes a quick thinker and an agile mouse-clicker to be one of the first 10 correct respondents to *Quantum*'s CyberTeaser Contest. Indeed, when faced with this month's calendrical question (brainteaser B246 in this issue), the following 10 readers didn't need a month of Sundays to figure it out.

Bruno Konder (Rio de Janeiro, Brazil)
Paul Williams (Red Deer, Canada)
Liam Hardy (Union City, Cal.)
Manny Dekermenjian (Sunnyvale, Cal.)
Theo Koupelis (Wausau, Wisc.)
Leo Borovskiy (Brooklyn, N.Y.)
H. Scott Wiley (Weslaco, Tex.)
John Fernandes (Fremont, Cal.)
Eli Bachmupsky (Kfar-Saba, Israel)
Farokh Jamaly Aria (Deer Park, Tex.)

Each winner will receive a free copy of the September/October issue and a *Quantum* button. Everyone who submitted a correct answer in the allotted time was eligible to win a copy of *Quantum Quandaries*, a collection of the first 100 *Quantum* brainteasers.

Hankering for a prize of your own? Keep an eye out for the new contest problem at www.nsta.org/quantum and click the Contest button.

Review

Russians in Space, CD-ROM (Mac, Windows), $44.95 plus shipping and handling. Published in 1997 by the Ultimax Group, 112 Mason Lane, Oak Ridge, TN; 1-800-ULTIMAX; http://catalog.com/ultimax/.

Forty years ago the *Sputnik* spacecraft was launched by the Soviet Union. This heralded the human race into the space age, and much has happened since that first initial foray into outer space. The CD-ROM *Russians in Space* chronicles the history, programs, technology, and other aspects of the Soviet space program.

CD-ROMs can function as valuable encyclopedias with their ability to contain massive collections of sound, graphics, and text that can help enrich learning. This CD helps the learning process by providing brilliant color graphics, music, narration, textual information, video clips, and a variety of topics related to the Soviet space program.

Russians in Space is divided into

four main sections. The section called "Personal" includes information about the development of the space program during the eighteenth and nineteenth centuries, including key figures, jet propulsion information, and the engineers who designed the rockets and made them fly. The "Programs" section details the Mars and Venus missions and the attempts to land on and explore planets with unmanned spacecraft. The "Tech" section highlights the spacecraft from the first *Sputnik* to the *Buran* space shuttle look-alike. The "Basic" section covers artificial and spy satellites, orbital stations, ground control, and research programs.

The top main tool bar allows the user to access additional information via a timeline and history of the space program. Navigation and searching tools within this menu bar (unfortunately sometimes covered by video/screens) also can assist in finding specific information.

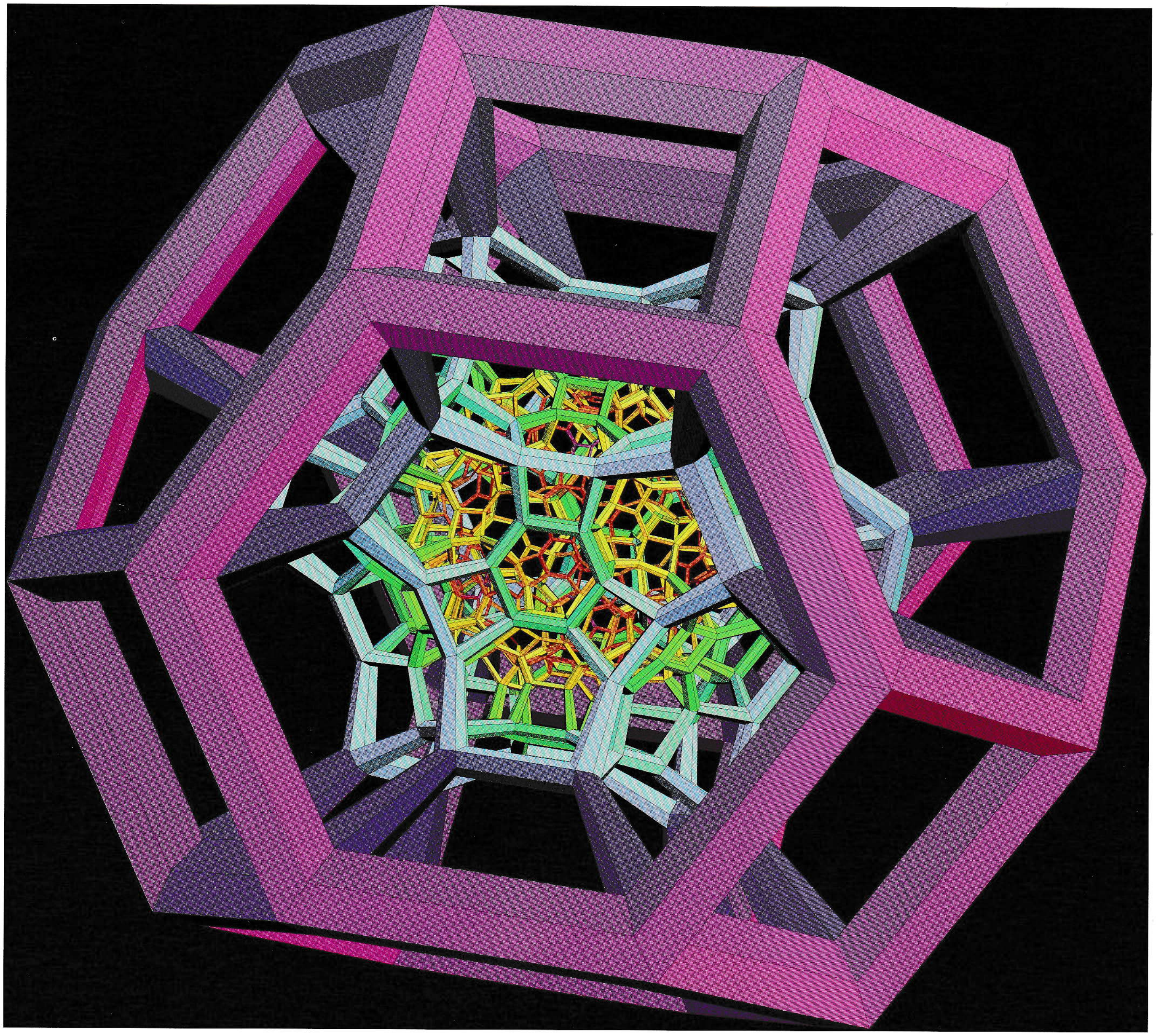
The CD is quite engaging and well designed. Educators at the elementary to high school level who would like to survey information about the Russian space program will find the CD useful if they are willing to do a little work to integrate the CD into their program to meet their needs and objectives.

—Eric Flescher

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION (Required by 39 U.S.C. 3685). (1) Publication title: Quantum. (2) Publication No. 008-544. (3) Filing Date: 10/1/98. (4) Issue Frequency: Bimonthly. (5) No. of Issues Published Annually: 6. (6) Annual Subscription Price: $45.00. (7) Complete Mailing Address of Known Office of Publication: 175 Fifth Avenue, New York, NY 10010-7858. Contact Person: Joe Kozak, Telephone 212-460-1500. (8) Complete Mailing Address of Headquarters or General Business Office of Publisher: 175 Fifth Avenue, New York, NY 10010-7858. (9) Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor: Publisher: Gerald F. Wheeler, National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201-3000 in cooperation with Springer-Verlag New York Inc., 175 Fifth Avenue, New York, NY 10010-7858. Editors: Larry D. Kirkpatrick, Mark E. Saul, Albert L. Stasenko, Igor F. Sharygin, National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201-3000. Managing Editor: Mike Donaldson, National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201-3000. (10) Owner: National Science Teachers Association. (11) Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities: None. (12) The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes: Has Not Changed During Preceding 12 Months. (13) Publication Name: Quantum. (14) Issue Date for Circulation Data Below: September/October 1998. (15) Extent and Nature of Circulation: (a.) Total No. Copies (Net Press Run): Average No. Copies Each Issue During Preceding 12 Months, 8708; Actual No. Copies of Single Issue Published Nearest to Filing Date, 7300. (b.) Paid and/or Requested Circulation: (1) Sales Through Dealers and Carriers, Street Vendors, and Counter Sales: Average No. Copies Each Issue During Preceding 12 Months, 1435; Actual No. Copies of Single Issue Published Nearest to Filing Date, 944. (2) Paid or Requested Mail Subscriptions: Average No. Copies Each Issue During Preceding 12 Months, 5166; Actual No. Copies of Single Issue Published Nearest to Filing Date, 5142. (c.) Total Paid and/or Requested Circulation: Average No. Copies Each Issue During Preceding 12 Months, 6601; Actual No. Copies of Single Issue Published Nearest to Filing Date, 6086. (d.) Free Distribution by Mail: Average No. Copies Each Issue During Preceding 12 Months, 191; Actual No. Copies of Single Issue Published Nearest to Filing Date, 180. (e.) Free Distribution Outside the Mail: Average No. Copies Each Issue During Preceding 12 Months, 622; Actual No. Copies of Single Issue Published Nearest to Filing Date, 630. (f.) Total Free Distribution: Average No. Copies Each Issue During Preceding 12 Months, 813; Actual No. Copies of Single Issue Published Nearest to Filing Date, 810. (g.) Total Distribution: Average No. Copies Each Issue During Proceeding 12 Months, 7414; Actual No. Copies of Single Issue Published Nearest to Filing Date, 6898. (h.) Copies Not Distributed: (1) Office Use, Leftovers, Spoiled: Average No. Copies Each Issue During Preceding 12 Months, 1294; Actual No. Copies of Single Issue Published Nearest to Filing Date, 404. (2) Return from News Agents: Average No. copies Each Issue During Preceding 12 Months, 0; Actual No. Copies of Single Issue Published Nearest to Filing Date, O. (i.) (Sum of 15g, 15h(1), and 15h(2)): Average No. Copies Each Issue During Preceding 12 Months, 8708; Actual No. Copies of Single Issue Published Nearest to Filing Date, 7300. Percent Paid and/or Requested Circulation: Average No. Copies Each Issue During Preceding 12 Months, 89.03; Actual No. Copies of Single Issue Published Nearest to Filing Date, 88.25. (16) This Statement of Ownership will be printed in the Vol. 9 #2 issue of this publication. I certify that all information furnished on this form is true and complete.

Dennis Looney
Senior Vice President and Chief Financial Officer

GALLERY Q II



Inverted lattice of orthotetrakaidecahedrons (1998) by Michael Trott

BY JUXTAPOSING EXCELLENT ART AND ILLUSTRAtions with thought-provoking discussions of mathematics and science, *Quantum* has always striven to emphasize the importance of different modes of thinking and to draw connections between the seemingly disparate worlds of art and science. This image, created by Michael Trott using *Mathematica* software, goes one step further, obliterating the boundaries between those two worlds.

The basic geometric unit in this image is the orthotetrakaidecahedron, or truncated octahedron, which has the notable property of being a space filler—a solid that when packed together with same-size copies leaves no spaces in between. Trott gives a two-step "recipe" for creating this type of image: "First, create a regular lattice by repeating a certain shape; second, turn that lattice inside out."

Lattices are important in crystallography, a fact expanded on in "Lattices and Brillouin Zones" on page 4 in this issue. Look again at the image above. Are you looking at the fancy of an artist, the beauty of a mathematical pattern, the internal structure of a crystal? Where is the demarcation between art and science?