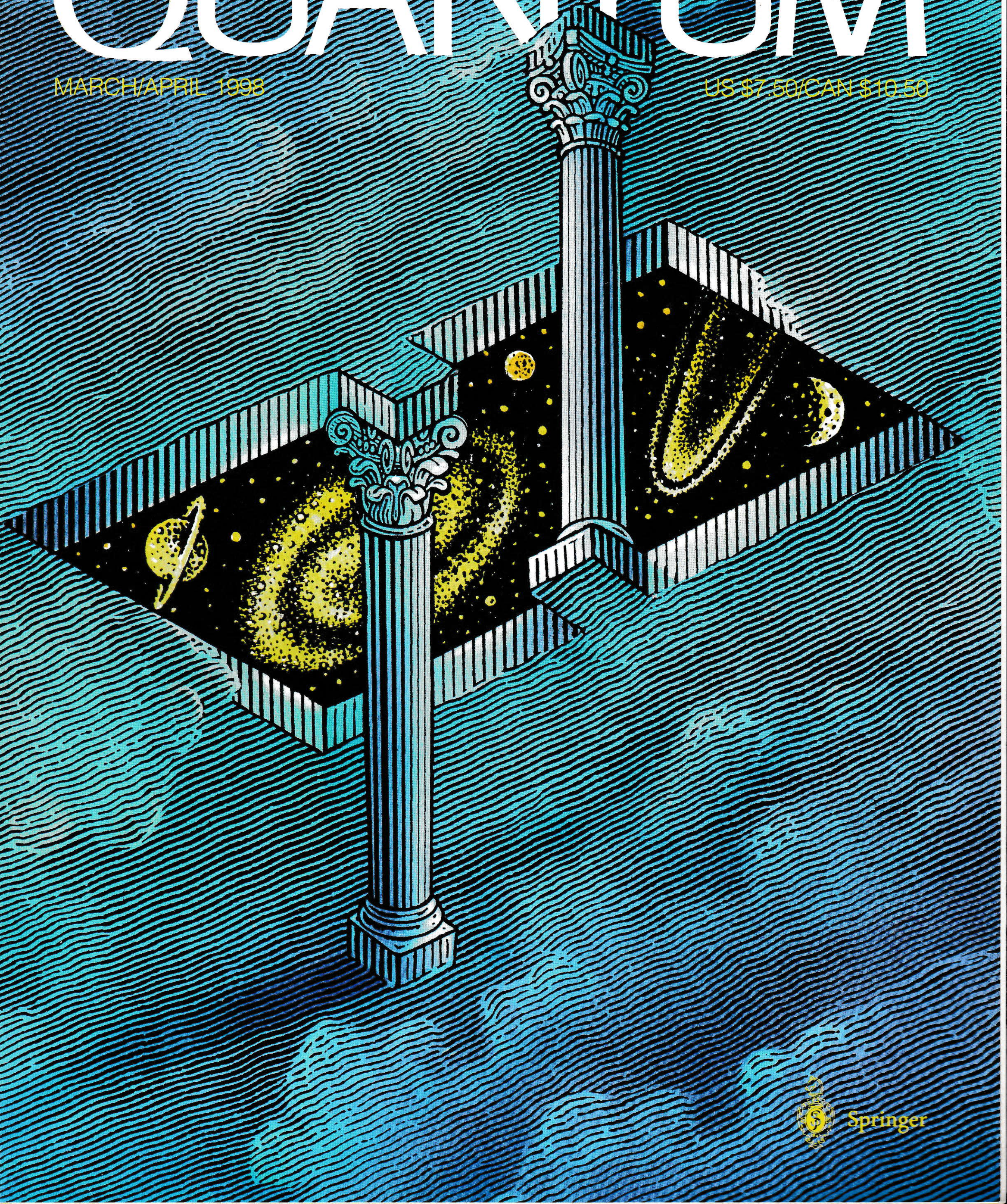
QUANTUM

MARCH/APRIL 1998

US $7.50/CAN $10.50



Springer

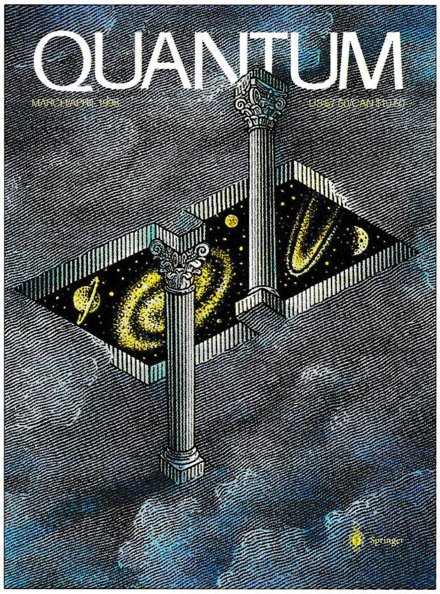
Oil on panel, 91 × 71 inches, Collection of The Salvador Dali Museum, St. Petersburg, Florida © 1997 Salvador Dali Museum, Inc.

Skull with Its Lyric Appendage Leaning on a Night Table Which Should Have the Exact Temperature of a Cardinal Bird's Nest (1934) by Salvador Dali

DOES YOUR HEART RESONATE WITH DREAD when you recall the piano lessons your parents forced you to take as a kid? They may have seemed deadly dull at the time, but other resonant experiences can pose a real (as opposed to surreal) threat to your health. To find out why an imperfect runway might one day strike a chord of terror in your heart, turn to page 45 to find out more about "The Horrors of Resonance."

QUANTUM

MARCH/APRIL 1998 VOLUME 8, NUMBER 4



Cover art by István Orosz

The expansion of the Universe has been a hot topic of discussion since the turn of the century. Many theories have been developed to explain how gravity and other factors influence universal contraction and expansion.

However, much like the artwork on this month's cover, upon closer examination these theories reveal incongruities that prevent us from getting a clear look at what is going on in the Universe.

To learn more about the historical efforts scientists have made to come to terms with our dynamic Universe, see page 10. Find out what all the *heated* debate is about.

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

Science with charm

Communicating the simplicity of physics

ELEMENTARY PARticles with spin are not spinning. Those with color have no color. Strange particles are neither more strange nor less strange than other elementary particles. Particles with charm are not charming, yet physicists with charm are charming.

If I pick up a box from the floor, move it across the room, and then place it on the floor again, I will not have done any work even though I may be paid for my work. Energy is always constant, so how can we be running out of energy? A 50 kilogram person exerts more pressure on the floor than a 1,000 kilogram elephant, although the elephant exerts a greater force on the floor.

Science, of course, has its own language, as is true with most specialized areas of knowledge. Yet, science sometimes usurps the layperson's language and attaches a very special meaning. In cases of words like *charm, strangeness,* and *color,* the physicist attaches specific meanings of no relevance to the language of the general populace. That wouldn't be much of a problem except that it reinforces the popular belief that physics is incomprehensible to all but a select few. It almost seems as though physics is mocking the language of the nonphysicist. Does the language difference exacerbate the alienation between those who understand two languages and yet speak the same language? Or does the use of common words give physics a more friendly face?

One recent Nobel Prize winner in physics gushed enthusiastically in response to a reporter's question that he began to study physics because of its simplicity. Yet most students who take a physics course don't recognize simplicity among its attributes. One person sees simplicity in the same evidence that daunts another person with its overwhelming complexity.

While physicists focus their attention on concepts and ideas, many students in our classrooms see physics as a collection of formulas with an array of confusing variables. Regardless of what the teacher encourages, many of our students grope for the single formula that connects the given data with the unknown quantity. While practicing physicists focus on principles and understanding, it is the "tyranny of technique" that dismays many students of physics.

Wouldn't it be wonderful if students taking their first physics course received a hint of why a Nobel laureate finds simplicity a primary attribute of our discipline? Is there some way to demonstrate simplicity by means other than a tortuous trail through complexity? Why must simplicity be evident only to the most passionate practitioners?

—Bernard V. Khoury

Bernard V. Khoury *is the Executive Officer of the American Association of Physics Teachers.*

Art by Michael Jones

Be a factor in the QUANTUM equation!

Have you written an article that you think belongs in *Quantum*? Do you have an unusual topic that students would find fun and challenging? Do you know of anyone who would make a great *Quantum* author? Write to us and we'll send you the editorial guidelines for prospective *Quantum* contributors. Scientists and teachers in any country are invited to submit material, but it must be written in colloquial English and at a level appropriate for *Quantum*'s predominantly student readership.

Send your inquiries to

Managing Editor
Quantum
1840 Wilson Boulevard
Arlington VA 22201-3000

What's happening?

Summer study ... competitions ... new books ... ongoing activities ... clubs and associations ... free samples ... contests ... whatever it is, if you think it's of interest to *Quantum* readers, let us know about it! Help us fill Happenings and the Bulletin Board with short news items, firsthand reports, and announcements of upcoming events.

What's on your mind?

Write to us! We want to know what you think of *Quantum*. What do you like the most? What would you like to see more of? And, yes—what *don't* you like about *Quantum*? We want to make it even better, but we need your help.

What's our address?

Quantum
National Science Teachers Assoc.
1840 Wilson Boulevard
Arlington VA 22201-3000

Be a factor in the QUANTUM equation!

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.

Publisher
Gerald F. Wheeler, Executive Director, NSTA

Associate Publisher
Sergey S. Krotov, Director, Quantum Bureau,
Professor of Physics, Moscow State University

Founding Editors
Yuri A. Ossipyan, President, Quantum Bureau
Sheldon Lee Glashow, Nobel Laureate (physics), Harvard University
William P. Thurston, Fields Medalist (mathematics), University of California, Berkeley

Field Editors for Physics
Larry D. Kirkpatrick, Professor of Physics, Montana State University, MT
Albert L. Stasenko, Professor of Physics, Moscow Institute of Physics and Technology

Field Editors for Mathematics
Mark E. Saul, Computer Consultant/Coordinator, Bronxville School, NY
Igor F. Sharygin, Professor of Mathematics, Moscow State University

Managing Editor
Kenneth L. Roberts

Associate Managing Editor
Mike Donaldson

Editorial Assistants
Jennifer M. Wang
Cara Young

Art Director
Sergey Ivanov

Editorial Advisor
Timothy Weber

International Consultant
Edward Lozansky

Assistant Manager, Magazines (Springer-Verlag)
Madeline Kraner

Advertising Managers
Paul Kuntzler (Washington office)
Brian Skepton (New York office)

Advisory Board
Bernard V. Khoury, Executive Officer, AAPT
John A. Thorpe, Executive Director, NCTM
George Berzsenyi, Professor of Mathematics, Rose-Hulman Institute of Technology, IN
Arthur Eisenkraft, Science Department Chair, Fox Lane High School, NY
Karen Johnston, Professor of Physics, North Carolina State University, NC
Margaret J. Kenney, Professor of Mathematics, Boston College, MA
Alexander Soifer, Professor of Mathematics, University of Colorado–Colorado Springs, CO
Barbara I. Stott, Mathematics Teacher, Riverdale High School, LA
Ted Vittitoe, Retired Physics Teacher, Parrish, FL
Peter Vunovich, Capital Area Science & Math Center, Lansing, MI

Editorial Consultants
Yuly Danilov, Senior Researcher, Kurchatov Institute
Irina Oleynik, Managing Editor, Quantum Bureau

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 8 (6 issues) will be published in 1997–1998. *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.

Copyright © 1998 NSTA.
Printed in U.S.A.

Subscription Information:

North America: Student rate: $17.70; 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: $45. 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) Brian Skepton (212) 460-1575; 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.

Ramanujan the phenomenon

India's inspired mathematician

by S. G. Gindikin

IN EARLY 1913 PROFESSOR Godfrey H. Hardy at Cambridge University received a letter from Madras, India. Although only 36 years old then, Hardy had written a series of marvelous mathematical papers and was recognized as one of the world's leading experts in calculus and number theory. His correspondent, Srinivasa Ramanujan, worked as a clerk in the accounting department of the Madras post office, earning a mere 20 pounds per year. He wrote that he hadn't graduated from any university and that he studied mathematics after school alone, not by following the traditional system, but by pursuing his own ways.

A letter of this sort probably wouldn't itself have made much of an impression on Hardy. But accompanying the letter was a list of formulas that Ramanujan proposed for publication if Hardy found them interesting (Ramanujan couldn't publish them himself because of his poverty). When Hardy examined the formulas, he grew excited. He understood that he had come across an outstanding talent. He wrote Ramanujan a letter expressing his interest, and a lively correspondence sprang up between them (it seems amazing how fast the mail traveled between England and India in that time).

Ramanujan's passport photo from 1919, when he returned to India.

Through this correspondence, Hardy accumulated about 120 formulas.

Ramanujan's formulas were primarily concerned with relations between infinite radicals (see inset 2); infinite series, products, and continued fractions (see insets 1, 3, 4); and identities with integrals. Hardy immediately recognized that the formulas went far beyond the limits of elementary mathematics. Further, he posed a sequence of questions: Are they already known? If they are, then did the author of the letter obtain them independently? and If they are not known, then are they correct? Hardy soon realized that he was in a peculiar situation. As a leading expert in calculus, he was dealing with a collection of formulas completely unfamiliar to him.

Hardy was impressed by the relations involving infinite series (inset 1). Studying them, he came to the conclusion that "Ramanujan must possess much more general theorems and was keeping a great deal up his sleeve."[1]

But most of all Hardy was stunned by the relations involving infinite continued fractions (an example of this type of relation, which Ramanujan found later, is given in inset 3). Said Hardy, "[these relations] defeated me completely; I had never seen anything in the least like them before. A single look at them is enough to show that they could only be written down by a mathematician of the highest class."

The miracle from Kumbakonam

So, how did Ramanujan develop into the mathematician who surprised Hardy to such an extent? Srinivasa Aiyangar Ramanujan was born on December 22, 1887, in the

[1]Quotes in this article are taken from Robert Kanigel's book *The Man Who Knew Infinity* (1991, Charles Scribner's Sons).

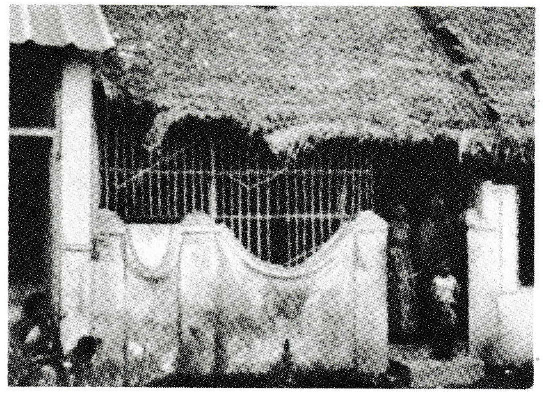
village of Erode in southern India. He spent most of his childhood in the small town Kumbakonam (260 km from Madras), where his father worked as an accountant in a small textile store. Ramanujan belonged to the Brahmin caste, but a long time had passed since his family's wealth had been dissipated. His parents, and especially his mother, were very religious, and Ramanujan was brought up in full accordance with the traditions of his caste. Growing up in a town where every stone was connected to the ancient religion and among people who always remembered that they belonged to the highest caste played a great role in Ramanujan's development as mathematician.

Ramanujan went to school when he was five years old and graduated from elementary school at age 10. At that time he started to show outstanding talents and received a scholarship that covered one-half of the tuition fees. When Ramanujan was 14, one student from Madras gave him the two volumes of *Loney's Guide in Trigonometry*. Ramanujan soon mastered trigonometry so that he could advise the student in solving problems. The first tales and legends about Ramanujan are about this period of his life. For instance, it is said that he discovered Euler's formula by himself and was very disappointed to find it in the second volume of Loney's book.

Ramanujan thought that mathematics, as well as the other sciences, contained some inherent "higher truth" that one should look for, and asked his teachers about it. The teachers, however, only gave him unconvincing references to the Pythagorean theorem and percentage as a ratio.

The two-volume *Synopsis of Elementary Results in Pure and Applied Mathematics,* written by the English mathematician George Shoobridge Carr in 1880–1886, fell into Ramanujan's hands in 1903, when he was 16. The influence it had on Ramanujan's mathematical development was enormous. It accumulated 6165 theorems and formulas, which were presented with minimal explanations and almost no proofs. The book was mainly devoted to algebra, trigonometry, calculus, and analytical geometry.

According to people who knew Ramanujan during that time,Carr's book prompted the young mathematician to derive all the formulas himself. The scope of his major interests was gradually shifting. He worked with magic squares and attempted to square the circle (according to one legend, he found π with a precision that allowed one to calculate the length of the equator with an error of only 1 to 2 meters). Finally, he turned to infinite series. This was the beginning of his life in real mathematics.



Ramanujan's childhood home in Kumbakonam, South India.

Carr's book well fit the purpose of forming Ramanujan's views on mathematics. But its influence had yet another consequence. Because the book contained no rigorous proofs, Ramanujan developed rather strange methods to establish mathematical truths. Besides this, living in India, he was deprived of any suitable manuals that could teach him to make his reasoning strict. About Ramanujan's style of demonstrating the correctness of a formula, Hardy said, "His ideas as to what constituted a mathematical proof were of the most shadowy description. All his results, new or old, right or wrong, had been arrived at by a process of mingled argument, intuition, and induction, of which he was entirely unable to give any coherent account."

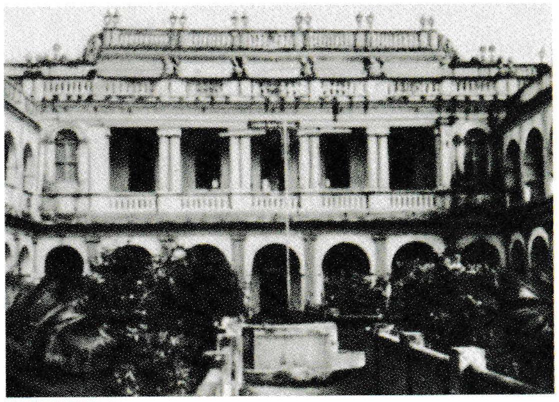
Inset 1. *One infinite sum calculated by Ramanujan.*

$$1-5\left(\frac{1}{2}\right)^3+9\left(\frac{1\cdot 3}{2\cdot 4}\right)^3-13\left(\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\right)^3+\ldots=\frac{2}{\pi}.$$

This magnificent formula was in the list supplied with Ramanujan's first letter to Hardy. Hardy spent a lot of time wondering how on Earth the sum of the alternating series $a_0 + a_1 + a_2 + \ldots$, where a_n was given by

$$a_n=(-1)^n(4n+1)\left(\frac{1\cdot 3\cdot 5\cdot\ldots\cdot(2n-1)}{2\cdot 4\cdot 6\cdot\ldots\cdot(2n)}\right)^3$$

could be equal to 2/p. The reader can use a calculator to check that this formula is valid as an approximate equality. There is no elementary way to prove the exact identity.



Town High School in Kumbakonam.

Ramanujan's life in mathematics was almost completely determined during these years. He would never change the direction of his search or his way of thinking. We can only regret that Ramanujan developed in such a harsh environment. Had the circumstances been different, he would have undoubtedly become a better trained mathematician. But can we be sure that he would become such a unique thinker? Could Ramanujan have understood or discovered as much as he did if he had been taught the rules of mathematical behavior early on, carried his results to publications with rigorous proofs, and based his reasoning on the whole complex of human knowledge rather than a relatively small number of facts?

From numbers to formulas

An important feature of the formation of Ramanujan's approach to mathematics is that he coupled the initial stock of mathematical facts (which he learned from Carr's book), with a great supply of observations of concrete numbers. He had collected these numerical facts since childhood. A schoolmate recalled that Ramanujan had remembered an enormous number of digits in the decimal notations of π and e. He possessed a wonderful ability to derive arithmetical regularities from observations about a huge stock of numerical data, an art mastered by Euler and Gauss but mainly forgotten by the beginning of the twentieth century.

Many facts in Ramanujan's store of numbers were discovered under casual circumstances. Hardy recollected later how he had visited Ramanujan in a hospital and remarked that he had come by a taxi with the "dull" number 1729. Ramanujan grew excited and exclaimed: "No, Hardy. It is a very interesting number. It is the smallest number expressible as the sum of two cubes in two different ways." ($1729 = 1^3 + 12^3 = 9^3 + 10^3$).

Ramanujan rapidly enlarged the stock of facts he had taken from Carr's book. In doing so, he rediscovered at a surprising rate results that belonged to Euler, Gauss, and Jacobi. Similarly, before him young Gauss in Braunschweig, deprived of mathematical literature, reconstructed in a short time the facts that took his great predecessors decades to establish.

Gradually, Ramanujan put aside his collection of numerical observations as he became fascinated by the world of formulas. Formulas were not just auxiliary means of proofs or calculations for him; the internal beauty of a formula was of an infinite value for Ramanujan.

Choice of career

In 1904 Ramanujan entered Madras University. There he made progress not only in mathematics but also in English. However, mathematics was beginning to possess his mind, and this was soon reflected in his grades. He couldn't even finish the first year at the university, went traveling with a friend, later attempted to return to the university, and then tried to take an external degree in 1907. Yet it was all in vain.

In 1909 Ramanujan married. His wife was only nine years old then, and she survived until 1987, tenderly saving recollections about her great spouse. Ramanujan had to look for a means of living, but he couldn't find any suitable job. In 1910 he showed his mathematical results to Ramaswary Iyer, founder of the Indian Mathematical Society, and then to Seshu Iyer, an instructor at Kumbakonam College, and Rama-

Inset 2. *Infinitely iterating radicals.*

$$\sqrt{1+2\sqrt{1+3\sqrt{1+4\sqrt{1+\dots}}}} = 3.$$

Ramanujan found this nice formula when he was yet a schoolboy by writing down the sequence of evident equalities

$$n(n+2) = n\sqrt{1+(n+1)(n+3)} = n\sqrt{1+(n+1)\sqrt{1+(n+2)(n+4)}} = \dots$$

and substituting $n = 1$ in it. The question of whether it was legal to pass to the limit here did not trouble Ramanujan. The reader can try to prove in the same way the following similar formula:

$$\sqrt{6+2\sqrt{7+3\sqrt{8+4\sqrt{9+\dots}}}} = 4.$$

Inset 3. *Numerical identity with an infinite sum and a chain fraction.*

$$1+\frac{1}{1\cdot 3}+\frac{1}{1\cdot 3\cdot 5}+\frac{1}{1\cdot 3\cdot 5\cdot 7}+\frac{1}{1\cdot 3\cdot 5\cdot 7\cdot 9}+\ldots+\cfrac{1}{1+\cfrac{1}{1+\cfrac{2}{1+\cfrac{3}{1+\cfrac{4}{1+\ldots}}}}}=\sqrt{\frac{\pi e}{2}}.$$

This is probably Ramanujan's most beautiful formula—a true masterpiece of mathematical fine art. It draws a sudden connection between an infinite series and a chain fraction. It is wonderful that neither the series nor the chain fraction can be expressed through the well-known constants π and e, and their sum mysteriously proves to be equal to $\sqrt{\frac{\pi e}{2}}$!

chandra Rao, an important functionary who had studied mathematics in a university. Later they became Ramanujan's biographers.

At first, Rao used his own money to support Ramanujan, then helped him get the position in the post office. In 1911 a small report concerning Ramanujan's results, written by Seshu Iar, was published, and some time later Ramanujan's own article appeared. Some influential English officials then started playing a part in Ramanujan's future. In May 1913 he obtained a two-year scholarship of 75 rupees (5 pounds) per month, which was enough to conduct a simple life. Ramanujan abandoned the career of a clerk and became a "professional mathematician."

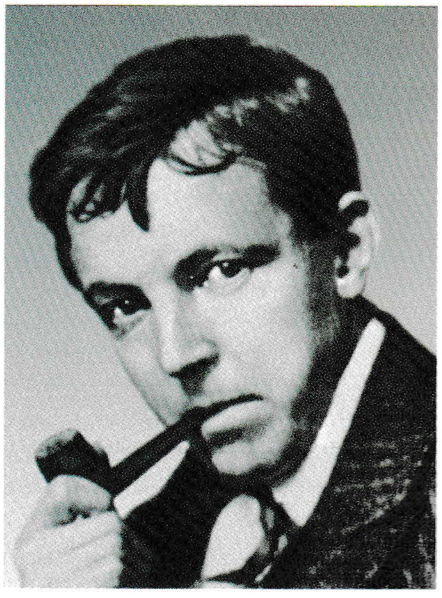
Thus, Ramanujan had found some recognition, if not understanding of his work, from the people around him. We recall that in early 1913 he wrote Hardy. What did he expect? Did he hope to find someone able to understand and appreciate his results and help him direct his further investigations? We'd rather think that his purpose was more prosaic: Ramanujan looked not for recognition or fame but simply for a way to make a living.

We should admit that as far as mathematics was concerned, his choice of addressee was very fortunate: There was hardly another mathematician in the world who could comprehend Ramanujan's results as quickly and as thoroughly. Soon Hardy realized that he had to do more than just appreciate the results of an unknown amateur—he had to save an immense talent. At the same time Hardy was growing convinced that Ramanujan had revealed only a small fraction of the secrets he had discovered. Hardy thought that Ramanujan had obtained some very general results and had shown only particular instances of these results. But what really worried Hardy was that he couldn't reconstruct Ramanujan's methods. He was eager to know the techniques his correspondent employed. However, Ramanujan firmly refused to describe his method. He wrote in his letter dated February 27, 1913:

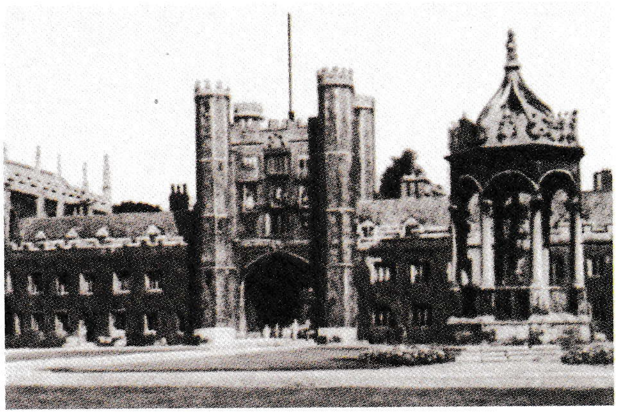
"You will not be able to follow my methods of proof if I indicate the lines on which I proceed in a single letter. You may ask how you can accept results based upon wrong premises. What I tell you is this: Verify the results I give and if they agree with your results, got by treading on the groove in which the present day mathematicians move, you should at least grant that there may be some truths in my fundamental basis."

Hardy supposed that Ramanujan was afraid his methods would be used by other men, so he tried to dismiss Ramanujan's apprehensions, but the answer he received on April 17 was: "I am a little pained to see what you have written.... I am not in the least apprehensive of my method being utilised by others. On the contrary my method has been in my possession for the last eight years and I have not found anyone to appreciate the method. As I wrote in my last letter I have found a sympathetic friend in you and I am willing to place unreservedly in your possession what little I have."

Hardy was convinced that Ramanujan had to meet active mathematicians. This couldn't happen if he remained in India, and thus he had to move to England immediately. Hardy managed to arrange for a scholarship in Cambridge for Ramanujan. But it was still necessary to



G. H. Hardy.



Hardy lived at New Court, Trinity College, during his association with Ramanujan.

convince Ramanujan, who was quite satisfied with his current situation, that he had to go. Besides this, his mother, whose consent was necessary, strongly objected to the journey. His friends tried to raise public interest. The Cambridge mathematician Neville, who visited Madras in 1914, worked energetically toward this purpose, even asking the dean of the local university to support his efforts, but all without any success.

But what scientists couldn't do was easily achieved by the goddess Namagiri. Ramanujan's mother saw her son in a dream, surrounded by Europeans in a large auditorium, and the goddess came and told her that she mustn't object to the trip. So, on March 17, 1914, Ramanujan left for England. He would get a two-year scholarship of 250 pounds per year. From this sum he would send 50 pounds to his mother. Soon after his arrival, the scholarship was increased by 60 pounds.

In Cambridge

When he arrived in Cambridge, Ramanujan was 27 years old. At the time of his life most important for the development of a mathematician, he had lived in India, out of any contact with serious scientists and without access to mathematical literature. People in different countries and in different periods consider themselves grown up at different ages. For India at the beginning of our century, life expectancy being very low, 27 years was the age of a mature man. Ramanujan's widow recalled that Ramanujan had been fond of casting horoscopes and his own predicted that he would die before he was 35.

Hardy had to make an important decision: Was it necessary to interrupt Ramanujan's studies to let him learn modern mathematics? Hardy found, as it seems now, the only possible solution. He decided not to change the style and direction of Ramanujan's investigations but rather to make some corrections to it by taking into consideration modern achievements in mathematics, trying to explain something new, and proposing suitable literature.

Ramanujan worked intensely and fruitfully. He and Hardy had many common interests. Ramanujan's marvelous intuition, coupled with Hardy's refined technique, produced wonderful results. Recognition came to Ramanujan in 1918 when he became a professor in Cambridge University and the first Indian to be chosen as a member of the Royal Society of London.

One can't say that Ramanujan led an easy life. He faithfully observed all the restrictions of his religion, just as he had promised to his parents. In particular, he was a vegetarian and had to do his own cooking. He refused to violate these strictures even when he fell ill with tuberculosis in 1917. The irregularity of his diet may have made the disease progress faster (Ramanujan had held this opinion himself, according to his widow). Ramanujan spent his last two years in England in hospitals and sanitariums, compelled to decrease the intensity of his studies.

Hardy was doing a great deal for Ramanujan. He looked after his studies, tried to fill the gaps in his mathematical education, and took care of his social status and everyday life. Ramanujan was full of gratitude and love to Hardy.

During his illness, Ramanujan began considering the idea of returning to his native land. In early 1919 his physical conditions improved enough to allow him to make the long voyage. He had received an appointment at the University of Madras; his fame was reaching to India. Ramanujan wrote a letter of thanks to the dean in which he apologized

Inset 4. *Ramanujan-Roger's identity.*

$$1+\frac{x}{1-x}+\frac{x^4}{(1-x)(1-x^2)}+\frac{x^9}{(1-x)(1-x^2)(1-x^3)}+\ldots=\frac{1}{(1-x)(1-x^6)(1-x^{11})\cdot\ \ldots\ \cdot(1-x^4)(1-x^9)(1-x^{14})\ldots}$$

Ramanujan found this identity in 1911, but he couldn't prove it then. Hardy couldn't do it either. In 1917, examining mathematical journals (which he seldom did), Ramanujan ran across an 1894 article of the English mathematician Rogers. The article, which until then had remained unnoticed, contained a complete proof of the formula. Later it turned out that this identity is closely connected with the number $p(n)$ of the ways one can represent the number n in the form of a sum (see inset 5). And about ten years ago it appeared in the field of statistical physics.

Inset 5. *Hardy-Ramanujan's theorem.*

This theorem gives an estimate of the number $p(n)$ of the ways one can represent number n in the form of a sum of natural terms. (For example, $p(5) = 7$, since $5 = 4 + 1 = 3 + 2 = = 3 + 1 + 1 = 2 + 2 + 1 = 2 + 1 + 1 + 1 = 1 + 1 + 1 + 1 + 1$.) Namely,

$$p(n) \cong A_n e^{\pi\sqrt{\frac{2}{3}\left(n-\frac{1}{24}\right)}},$$

where

$$A_n = \frac{1}{2\pi\sqrt{2}}\left(\frac{\pi}{\sqrt{6}\left(n-\frac{1}{24}\right)} - \frac{1}{2\left(n-\frac{1}{24}\right)^{\frac{3}{2}}}\right)$$

is a function of n. For instance, when $n = 200$ this "approximate" formula of Hardy and Ramanujan gives $p(200) = 3{,}972{,}999{,}029{,}388$. This is the exact answer! The most mysterious part of this formula for $p(n)$ is the small "correction" $(-1/24)$, suggested by Ramanujan. Nobody, neither Hardy nor even Ramanujan, could explain how it appeared. Another interference of the goddess Namagiri? Yet it was this mysterious correction that made the formula work. Still, Hardy and Ramanujan didn't stop with the approximate formula. They later found an exact *equality* for the number $p(n)$.

for working less intensely recently because of his illness. Still he was not to enter upon his duties in the university. He had less than one year left to live in his native land. After three months in Madras, Ramanujan moved to Kumbakonam. In January 1920 he sent his last letter to Hardy, saying that he was working on a new class of theta functions. Neither his doctors nor his relatives could persuade the morbidly ill scientist to stop his studies. Ramanujan died on April 26, 1920. He was not even 33 years old.

Ramanujan's legacy

The news of Ramanujan's death struck his friends both in India and in England. They felt it was their duty to understand the astonishing phenomenon of Ramanujan. Hardy wrote:

"It is possible that the great days of formulae are finished and that Ramanujan ought to have been born 100 years ago; but still he was by far the greatest formalist of his time."

Friends and colleagues tried to appreciate Ramanujan's place in modern mathematics. They had no doubt concerning his wonderful talents and the amazing beauty of his formulas. But everyone agreed that Ramanujan's choice of subjects makes it difficult for him to take his rightful place in the history of mathematics.

More than 75 years have elapsed since Ramanujan's death, and we now see what Hardy and his contemporaries could not have foreseen. Ramanujan's genius proved consonant not only with the past but also with the future of mathematics. Ramanujan's arithmetical identities often turned out to take central places in the new stage of algebraic number theory, and we can only wonder how he could envision the identities when he didn't know any of the facts one must know to understand them. Later, interest in concrete explicit formulas revived in both pure and applied mathematics.

Modern mathematical and theoretical physics at times resort to very abstract branches of mathematics, where important roles are played by refined formulas. Two relatively fresh examples connected with Ramanujan follow.

Rodney J. Baxter, who became famous for his constructions of exactly integrable models of statistical mechanics, suddenly discovered that he was constantly dealing with Rogers-Ramanujan and Ramanujan's identities while studying the model of a "rigid hexagon" (inset 4).

Nobel prize winner Steven Weinberg recalled that while studying in the early 1970s the now-popular string theory, he faced the problem of estimating the number of decompositions $p(n)$ for large n. It turned out that all the formulas he needed were found by Hardy and Ramanujan in 1918 (inset 5).

The innate beauty of Ramanujan's formulas has endowed them with the wonderful property of turning up now and then under the most unusual circumstances.

For further reading

Ramanujan: Letters and Commentary. Bruce C. Berndt and Robert A. Rankin. Providence, R.I.: American Mathematical Society, 1995.

The Man Who Knew Infinity: A Life of the Genius Ramanujan. Robert Kanigel. New York: Charles Scribner's Sons, 1991.

The thermodynamic Universe

Does time have a beginning and an end?

by I. D. Novikov

WE LIVE IN THE EXPANDING Universe. This fact was theoretically predicted by the Soviet mathematician Alexander Friedmann and confirmed experimentally at the end of the 1920s by the American astronomer Edwin Hubble. The fact that the matter in the surrounding Universe is in constant motion is of fundamental importance to our understanding of physical processes in the Universe.

In this article we discuss some general features of the thermal processes in the macroscopic world and their implications for the evolution of the Universe. The first attempts to apply the laws of thermodynamics discovered in the nineteenth century to the entire Universe led to strange inferences and downright paradoxes. Before considering these cosmological problems, let's briefly recall the essence of Friedmann's theoretical prediction.

Friedmann's cornerstone idea was a stroke of genius, remarkable for its simplicity. It says that on a very large scale (we now know that this means distances greater than hundreds of millions of light years) matter is distributed homogeneously in the form of galaxies and galactic clusters. Gravitational forces attract these huge masses and determine their motions. Whether this attraction ultimately results in expansion or contraction of the system of interacting masses depends on the initial conditions. For example, it depends on whether or not some undetermined forces imparted initial divergent velocities to the substance that later formed the matter of the Universe (that is, upon whether or not the Big Bang occurred). It also could be possible that at some initial moment the matter was extremely spread out, but the gravitational forces caused it to contract with constantly increasing velocity.

Alexander Friedmann used the relativistic theory of gravitation elaborated by Einstein, who generalized Newton's law of universal gravitation for the case of superstrong fields. However, the important results of Friedmann's work can be explained in the framework of Newton's law.

What we need is the basic law of motion for matter in the Universe. This law can be deduced in the following way. Consider in the Universe a large spherical volume of radius R equal to many hundreds of light years. On such a large scale the distribution of matter can be considered homogeneous. Denote the mass in this volume by M. Astronomers have found that galaxies move away from each other, so the chosen volume expands and its boundary moves radially outward. How does the speed of this expansion vary with time?

According to Newton's law of universal gravitation, a galaxy of mass $m << M$ located at the boundary of our sphere is attracted by the gravitational force due to the entire mass within the sphere:

$$F = \frac{GMm}{R^2}.$$

This force slows down the expansion. We do not consider the gravitational forces due to the matter extending to vast distances in every direction outside the sphere because all these forces cancel (we shall not prove it here).

Now we can easily write down the law for the motion of a galaxy at the sphere's boundary—that is, the law of motion of the boundary itself. To do this, we need the total energy E of the galaxy. It is composed of kinetic energy $E_k = mv^2/2$ and gravitational potential energy $E_p = -GMm/R$ (note its negative value):

$$E = E_k + E_p = \frac{mv^2}{2} - \frac{GMm}{R}. \qquad (1)$$

Due to conservation of energy, the total energy E = const.

Equation (1) shows that when $E > 0$, the sphere's radius can increase infinitely: Although v decreases during this process, it does not reach zero. At $R \to \infty$, E is entirely determined by the kinetic energy. On the other hand, when $E < 0$, the gravitational forces arrest the expansion of the sphere, and v becomes zero at a maximum radius

$$R_{max} = \frac{GMm}{|E|}. \qquad (2)$$

Then the sphere begins to shrink.

Because we have chosen our sphere arbitrarily, and since matter is distributed homogeneously in the Universe, the evolution of the sphere's boundary describes the motion of any large mass in the Universe. Dividing equation (1) by m, we obtain

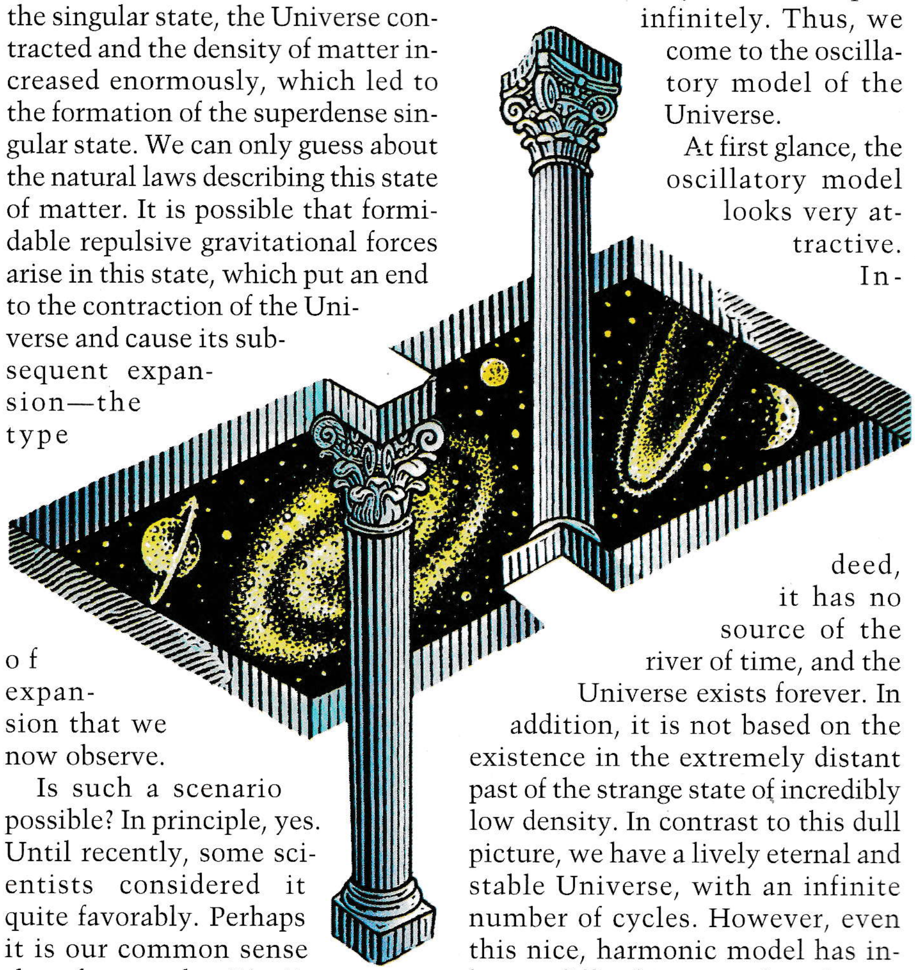
$$\frac{v^2}{2} - \frac{GM}{R} = \text{const}', \qquad (3)$$

where const′ = E/m. This equation describes the evolution of the distance R between any remote galaxies or protomatter particles from the time when no galaxies were yet in the sky. Therefore, equation (3) is the basic law of motion for matter in the Universe.

According to Einstein's theory, gravitation modifies the geometry of space, making it "curved." If const′ < 0 (that is, $E < 0$), the geometry is spherical, where parallel "straight lines" cross each other and space itself is closed, and thus has a finite volume. The curvature of space is described by the Universe's radius of curvature l. The entire volume V of the closed Universe is on the order of l^3. The radius l varies with time according to equation (3), as does any other distance in the Universe. Extrapolating the changes of R or l into the past, we conclude that the expansion formally began from a sizeless point of matter that had infinitely large density ρ, or in other words, from the singular state. We say "formally" to stress that, because some unknown physical laws probably play a major role at such a huge density ($\rho \sim 10^{94}$ g/cm^3 according to some estimates), such a conclusion is only a mathematical idealization. In particular, such conditions should produce the vacuum states of matter that generated the enormous repulsive forces necessary for the Big Bang.

What was before the Big Bang? Nobody knows for sure, but some possibilities are considered by scientists.

All that we have said makes it possible to "look" into the past and describe the evolution of the Universe in the following way. Before the singular state, the Universe contracted and the density of matter increased enormously, which led to the formation of the superdense singular state. We can only guess about the natural laws describing this state of matter. It is possible that formidable repulsive gravitational forces arise in this state, which put an end to the contraction of the Universe and cause its subsequent expansion—the type of expansion that we now observe.



Art by István Orosz

Is such a scenario possible? In principle, yes. Until recently, some scientists considered it quite favorably. Perhaps it is our common sense that favors the Big Bang model. Indeed, in this model time flows from negative infinity to positive infinity. Though the singular state is somewhat "unclear," the infinite river of time has neither source nor sink, a view that seems obvious or intuitive to most.

However, this simple model has an important defect. Indeed, it supposes that in the extremely distant past the Universe did contract from an infinitely spread out state. This state of near-zero density seems too "simple" and "primitive" for our splendid Universe.

Scientists tried to clear this "hurdle" in the following way. Assume that the constant value const′ in equation (3) is negative. This means that expansion of the Universe will be followed by contraction. If we assume that in its turn this contraction would be followed by a new expansion period after passing the singular state (we don't know whether this is true or not), then the entire cycle would repeat infinitely. Thus, we come to the oscillatory model of the Universe.

At first glance, the oscillatory model looks very attractive. Indeed, it has no source of the river of time, and the Universe exists forever. In addition, it is not based on the existence in the extremely distant past of the strange state of incredibly low density. In contrast to this dull picture, we have a lively eternal and stable Universe, with an infinite number of cycles. However, even this nice, harmonic model has inherent difficulties rooted in physics developed in the middle of the nineteenth century.

In 1850 the German physicist Rudolf Clausius and in 1851 the English physicist William Thomson

(Lord Kelvin) independently discovered the second law of thermodynamics. In the form given by Lord Kelvin, it proclaimed that for a system in a stable state, no process is possible whose sole result is the absorption of heat from a reservoir and the conversion of this heat into work. Thus, mechanical work cannot be performed exclusively by extracting thermal energy from a heat reservoir. In other words, one cannot convert the entire amount of thermal energy into mechanical energy. This means that in the end all forms of energy in an isolated system will be transformed into thermal energy, and this thermal energy will be uniformly distributed within the system, a state of thermal equilibrium.

In practice we know this law well. In any mechanical system there is friction, which transforms mechanical energy into thermal energy. It is true that in engines, the reverse process takes place, which converts thermal energy into mechanical work. However, this is possible only if the two heat reservoirs have different temperatures, otherwise the heat engine will not work. Energy is needed to maintain this temperature difference, and part of this energy is again transformed into thermal energy. Therefore, we have the continual process of converting all types of energy into thermal energy, which leads to the irreversible accumulation of thermal energy and the elimination of all kinds of energy except thermal energy. Clausius later formulated the second principle of thermodynamics mathematically.

Heat death

Thomson and Clausius understood the importance of the new law they discovered for the theory of the evolution of the Universe. Indeed, the entire Universe should be considered as an isolated system that does not exchange energy with any other hypothesized reservoirs. Therefore, all types of energy in the Universe ultimately should be transformed into thermal energy, which will be equally distributed. Although this process doesn't violate conservation of energy and the energy doesn't disappear, the energy becomes less valuable and cannot be turned into mechanical work. This means that macroscopic motion will cease to exist in the Universe. This unpleasant scenario is known as the "heat death" of the Universe.

However, the Universe we live in is far from being in the state of heat death! We can even think in metaphysical terms and suppose either that somebody or something is interfering with the evolution of the Universe and defends it from heat death or that the Universe is rather young and just has not had enough time to reach its thermal demise.

Now we'll see how this gloomy prognosis was refuted by science. The thermodynamic ideas of Clausius and Thomson were further developed by the Austrian physicist Ludwig Boltzmann. He revealed the physical meaning of the second law of thermodynamics. Thermal energy is the chaotic (stochastic) motion of atoms and molecules that make up matter. Therefore, the conversion of mechanical energy of some parts of a system into thermal energy means the transformation of organized (macroscopic) motion into chaotic motion—that is, an increase of chaos in the system. The same is true for the other forms of motion and energy. Due to the statistical laws, this increase of chaos is inevitable, provided the system is unaffected by any external organizing forces.

Boltzmann showed that chaos can be measured. The parameter that describes its value is entropy, which was introduced earlier by Clausius. The larger the chaos, the higher the entropy. The conversion of macroscopic motion into thermal energy is inevitably accompanied by an increase in entropy. When all forms of energy have changed into thermal energy and the thermal energy is uniformly spread throughout the system, the amount of chaos will not change any more, and the state of maximum entropy is achieved.

This statistical interpretation means that the second law of thermodynamics does not strictly hold all the time, and in principle, violation of this law is possible. Indeed, the law of increasing entropy says that the particles in an isolated system evolve into more and more probable states of the chaotic motion. However, some random deviations, or fluctuations, are possible in a statistical system.

For example, due to random collisions in a small volume, the atoms of a gas can stochastically acquire momentum in some direction. This means that the atoms will acquire translational motion in this direction, and thus macroscopic motion arises from thermal energy! Of course, such events are extremely rare. And such a special case of macroscopic fluctuation would be much rarer if we took a larger volume of gas.

As a rule, the entropy of an isolated system always increases, and the system evolves into the most probable state (one with maximal entropy), where it remains for an infinitely long time. However, some deviations from equilibrium may occur very rarely in one place or another in the system—though, as a rule, they will be rather small.

Theoretical resurrection

Boltzmann sought escape from the gloomy forecast of the heat death of the Universe as follows. The infinite Universe, he thought, exists in the most probable state of thermodynamic equilibrium with maximum entropy. However, rare fluctuations from this state are possible in any part of the Universe. It is true that a marked fluctuation in a large volume is an extremely rare event. Nevertheless, if we have infinite time for observation, we can wait until a large fluctuation occurs. According to Boltzmann, we live within just such a giant fluctuation.

Until the discoveries of Friedmann and Hubble, the Boltzmann fluctuation hypothesis was the single attempt to refute the heat death prognosis on the basis of nineteenth-century classical physics. These discoveries radically modified our views of the direction and final state of the evolution of the Universe. First of all, it was understood that gravitation plays the dominant role in development of the Universe. This major factor was entirely ignored in the theory of heat death, which was a mistake.

In the common reasoning on the conversion of all types of energy into thermal energy and the resulting dying out of all the processes in any isolated system, it was supposed that the total amount of energy in the system doesn't change. Why not? a reader may ask—the system is isolated, after all. Where does the extra energy come from to maintain the macroscopic motion in the system?

Of course, conservation of energy is an unshakable law. However, when applying it to the Universe, we must take into account the gravitational potential energy. This energy is of a particular kind—it is negative (see equation (1)). How is this fact reflected in the processes of the Universe?

Let's consider the following example: a spherical region of space filled with gaseous particles that interact with each other via gravitational forces. We'll assume that initially the gas is cold and scattered in space, the gravitational forces between the particles are extremely small, and the gravitational potential energy is virtually zero. However, weak gravity is not the same as zero gravity, and in due course gravity collects the scattered gas into a ball, which continually contracts under the influences of the forces of gravity.

Thus, the gas particles acquire more and more speed, and consequently, their kinetic energy increases. This positive energy component grows at the expense of the negative gravitational potential energy. Due to conservation of energy, its gravitational component decreases. This means that the absolute value of the gravitational potential energy grows (since $E_p < 0$). Thus, the gravity-induced compression of the system results in an increase of its positive (kinetic) energy. This fact was not taken into account in the earlier theories, which neglected gravitation. If positive energy can grow in an isolated system, the increase of entropy doesn't necessarily lead to fading of the macroscopic processes.

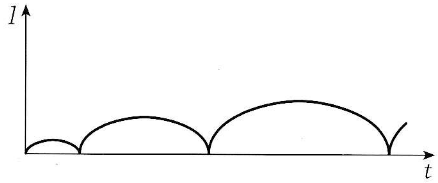


Figure 1. *According to the work of American physicist R. Tolman, this curve shows how the radius of the Universe varies with time.*

Therefore, the heat death theory in the form given in the nineteenth century, when the dynamic nature of the Universe was not known, was a mistake. Now we'll see how gravitation "works" in the oscillatory model of the Universe and how it disproves the heat death theory.

Paradoxically, amplitude increases

According to the second law of thermodynamics, which is assumed to be valid throughout the Universe, thermal energy and entropy accumulate in every cycle of the oscillatory Universe. For example, thermal energy accumulates on a large scale through the radiance of stars, which convert nuclear energy into radiant energy. We assume that entropy doesn't radically decrease in the singular state.

Thus, entropy grows from cycle to cycle. At first glance, it should lead to a decrease of the oscillation amplitude and finally to a standstill of the Universe. It seems that the life of the Universe should be similar to the damping of a pendulum, whose energy is gradually converted into thermal energy by friction in its suspension. In reality, this model predicts quite another phenomenon—an increase of the Universe's oscillation amplitude! Let's explain this.

Look at equation (3). In every cycle at the moment of maximum expansion of the Universe, when expansion turns into contraction, the velocity v of the matter composing the sphere becomes zero. Inserting $v = 0$ into equation (3), we get the condition for this moment:

$$\frac{GMm}{R_{max}} = -\text{const}'.$$

(Remember, in our model const' < 0). Now we insert the formula $M = (4/3)\pi R^3_{max}\rho_*$ into this expression, where ρ_* is the density of the matter at the moment of maximum expansion. This yields $\rho_* R^2_{max} = \text{const}''$. Finally, we recall that the Universe's radius of curvature l varies with time in the same way as the sphere's radius R (figure 1). Thus, for the radius of curvature l_{max} at the moment of maximum expansion of the Universe we have

$$\rho_* l^2_{max} = \text{const}'''. \qquad (4)$$

As we noted earlier, the volume V of the closed Universe is on the order of l^3. To an order of magnitude, at the moment of maximum expansion this volume is equal to l^3_{max}, and the total mass of matter in the Universe is then given by

$$\rho_* l^3_{max} = M. \qquad (5)$$

Dividing equation (5) by equation (4) we get:

$$l_{max} = \frac{M}{\text{const}'''}.$$

This formula tells us that l_{max} is proportional to the total mass of matter in the Universe. However, according to Einstein's principle of the equivalence of mass and energy, this mass is composed of the sum of the masses of the particles, the kinetic energy of their motions, and the energy of the photons. As the thermal energy constantly grows, the total mass of the Universe grows. This mass is proportional to l_{max}, so the amplitude of the Universe's oscillation will increase with time. Instead of fading, the oscillation's amplitude increases! This conclusion was reached in 1934 by the American physicist Richard Tolman.

Where is the energy necessary for the build-up of oscillations taken from? It is taken from the negative potential energy of the gravitational field (the total amount of energy doesn't vary according to conservation of energy).

Can our Universe be described by this model? Probably not. Indeed, although there is no similarity in this oscillatory model to the old concept of heat death, it predicts the constant growth of thermal energy and entropy in the Universe, which is at odds with current knowledge. Therefore, the number of oscillations should be limited. If so, the most attractive feature of the oscillatory Universe—the infinite time of its existence in the past—disappears.

Having shattered the myth of heat death, scientists encountered the no less enigmatic problem of the beginning of universal expansion. This problem is now the focus of attention of astronomers and cosmologists. Such is the path of science, in which one solution of nature's mystery inevitably leads to a more perplexing problem.

What to read in *Quantum* on cosmology and astrophysics:

Y. Zeldovich, "A Universe of Questions," January/February, 1992, 6–11.

Y. Solovyov, "The Universe Discovered," May/June 1992, 12–18.

W. A. Hiscock, "The Inevitability of Black Holes," March/April 1993, 26–29.

A. Sakharov, "Does Elementary Length Exist?," May/June 1997, 14–20.

Sir Isaac Newton

by David Arns

Under a spreading apple tree,
The village genius stands:
His mind conceives of wondrous things,
He writes them with his hands;
His fame goes forth to all the world—
He's known in many lands.

A tiny babe on Christmas Day
in 1642
Was born to Mrs. Newton
while outside, the cold winds blew.
And on the farm, through childhood,
precocious Isaac grew.

And after chores, he built devices
to see just how they worked,
To see what laws of nature
underneath the workings lurked.
(When people called them "toys," that's what
got Isaac *really* irked.)

His mother saw he was no farmer,
sent him off to school;
He quickly showed at Cambridge
that he was nobody's fool:
He began to bring to light the laws
that all of nature rule.

In one chapter in his story
(though apocryphal, it's said),
An apple, falling from a tree
impacted on his head,
Which drew his thoughts to gravity,
and we all know where *that* led.

He wondered if, by any chance,
the self-same gravitation
That pulls an apple to the ground,
affected all creation:
The Moon, the planets, and the Sun...
Thus went his cogitation.

He determined that the gravity
of Earth indeed controls
The orbit of the Moon, as 'round
the Earth it ever rolls.
Now, *describing* it mathematically
was one of Newton's goals.

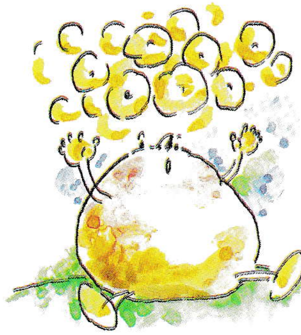
He discovered that the math you need
to show the laws of nature,
Surpassed the knowledge of that day;
the cosmos' legislature
Required new math, so Newton wrote
his "fluxions" nomenclature.

He talked of falling bodies
and his famous Laws of Motion,
And of colors seen in bubbles
and the tides upon the ocean.
And his crowning jewel, *Principia,*
created *great* commotion.

Yes, Newton's brilliant mind, it was
a trunk with many twigs—
His mind branched out in every way
(right through his powdered wigs).
His greatest contribution, though,
was cookies made from figs.

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.*

Just for the fun of it!



B226

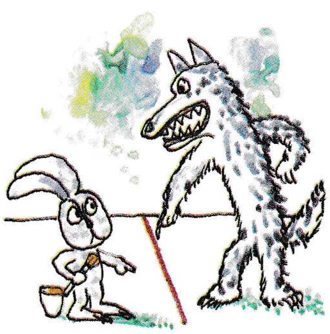
Doughnuts to dollars. Twenty-five doughnuts cost as many dollars as doughnuts can be bought for one dollar. What's the price of one doughnut?

B227

Carousel count. Thirty children ride a carousel swing. Every girl rides behind a boy, half of the boys ride behind a boy, and all the other boys ride behind girls. How many boys and girls are there?

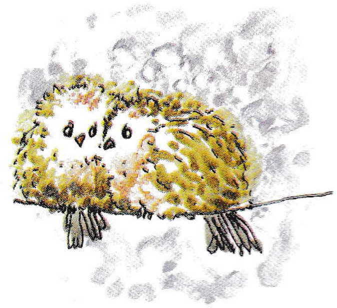
B228

Truancy treats. Boris came to school 35 minutes after the first class had started. So, he decided to go to the nearest shop and buy an ice cream cone. Unfortunately, when he came back, the second class had already begun. He immediately ran for another ice cream cone and spent as much time at the shop as before. When he returned the second time, he saw that there were 50 minutes left before the start of the fourth class. Does he have enough time to buy and eat a third ice cream cone if every class (including the break after it) takes 55 minutes?



B229

Quirky quadrilateral. Is there a quadrilateral that can be divided by one straight cut into two congruent parts, but neither its diagonals nor the segments connecting the midpoints of the opposite sides divide it into equal parts?



B230

Fluffy feathers. Why do birds fluff their feathers when the weather is very cold?

ANSWERS, HINTS & SOLUTIONS ON PAGE 52

Art by Pavel Chernusky

Numeral roamings

Exploring nontraditional mathematical operations

by A. Egorov and A. Kotova

MOST READERS WILL PROBably be surprised (or even troubled) by the following equalities:

$$2 \cdot 2 = 1,$$
$$2 \cdot 5 = 3,$$
$$3 \cdot 2 = 0.$$

These statements are perfectly correct, if we are using the multiplication sign and the symbol for equality to mean something different from what we were taught in school.

Arithmetic of numerals

Let's begin with an example. Consider the set of numerals 0, 1, …, 9. Let's say that the *sum* (or the *product*) of two numerals is the last digit of their sum (or product). Then

$$2 + 5 = 7,\ 7 + 6 = 3,\ 5 + 5 = 0;$$
$$7 \cdot 7 = 9,\ 2 \cdot 5 = 0,\ 8 \cdot 8 = 4.$$

Such arithmetical operations are by no means worse than the usual addition and multiplication of integers to which we are accustomed. Indeed, for any set of three numerals a, b, and c, the following identities hold:

1. $a + b = b + a$;
2. $(a + b) + c = a + (b + c)$;
3. $a + 0 = a$;
4. For every numeral a, there exists a numeral $(-a)$ such that $a + (-a) = 0$ (for example, $-4 = 6$, $-5 = 5$, $-1 = 9$);
5. $ab = ba$;
6. $a \cdot 1 = a$;
7. $a(b + c) = ab + ac$.

	0	1	2	3	4	5	6	7	8	9
0	0	1	2	3	4	5	6	7	8	9
1	1	2	3	4	5	6	7	8	9	0
2	2	3	4	5	6	7	8	9	0	1
3	3	4	5	6	7	8	9	0	1	2
4	4	5	6	7	8	9	0	1	2	3
5	5	6	7	8	9	0	1	2	3	4
6	6	7	8	9	0	1	2	3	4	5
7	7	0	9	0	1	2	3	4	5	6
8	8	9	0	1	2	3	4	5	6	7
9	9	0	1	2	3	4	5	6	7	8

	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9
2	0	2	4	6	8	0	2	4	6	8
3	0	3	6	9	2	5	8	1	4	7
4	0	4	8	2	6	0	4	8	2	6
5	0	5	0	5	0	5	0	5	0	5
6	0	6	2	8	4	0	6	2	8	4
7	0	7	4	1	8	5	2	9	6	3
8	0	8	6	4	2	0	8	6	4	2
9	0	9	8	7	6	5	4	3	2	1

Figure 1

The truth of these identities follows directly from the analogous properties of the operations with ordinary numbers. One can subtract numerals as well as add them, if we make the definition:

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

For instance,

$$2 - 7 = 2 + (-7) = 2 + 3 = 5,$$
$$4 - 6 = 4 + (-6) = 4 + 4 = 8,$$

and so on.

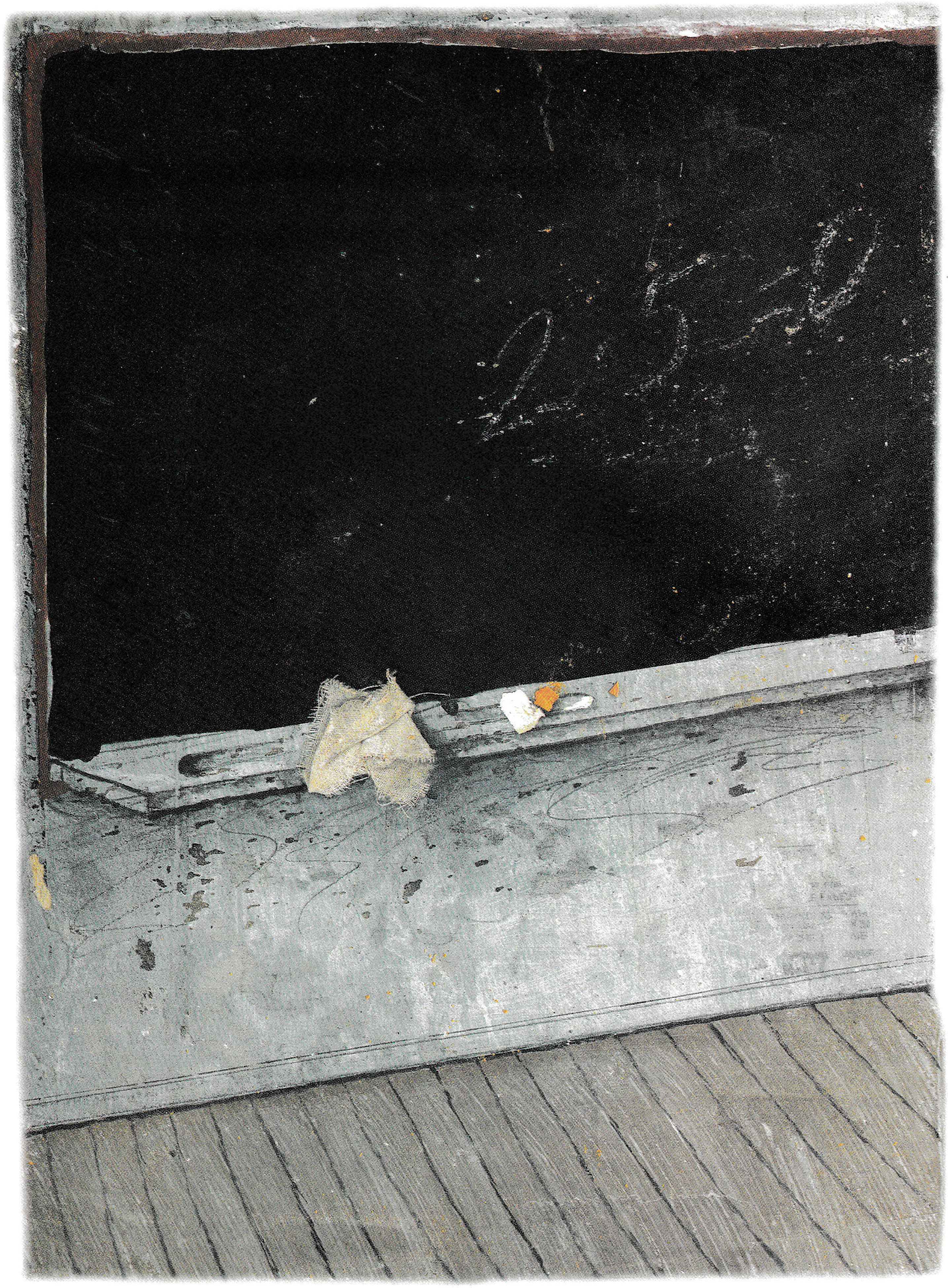
We can form an addition table and a multiplication table of our numeral arithmetic (fig. 1). A number at the intersection of a row and a column is the sum (or the product) of the numbers at the head of the row and the column. Readers will undoubtedly note the striking difference between the "new" arithmetic and the traditional one. The product of two nonzero numerals can equal zero!

When this happens, we say that the arithmetic contains "zero divisors"—numbers $a \neq 0$ and $b \neq 0$ such that $ab = 0$.

Exercises

1. Find the last digit in the number: (a) 7^{1993}; (b) $2^{7^{1993}}$; (c) 3^{1993}.

Art by Yury Vashchenko



2. Prove that the product of the two last digits of the square of an integer is even.

3. Solve the equation $x^2 - 1 = 0$ in numeral arithmetic.

Arithmetic of remainders modulo *m*

The units digit of an integer is just its remainder upon division by 10. So our numeral arithmetic is really the arithmetic of remainders when numbers are divided by 10. This leads to a natural generalization of our numeral arithmetic.

Let $m > 1$ be an arbitrary natural number. Any integer, when divided by m, leaves some remainder. There are m different remainders:

$$0, 1, 2, \ldots, m - 1.$$

Note that the remainder from division by m coincides with the last numeral in the m-based notation of the number divided.

We now introduce the addition and multiplication of remainders upon division by m. We'll call the *sum* of two remainders a and b the remainder of the division by m of the usual sum of the numbers a and b, with a similar definition for products.

It's not hard to see that all seven previously mentioned properties of addition and multiplication of numerals hold for the operations with remainders. Of course, in the arithmetic of remainders, we can subtract, too. *Subtract* is defined, as before, by the rule $a - b = a + (-b)$.

The arithmetic of remainders of division by m (modulo m) is traditionally denoted by $\mathbf{Z}_m$.

Figure 2 shows multiplication tables for $\mathbf{Z}_5$ and $\mathbf{Z}_6$ (we do not show the corresponding addition tables since their properties are similar for all moduli). We see that $\mathbf{Z}_5$ does not contain zero divisors yet $\mathbf{Z}_6$ does. What's going on? The following exercises will help illustrate this phenomenon.

	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Figure 2

Exercises

4. Draw multiplication tables for the arithmetics of remainders modulo 7, 8, 9, 11, 12, and 13.

5. Solve the equations $x^2 = 1$ and $x^2 = -1$, when x belongs to one of the modular systems listed in exercise 4. Use this result to find all integers x for which $x^2 - 1$ or $x^2 + 1$ is divisible by the corresponding modulus.

6. Consider the arithmetic of remainders modulo 100. Find all zero divisors in it. How many are there? Prove that if a is not a zero divisor, then $a^{40} = 1$. (One corollary of this fact is that for any integer a coprime with 100, $a^{40} - 1$ is divisible by 100.)

7. Prove that the sum

$$1 + 2^{1997} + \ldots + 1996^{1997}$$

is divisible by 1997.

Prime moduli

Let p be an arbitrary prime number. Consider the remainders upon division by p:

$$0, 1, 2, \ldots, p - 1.$$

Theorem 1. The set of nonzero remainders modulo p (where p is prime) contains no zero divisors.

Proof. Suppose there exist two remainders a and b modulo p such that $a \neq 0$, $b \neq 0$, and $ab = 0$ in $\mathbf{Z}_p$.

This means that the number ab is divisible by p. Since p is prime, this means that one of the numbers a or b is divisible by p, which is impossible, since $0 < a < p$ and $0 < b < p$.

In the following discussion we denote the set of all nonzero remainders in $\mathbf{Z}_p$ by $\hat{\mathbf{Z}}_p$.

It follows directly from theorem 1 that if $ab = ac$ and $a \neq 0$, then $b = c$ (the reader is invited to check this). But this means that all elements of $\hat{\mathbf{Z}}_p$ of the form

$$a, 2a, \ldots, (p - 1)a$$

are different, and thus exactly one of them is equal to 1. This means that $\hat{\mathbf{Z}}_p$ contains some remainder b such that $ab = 1$.

We denote this unique element by a^{-1}, or $1/a$, and call this element of $\mathbf{Z}_p^*$ the *multiplicative inverse* of a. It's clear that if $a^{-1} = b^{-1}$ for two remainders a and b, then $a = b$.

Thus, if p is prime, we can define division in $\mathbf{Z}_p$: for $b \neq 0$, a/b is defined as $a \cdot b^{-1}$.

Exercises

8. Prove that

(a) $(a^{-1})^{-1} = a$;

(b) $(-a^{-1}) = -a^{-1}$;

(c) $(ab)^{-1} = a^{-1}b^{-1}$;

(d) every equation $ax = b$ in which $a \neq 0$ has a unique root.

9. Prove that for any prime p, the number

$$(p - 1)!(1 + 1/2 + 1/3 + \ldots + 1/(p - 1))$$

is divisible by p.

10. Prove that a remainder from $\mathbf{Z}_m$ is invertible (has an inverse element) if and only if it is not a zero divisor (that is, if it is coprime with m).

Wilson's theorem

The arithmetic of remainders modulo a prime number allows one to prove the following criterion of primality of an integer p.

Wilson's theorem. An integer number p is prime if and only if the number $A = (p - 1)! + 1$ is divisible by p.

Proof. Let p be a prime number. Let's show that in this case A is divisible by p. When $p = 2$ this statement is evident. If $p > 2$, let's note that for every nonzero remainder modulo p, its inverse is defined, and that remainders a and a^{-1} are different if $a \neq 1$ and $a \neq p - 1$. Indeed, if $a = a^{-1}$, then

$$0 = 1 - 1 = a \cdot a^{-1} - 1 = a^2 - 1$$
$$= (a - 1)(a + 1).$$

But this can only hold true if $a = 1$ and $a = p - 1$, since otherwise neither $a + 1$ nor $a - 1$ is divisible by p.

Thus we can combine all remainders in the product $1 \cdot 2 \cdot \ldots \cdot (p-1)$, except 1 and $p-1$ into pairs consisting of the remainder and its inverse. And therefore,

$$1 \cdot 2 \cdot \ldots \cdot (p-1) = 1 \cdot [2 \cdot 2^{-1}] \cdot [3 \cdot 3^{-1}] \cdot \ldots \cdot [((p-1)/2) \cdot ((p-1)/2)^{-1}] \cdot (p-1) = 1 \cdot (p-1) = p-1 = -1.$$

So, in $\mathbf{Z}_p$ the equality $(p-1)! = -1$ holds, which means that the integer number $(p-1)! + 1$ is divisible by p.

Now let p be a composite number. That is, suppose $p = kd$ for integers k and d. Then one of the factors in $(p-1)!$ is d, and $(p-1)!$ can be written in the form nd, for some integer n. However, then $(p-1)! + 1 = nd + 1$ cannot be a multiple of d, and thus is not divisible by p.

Unfortunately, this criterion for primality is a nice mathematical fact but hardly a convenient one: It is very difficult to check this criterion when p is not small. For example, in order to check that the number 1997 is prime, we would have to calculate the huge number $1996! + 1$ and divide it by 1997.

Exercises

11. Prove that the equation $x^2 + 1 = 0$ has roots in $\mathbf{Z}_p$ if $p = 4k + 1$.

12. Prove that the numbers
(a) $91! \cdot 1901! - 1$
(b) $92! \cdot 1990! + 1$
are divisible by 1993.

Periodicity of powers

Let $a \neq 0$ be an element of $\hat{\mathbf{Z}}_p$, where p is prime number. What can we say about $a^2, a^3, a^4, \ldots$?

Since there are only $p-1$ different elements in $\mathbf{Z}_p^*$, we conclude that one can find two equal elements among the first p terms of the sequence $a, a^2, \ldots, a^m, \ldots$. Let them be a^k and a^l. Then $a^{k-l} = 1$.

Thus, one of the powers of a is equal to 1.

Let $a \neq 1$ and let d be the least natural number such that $a^d = 1$. Clearly, $d \neq 1$; all the powers $a, a^2, \ldots, a^d$ are different; and $a^{k+d} = a^k$ for all integers k. That is, the sequence of powers of the remainder a is periodic (with period d).

Definition. We call the number d defined in the preceding paragraph *the order of* a *modulo* p, which is denoted by $d_p(a)$.

Let's point out some important properties of the order of an element.

1. If $d_p(a) = d$ and $a^m = 1$, then m is divisible by d;
2. If $d_p(a) = m$ and $m = kl$, $d_p(a^k) = l$;
3. $d_p(a) = d_p(a^{-1})$;
4. If $d_p(a) = m$ and $d_p(b) = n$, then $d_p(ab)$ divides the least common multiple of the numbers m and n;
5. If $d_p(a) = m$, $d_p(b) = n$, and the numbers m and n are coprime, then $d_p(ab) = mn$.

Exercises

13. Prove properties 2 and 3.

14. Prove property 4 and check that $d_p(ab)$ is not necessarily equal to the least common multiple of m and n.

Let's prove property 1. Suppose that $a^m = 1$ and m were not divisible by d (that is, $m = qd + r$, where $0 < r < d$). Then

$$1 = a^m = a^{qd+r} = a^{qd}a^r = (a^d)^q a^r = a^r.$$

By definition, d is the least power such that $a^d = 1$. But we've found a number r less than d and satisfying the same condition. Thus, we have a contradiction.

Here is an explicit proof of property 5. Let $d_p(ab) = d$. We can see that d does not exceed mn since

$$(ab)^{mn} = (a^m)^n (b^n)^m = 1 \cdot 1 = 1.$$

On the other hand, since $(ab)^d = 1$, we see that $a^d = (b^d)^{-1} = (b^{-1})^d$, and thus (see property 3) $a^{nd} = (b^{-1})^{nd} = 1$. Therefore, nd is divisible by m (property 1). But m and n are coprime, so m divides d. Similarly, we show that n divides d as well; that is, d is divisible by mn. But d is not greater than mn, so it is equal to mn.

We also call the order of a in $\mathbf{Z}_p$ the *exponent* to which the number a belongs modulo p.

Exercises

15. Find the order of the remainder 2 in $\mathbf{Z}_7$, $\mathbf{Z}_{11}$, $\mathbf{Z}_{13}$.

16. Prove that
(a) if $A = a^4 + b^4 + c^4 + d^4$ is divisible by 5, then it is divisible by 625, too.
(b) if $a^3 + b^3 + c^3$ is divisible by 7, then abc is divisible by 7 as well.
(c) if $a^2 + b^2$ is divisible by 7, then both a and b are divisible by 7.

17. Prove that for all natural n, $5^{2n+1} + 3^{n+2} \cdot 2^{n-1}$ is divisible by 19.

Euler's function

Let m be a natural number, not necessarily prime. Consider the set of all remainders modulo m coprime to m. (We use the symbol $\hat{\mathbf{Z}}_p$ for this set). Denote the number of elements in $\hat{\mathbf{Z}}_p$ by $\phi(m)$.

The function $\phi(m)$, which associates every natural number m with the number of natural numbers less than m and coprime with it, is called *Euler's function*. In particular, for every prime p

$$\phi(p) = p - 1;$$
$$\phi(p^2) = p(p-1);$$
$$\ldots$$
$$\phi(p^n) = p^{n-1}(p-1).$$

Theorem 2. For all coprime integers m and n,

$$\phi(mn) = \phi(m) \cdot \phi(n).$$

Because of this property, we say that Euler's function is *multiplicative*).

Proof. We can count the amount of natural numbers that are less than mn and coprime with it in the following way. Consider the table consisting of the numbers from 1 through mn, shown in figure 3. Every row of this table contains exactly $\phi(n)$ numbers coprime with n, and every column contains exactly $\phi(m)$ numbers coprime with m. Moreover, the position of the numbers coprime with n in a row does not depend on the choice of row. (We invite the reader to find why this happens.)

Since m and n are coprime, any number that appears in the table is coprime with mn if and only if it is coprime both with m and with n. Thus, it stands in a column consisting of the numbers coprime with n. But each column contains exactly $\phi(m)$ numbers coprime with m. Thus, there are $\phi(m)\phi(n)$ numbers of this sort. On the other hand, by defi-

1	2	3	...	n
$n+1$	$n+1$	$n+1$	...	$2n$
$2n+1$	$2n+1$	$2n+1$	...	$3n$
...	...	...	...	...
$(m-1)n+1$	$(m-1)n+1$	$(m-1)n+1$	...	mn

Figure 3

nition, $\phi(mn)$ is equal to this quantity.

So, $\phi(mn) = \phi(m) \cdot \phi(n)$.

Now we can obtain the formula for $\phi(m)$. Let's represent m in the form of a product of distinct primes:

$$m = p_1^{k_1} p_2^{k_2} \dots p_s^{k_s}.$$

Then

$$\phi(m) = \phi\left(p_1^{k_1} p_2^{k_2} \dots p_s^{k_s}\right)$$
$$= \phi\left(p_1^{k_1}\right) \cdot \dots \cdot \phi\left(p_s^{k_s}\right).$$

But $p_1, p_2, \dots, p_s$ are prime numbers, and thus

$$\phi(m) = p_1^{k_1-1}(p_1-1)p_2^{k_2-1}(p_2-1)\dots p_s^{k_s-1}(p_s-1)$$
$$= p_1^{k_1}\left(1-\frac{1}{p_1}\right)p_2^{k_2}\left(1-\frac{1}{p_2}\right)\dots p_s^{k_s}\left(1-\frac{1}{p_s}\right)$$
$$= m\left(1-\frac{1}{p_1}\right)\left(1-\frac{1}{p_2}\right)\dots\left(1-\frac{1}{p_s}\right).$$

For example,

$$\phi(20) = 20(1 - 1/2)(1 - 1/5) = 8.$$

Exercises

18. Prove that $\phi(n)$ is even when $n \neq 2$.

19. Find $d_{210}(11)$.

Another wonderful property of Euler's function is described by the following theorem:

Euler's theorem. If the number a is coprime with a natural number m, then $a^{\phi(m)} - 1$ is divisible by m.

Proof. Let's write down all the elements of $\hat{\mathbf{Z}}_m$ in a row:

$$1 = a_1, a_2, \dots, a_{\phi(m)}.$$

Since a is coprime with m, the remainder $\overline{a}$ when a is divided by m is not a zero divisor (that is, $\overline{a} \in \hat{\mathbf{Z}}_m$). So, if we show that $\overline{a}^{\phi(m)} = 1$ in $\hat{\mathbf{Z}}_m$, we are done.

To this end, we multiply each element of $\hat{\mathbf{Z}}_m$ by $\overline{a}$ to obtain a new row:

$$\overline{a}, \overline{a}a_1, \overline{a}a_2, \dots, \overline{a}a_{\phi(m)}.$$

All remainders in this row are different, and thus we've again obtained a complete set of elements of $\hat{\mathbf{Z}}_m$, although they are in some other order. Let's multiply these elements together:

$$a_1 \cdot a_2 \cdot \dots \cdot a_{\phi(m)}$$
$$= \overline{a}^{\phi(m)} \cdot a_1 \cdot a_2 \cdot \dots \cdot a_{\phi(m)}.$$

But the product $a_1 \cdot a_2 \cdot \dots \cdot a_{\phi(m)}$ is coprime with m, and therefore this equality means that $\overline{a}^{\phi(m)} = 1$.

One simple corollary of Euler's theorem is Fermat's "little" theorem: If p is a prime number and a is not divisible by p, then $a^{p-1} - 1$ is divisible by p. (We obtain this statement from Euler's theorem immediately, since $\phi(p) = p - 1$.)

Note: It follows from Fermat's "little" theorem and property 1 of the degrees that $p - 1$ is divisible by $d_p(a)$ for any nonzero remainder a.

Exercises

20. Prove that

(a) $2^{131} - 1$ is divisible by 263;

(b) $2^{3^n} + 1$ is divisible by 3^{n+1} and is not divisible by 3^{n+2}.

21. Prove that if a prime number p divides $x^2 + 1$ (where x is an integer greater than 1), then $p = 4k + 1$.

22. Prove that there are infinitely many prime numbers of the form $p = 4k + 1$.

23. A rational fraction p/q (where q is coprime with 10 and $q > 10$) is written in the form of an infinite periodic decimal fraction with period m. Prove that m divides $\phi(q)$.

QUANTUM

in every library and college bookstore!

That's our goal. So if you don't find *Quantum* where you expect to, **ask for it!** *Quantum* is a resource that belongs in every library and bookstore.

"A first-class 'new' magazine . . . one can appreciate the meaning of quality and imaginative challenge . . . it is for anyone with an interest in science, particularly math and physics. Highly recommended."—***Library Journal***

". . . full of stimulating articles . . . excellent mathematical reading for students at school and university."—***Nature***

"*Quantum*'s entry into the world of teens (and older folk, too) should be embraced and welcomed. Its relatively low subscription price makes it a bargain for the wealth of knowledge and recreational materials its readers receive."—***Journal of Negro Education***

"Translations are in excellent and easily understandable English; English-language articles are similarly well written. This wonderful magazine should be in every secondary school library and in college and public libraries as well."—***Magazines for Libraries***

"It should be in every high school library [and] in most public libraries . . . we owe it to our students to make *Quantum* widely available."—**Richard Askey, Professor of Mathematics at the University of Wisconsin, Madison**

Share the QUANTUM experience!

To order, call 1 800 SPRINGER (1 800 777-4643)

HOW DO YOU FIGURE?

Challenges in physics and math

Math

M226

Regal route. A king stands in the lower left corner of a 6×6 chessboard. In a move it can go either one square up, one square to the right, or one square up and one square to the right. How many different ways are there for the king to come to the upper right corner of the board?

M227

Murder by numbers. Each of infinitely many gangsters has a contract out on exactly one of the others. Prove that there is some infinite subset of these gangsters, none of whose members has a contract out on another.

M228

Planar figure. Find the area of the figure on the plane determined by the inequality

$$(y^3 - \arcsin x)(x^3 + \arcsin y) \geq 0,$$

where x and y are standard coordinates on the plane.

M229

Systematic mathematics. Solve the following system of equations:

$$\begin{cases} x + y + z + t = 6 \\ \sqrt{1-x^2} + \sqrt{4-y^2} \\ \qquad + \sqrt{9-z^2} + \sqrt{16-t^2} = 8. \end{cases}$$

M230

Segment ratio. In triangle ABC, $\angle B \neq 90°$, and $AB{:}BC = k$. Let M be the midpoint of AC. Lines symmetric to BM with respect to AB and BC meet line AC at point D and E respectively. Find the ratio $BD{:}BE$.

Physics

P226

Crawly critters. A cockroach and two beetles crawl on a large horizontal table. Each beetle can crawl at a speed of up to $v = 1$ cm/s. At the initial moment the insects are at the vertices of an equilateral triangle. What must the cockroach's maximum velocity be in order to preserve the equilateral shape of the triangle for any movements of the beetles? (A. Korshkov)

P227

Ball in a chute. There is no slipping when a ball rolls with velocity v along the rib of a right angle chute ACB (fig. 1). The distance AB is equal to the ball's radius. What points of the ball have the maximum velocity? (S. Krotov)

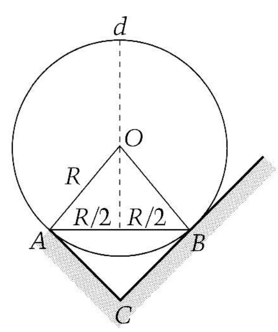


Figure 1

P228

Evaporating drop. In the nineteenth century, the Russian scientist B. I. Sreznevsky studied the evaporation of liquid drops in air. Assume such evaporation proceeds at constant temperature difference due to heat transfer from the surroundings. Find the dependence of the drop's radius on time assuming the thermal flow per unit area of a spherical drop is directly proportional to the temperature difference and inversely proportional to the drop's radius. What time is needed to completely evaporate a drop whose radius decreases by half in 10 minutes? (A. Stasenko)

P229

Lamp in an electric circuit. A lamp in an electric circuit (fig. 2) glows with the same brightness whether or not switch S is open or closed. The parameters of the circuit are $R_1 = R_3 = 90\ \Omega$, $R_2 = 180\ \Omega$, and $V = 54$ V. Find the voltage drop across the lamp. (V. Chivilev)

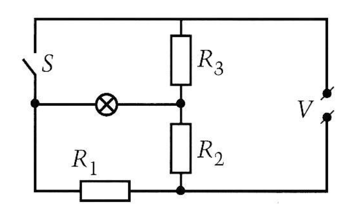


Figure 2

P230

Atom emits a quantum. An excited hydrogen atom radiates light. Find the change in the wavelength of the light due to the recoil of the nucleus caused by an emitted quantum of light. (V. Mozhayev)

ANSWERS, HINTS & SOLUTIONS ON PAGE 49

Does a falling pencil levitate?

When the normal becomes abnormal

by Leaf Turner and Jane L. Pratt

Although a pencil is about the simplest mechanical system that one can imagine, it can exhibit rich, intricate physics. How many times have you put the point of your pencil down on your desk and let the pencil fall from a vertical position? No doubt the fall was over too quickly for you to notice whether the point ever left the surface of the desk. Can such "dynamical levitation" occur?

Before addressing this question, we recount an apocryphal story of another piece of physics engendered by a falling pencil. Fred Hoyle had the frequent experience of dropping something (in his case, a pencil) and then not being able to find it. He whimsically suggested that it was imagining the time reversal of this experience that led him to his hypothesis of the continuous creation of matter! However, our falling pencil will not have such momentous consequences.

Static friction

We'll think of the pencil as an infinitely thin rod whose total mass m is uniformly distributed along its length L. Its center of gravity CG is then at its midpoint. When the pen-

Art by Jose Garcia

cil is placed vertically on a horizontal surface, it starts to fall, initially without sliding and eventually with sliding. But does it slide in the same direction as the horizontal component of velocity of its CG, or does it slide in the opposite direction? Does the point of the pencil ever rise off the table (levitate)? The only free parameters at our disposal are the coefficients of static and kinetic friction, μ_s and μ_k, respectively. Indeed, were there no friction, there would be no horizontal force on the pencil, and the pencil's CG would fall straight down.

As the pencil starts to fall, static friction keeps the pencil from sliding. If θ is the angle the pencil makes with the horizontal and if ω and α are the angular velocity and angular acceleration of the pencil, then we can relate the vertical and horizontal components of the acceleration of the CG prior to the onset of sliding to α, ω, and θ. The CG receives two contributions to its acceleration: the angular $(-L\alpha/2)$ and the centripetal $(L\omega^2/2)$ as shown in figure 1. If we let $\parallel$ and $\perp$ represent the directions parallel and perpendicular to the table, an examination of figure 1 shows that

$$a_\parallel = -\frac{L\omega^2}{2}\cos\theta - \frac{L\alpha}{2}\sin\theta, \quad (1a)$$

$$a_\perp = -\frac{L\omega^2}{2}\sin\theta + \frac{L\alpha}{2}\cos\theta. \quad (1b)$$

(Note that α is a negative quantity.)

Since no energy is dissipated prior to the onset of sliding, we can use conservation of energy to find how ω^2 depends on θ. Since the pencil is rotating about its point, its kinetic energy is $I\omega^2/2$, where $I = mL^2/3$. Recalling that its CG is a distance $L/2$ from the pencil point, we can see that the pencil's potential energy is $(mgL/2)\sin\theta$. If the pencil falls from an essentially stationary vertical position, its total mechanical energy must be $mgL/2$. We thus find that

$$\frac{1}{6}mL^2\omega^2 + \frac{mgL}{2}\sin\theta = \frac{mgL}{2},$$

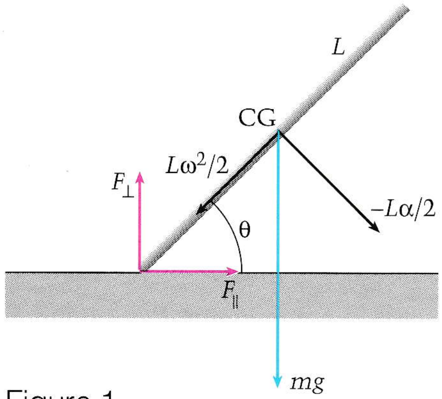


Figure 1

or, equivalently,

$$\omega^2 = \frac{3g}{L}(1-\sin\theta). \quad (2)$$

We can obtain the angular acceleration α readily by taking torques about the (fixed) contact point. The only force not passing through the contact point is that of gravity. Thus, we find that

$$\tau = -I\alpha = -\frac{mL^2}{3}\alpha = \frac{mgL}{2}\cos\theta,$$

which specifies α as a function of θ:

$$\alpha = -\frac{3g}{2L}\cos\theta. \quad (3)$$

Equations 1a, 1b, 2, and 3 now determine $a_\parallel$ and $a_\perp$ as functions of θ.

We now evaluate the normal force $F_\perp(\theta)$ exerted by the table on the pencil when the pencil is inclined to the horizontal by θ. Newton's second law tells us that the sum of the vertical components of all forces is $ma_\perp$; namely,

$$F_\perp - mg = ma_\perp.$$

Inserting our expressions for ω^2 and α (equations 2 and 3) into equation 1b, we can find the value of $F_\perp$ at any θ:

$$F_\perp(\theta) = \frac{mg}{4}(1-3\sin\theta)^2.$$

We note that $F_\perp$ is zero when θ equals $\sin^{-1}(1/3)$, or about 19.5°.

Newton's second law also allows us to infer the actual horizontal frictional force, which is parallel to the surface of the table. Since the horizontal acceleration $a_\parallel$ is specified by equation 1a and since the only horizontal force inducing this acceleration is the friction exerted by the table on the pencil at their point of contact, $F_\parallel(\theta)$, we find that

$$F_\parallel = ma_\parallel,$$

so that

$$F_\parallel(\theta) = \frac{3mg\cos\theta}{2}\left(\frac{3}{2}\sin\theta - 1\right), \quad (4)$$

which we now obtain by inserting our expressions for ω^2 and α into equation 1a. Slipping starts at the critical angle θ_c when the magnitude of the horizontal force has become equal to the maximum force that static friction can exert: $\mu_s F_\perp$. The condition for this critical angle at the onset of sliding is thus given by

$$\mu_s = \mu(\theta_c),$$

where $\mu(\theta)$ merely represents the ratio of the magnitude of the parallel force $F_\parallel$ to the normal force $F_\perp$ at an angle θ—that is,

$$\mu(\theta) \equiv \frac{|F_\parallel(\theta)|}{F_\perp(\theta)} = \frac{|3\cos\theta(3\sin\theta - 2)|}{(1-3\sin\theta)^2}.$$

This expression is plotted in figure 2 as a function of θ in degrees.

Let's explore and analyze figure 2. As we consider pencils having progressively increasing values of μ_s starting from zero, the corresponding values of θ_c decrease from 90° to the angle θ_R at which the force ratio μ attains a relative maximum. Using the expression for μ above, we can use elementary calculus to find the right-hand extremity θ_R of the domain associated with the red region of figure 2. We evaluate $d\mu(\theta)/d\theta$, set the result equal to zero, and solve for θ. The solution yields $\theta_R = \sin^{-1}(9/11) \cong 54.9°$. At this value of θ_c,

$$\mu_s = \mu\left[\sin^{-1}\left(\frac{9}{11}\right)\right] = \frac{15\sqrt{10}}{128} \cong 0.371.$$

Each of the values of μ associated with the red region has already been attained on the domain $54.9° < \theta < 90°$. If a pencil possessed a μ_s equal to one

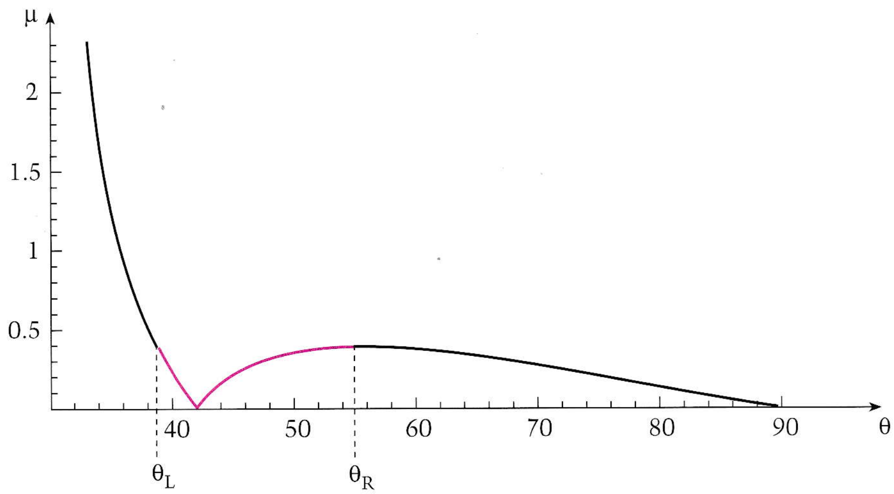


Figure 2

of these values, it would have already started sliding at an angle greater than 54.9°. When μ_s is slightly greater than $\mu(\theta_R)$, the red region is bypassed. Thus, a pencil cannot start sliding in the red region. Setting the numerator of $\mu(\theta)$ equal to zero, we observe that $\mu(\theta)$ equals 0 when $\theta = \sin^{-1}(2/3) \cong 41.8°$. We can use a programmable calculator to find the left-hand extremity θ_L of the red region. We merely solve for the θ value θ_L that satisfies $\mu(\theta_L) = 15\sqrt{10}/128$ on the interval, say, between 35° and 40° to find that $\theta_L \cong 38.8°$. Since the denominator of $\mu(\theta)$ approaches zero as θ approaches $\sin^{-1}(1/3) \cong 19.5°$, the left-hand portion of the curve has the line $\theta = \sin^{-1}(1/3)$ as its asymptote. Thus sliding will always start before a pencil ever reaches 19.5° from the horizontal!

Kinetic friction

But in which direction will the pencil point slide? If μ_s is less than about 0.371, the sliding will start at an angle greater than about 54.9°. Using equation (4), we observe that $F_\parallel$ will be positive. Conversely, if μ_s is greater than about 0.371, the sliding will start at some angle between about 38.8° and 19.5°. On this domain $F_\parallel$ is negative. But this can happen only if, in the first case, the pencil point is sliding in the direction opposite the motion of the CG, and, in the second case, in the direction of horizontal motion of the CG! The direction of slide depends on the coefficient of static friction! A small coefficient leads to a backward slide, and a large coefficient leads to a forward slide. This is plausible because, for the larger coefficient, the pencil will have had a chance to build up some significant horizontal momentum before sliding. But observe that the pencil never leaves the horizontal table under the influence of static friction.

Can it leave the table after the onset of sliding? Obtaining the answer to this question requires significantly more subtlety, both mathematically and physically because total mechanical energy is no longer conserved and all points of the pencil are free to accelerate!

When all points of an object are free to accelerate, to find α, the angular acceleration, it is expedient to take torques using the CG as origin. From figure 1, we see that

$$\tau_{cm} = F_\perp^k \frac{L}{2}\cos\theta - F_\parallel^k \frac{L}{2}\sin\theta = -I_{cm}\alpha. \quad (5)$$

The superscript k denotes the kinetic friction case. We observe that the frictional force is given by $F_\parallel^k = \pm\mu_k F_\perp^k$, where the sign is chosen to be in accordance with the preceding discussion. Using the fact that $I_{cm} = mL^2/12$ and the torque equation, the angular acceleration can be calculated in terms of θ and $F_\perp{}^k$:

$$\alpha = -6\frac{F_\perp^k}{mL}(\cos\theta \mp \mu_k \sin\theta). \quad (6)$$

If we insert this expression for α into equation 1b, we can find $F_\perp{}^k$ from Newton's second law: $F_\perp{}^k - mg = ma_\perp$. (Do you see that equation 1b, but not 1a, remains valid in the sliding context?) When the pencil is sliding,

$$F_\perp^k = m\left[\frac{g - \dfrac{L\omega^2}{2}\sin\theta}{1 + 3\cos^2\theta(1 \mp \mu_k \tan\theta)}\right].$$

If we next insert this result for $F_\perp{}^k$ right back into our expression for α, equation 6, we obtain

$$\alpha = -\frac{6\cos\theta}{L}\left[\frac{g - \dfrac{L\omega^2}{2}\sin\theta}{3\cos^2\theta + (1 \mp \mu_k \tan\theta)^{-1}}\right]. \quad (7)$$

This equation possesses the needed information about levitation of the bottom of the pencil! When we decipher its message, we'll obtain the information for all cases without doing any numerical computations whatsoever!

First we look at the message borne by the numerator within the square brackets, $g - (L\omega^2/2)\sin\theta$. Suppose that at a certain instant of time while the pencil is sliding, the table vanished. At that instant, the CG would accelerate straight downward due to gravity. When the table vanished, there no longer would be any torque about the CG (because the table exerted both a normal and a frictional force on the pencil's point), and the angular acceleration of the pencil would vanish. The point of the pencil would thus rotate uniformly with angular velocity ω about the CG. As a result, it would have only a centripetal acceleration $L\omega^2/2$ toward the CG. Note from figure 1 that the pencil is instantaneously tilted from the horizontal by the angle θ. The net acceleration of the pencil point with respect to the table equals the sum of the acceleration of the pencil point with respect to the CG and the acceleration of the CG with respect to the table. Thus, the pencil point would have an acceleration downward equal to $g - (L\omega^2/2)\sin\theta$, which is that numerator! Let us refer to this numerator as a "virtual" acceleration and denote it as a_v. Since θ and ω are continuous functions of time, we see that a_v is positive at the onset of sliding. (We merely insert ω^2 obtained from equation 2, valid until the onset of sliding, into the numerator to verify that a_v equals

$$g\left[\frac{3}{2}\left(\sin\theta - \frac{1}{2}\right)^2 + \frac{5}{8}\right]$$

and is thus positive when sliding starts.) Once the pencil starts sliding, it will levitate only if a_v changes sign. Let's find out what equation 7 is trying to tell us about that sign.

The rotational manifestation of the work-energy theorem is that the effect of a torque turning an object through an angle changes the rotational kinetic energy according to

$$K_{cm} - K_0 = -\int_{\theta_0}^{\theta} \tau_{cm}\, d\theta,$$

in which K_0 is the initial rotational kinetic energy, the energy when $\theta = \theta_0$. The minus sign highlights that while falling, the pencil's tilt θ decreases while the pencil's kinetic energy increases. Thus the rate of change of K_{cm} with angle θ is a negative quantity that is specified by the torque according to $dK_{cm}/d\theta = -\tau_{cm}$. We recall that the rotational kinetic energy K_{cm} of the pencil about its CG is equal to $(I_{cm}/2)\omega^2$. Then, using equation 5, we obtain

$$\frac{I_{cm}}{2}\frac{d\omega^2}{d\theta} = I_{cm}\alpha.$$

We can cancel out the moments of inertia and use equation 7 to find

$$\frac{1}{2}\frac{d\omega^2}{d\theta} = -\frac{6\cos\theta}{L} \times \left[\frac{g - \dfrac{L\omega^2}{2}\sin\theta}{3\cos^2\theta + (1 \mp \mu_k \tan\theta)^{-1}}\right]. \qquad (8)$$

This equation tells us how the angular velocity varies with θ. We want to focus on the sign of a_v, the numerator in the square brackets. Let's suppose it reverses. Then there is an angle θ_0 at which a_v passes through zero. At that angle the vanishing of a_v requires that

$$\omega^2 = \frac{2g}{L\sin\theta_0}.$$

At θ_0, in addition, because a_v is zero, we observe that equation 8 tells us that $d\omega^2/d\theta_0$ also equals zero. Finally, we note that at θ_0,

$$\frac{da_v}{d\theta_0} = \frac{d\left(g - \dfrac{L\omega^2}{2}\sin\theta_0\right)}{d\theta_0} = -\frac{L\omega^2}{2}\cos\theta_0 = -g\cot\theta_0,$$

a quantity that is clearly negative for values of θ_0 between 0° and 90°. Do you see the difficulty? As we have shown above, at the onset of sliding, the function a_v starts out positive. Therefore, if it were to pass through zero as the pencil falls through ever decreasing values of θ, it must do so with *positive* slope—that is, $da_v/d\theta_0$ must be positive.

The assumption that a_v can pass through zero, or that the pencil can levitate, has led us to a contradiction. To avoid the contradiction, we conclude that the pencil cannot levitate while sliding, and thus never levitates from a horizontal table.

Before concluding, let's note how the dynamics of a falling pencil distinguishes it from, say, a block sliding down an incline. In the latter case, as we increase the tilt of the incline, a critical angle is reached at which a block that had been held stationary by static friction suddenly starts to slide. At this point, the normal force on the block has no discontinuous change. This starkly contrasts with the behavior of the normal force acting on our falling pencil. Because of the difference between the static and kinetic friction dynamics of the *rotating* pencil, the two normal forces, $F_\perp$ and $F_\perp^k$, have a discontinuous jump at the changeover of the dynamics.

Uplifting, if nonlevitating

We have wallowed with pleasure in the dynamics of the simplest of physical systems, a falling pencil. We have uncovered some beautiful and rich physics merely with some trigonometry and a smidgen of calculus or a calculator. We have noted that switching over from the dynamics of static friction to the dynamics of kinetic friction, which is familiar to every student of introductory physics, leads to a discontinuity in the normal force acting on the falling pencil. And finally, we have shown that the pencil will never levitate from a horizontal table under the influence of static friction nor can it levitate once it has started to slide.

The same techniques and results can be shown to apply equally well to a tilted table. The critical difference is that when the table has a tilt, the initial value of the virtual acceleration at the onset of sliding can be negative, in which case the pencil will levitate instead of slide.

Such abnormal normal forces can occur only when the tilt of the table exceeds $\sin^{-1}(4/5)$ or about 53.13°. This is the first time we have ever seen a 3–4–5 right triangle, a favorite of the ancients, arise from the intrinsic nature of a physics problem. ◘

Leaf Turner *is a theoretical physicist at the Los Alamos National Laboratory, Los Alamos, New Mexico (e-mail: tleaf@lanl.gov).* **Jane L. Pratt** *is currently a student at Harvey Mudd College in Claremont, California.*

Acknowledgment
This work was initiated when Leaf Turner had been enjoying lunch hours as an Instructor of AP Physics C at Los Alamos High School. Jane Pratt participated under the auspices of the Los Alamos National Laboratory Undergraduate Student Program. This work was supported in part by the U. S. Department of Energy.

Revolutionary teaching

For the fatherland, sciences, and fame

by Yuri Solovyov

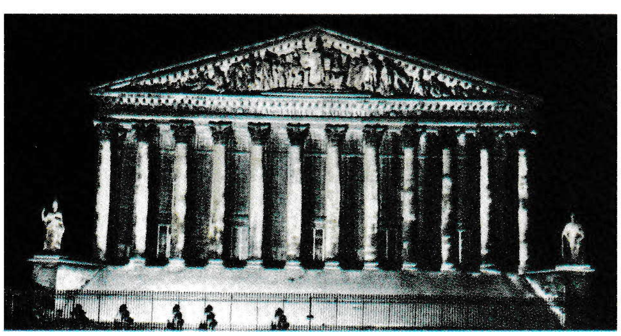
IN THE VERY CENTER OF THE Latin Quarter in Paris, not far from the Pantheon, the Sorbonne, and the Lycée de Louis le Grand, you will find the small and quiet Rue Descartes. For more than a century and a half this has been the location of the famous Ecole Polytechnique.

The school was founded in 1794, during the early and most chaotic years of the French Revolution. This upheaval affected every level of French society. Perhaps for the first time in history, politicians and public figures began to appreciate how much science influences politics, industry, and trade. The wars associated with the revolution and the fierce rivalry between France and England forced the government to get involved in the training of highly skilled specialists who could answer the many challenges faced by the young republic. Consequently, a considerable number of prominent scientists found themselves in the ruling institutions of the revolutionary state. For example, Gaspard Monge, the outstanding geometer, held the position of navy minister, and Lazare Carnot, the gifted mathematician and specialist in mechanics, was one of the founders of the republic's armed forces and its military industry.

These scientists' greatest contribution to the new republic, however, was in the field of higher education. Prior to the revolution, France's system of higher education was in shambles, primarily because it relied on 22 outdated universities hampered by medieval traditions and scholastics. The fields of mathematics, physics, and chemistry, in particular, were in a miserable state. The only exceptions in this archaic system were a few elite military engineering schools—the School for Bridges and Roads (Ponts et Chaussées), the Mézières School for Military Engineers (where Monge taught and from which Carnot graduated), and the Artillery Students' School.

L. Carnot (1753–1823)

G. Monge (1746–1818)



The front of the Bourbon Palace

During the first years of the Revolution, many higher educational institutions and specialized high schools were closed. In 1793 the supreme body of revolutionary power—the convention—began to reorganize the educational system in the country. The decree of 29 Frimaire (according to the new revolutionary calendar; we would give the date as 19 December, 1793) made primary education free and compulsory for all.

Around the same time, at the recommendation of Monge and other prominent scientists, a Commission on Public Works was formed. It was this commission that proposed the creation of a new school of higher education that would train specialists in mathematics and the sciences. Contrary to the ideals of education as the comprehensive development of the whole personality that had developed in the eighteenth century, the new school was supposed to aim toward the fastest possible achievement of results in mathematics, science, and technology. Material rewards and honor were promised its students to encourage them to achieve at the highest levels and as quickly as possible.

The committee supported Monge completely and commissioned him, as well as other scientists, to elaborate on the proposal. On 21 Ventose of the Year II of the Republic (March 11, 1794), the Committee adopted a decision to establish a new school for training engineers, the School of Public Works. At the end of the academic year, the name of the school was changed to Ecole Polytechnique to clearly describe its goals. The purpose of the school was given as "to train diverse engineers, to restore the teaching of the exact sciences, which has been interrupted by the crises of the Revolution, and to provide higher scientific education to young people, so that they can either be of use by the government in the work of the Republic, or bring enlightenment to their native cities and spread there useful knowledge."

The Committee provided for entrance examinations in 22 French cities to select 400 male persons between the ages of 16 and 20 who "have proven their devotion to Republican principles and displayed a good knowledge of arithmetic and the elements of algebra and geometry." The difficult situation of the country made it impossible to satisfy the rigid restrictions on the candidates' age, so the youngest of them turned out to be only 12, and others were well past 20. The first director of the school was the former head of the School of Bridges and Roads, Lamblardi. He was soon replaced by Monge, who remained at this post for four years.

The school was originally housed in the Bourbon Palace. Classes started on December 21, 1794. The course of study lasted three years, and included calculus, geometry, descriptive geometry, technical drawing, mechanics, physics, chemistry, architecture and military engineering. The professors were such prominent French scientists as Lagrange, Monge, Laplace, Le Peletier, Berthollet, and Neveux. Even the first entering class included such outstanding scientists as Biot (the astronomer and physicist), Poinsot (known for his work in geometry and theoretical mechanics), Malus (who discovered the polarization of light) and the archaeologist De Chezy (who deciphered the Assyrian cuneiform texts).

Monge devoted all his time and resources to the school. He created a course of descriptive geometry, the basis of many technical disciplines. "No one was as good a teacher as Monge," recalled his student Brisson, a well-known engineer. "His gestures, poses, the modulation of his voice—everything served to develop his thoughts. He always followed his listener's eyes, and could judge each one's degree of understanding. We got to know Monge, that finest of men, as one who is devoted to youth and to science. He was always among us: After lectures in geometry, calculus and physics, private conversations arose that reinforced our abilities still more. He was a friend to each of his charges, encouraged us in our work, and enjoyed our success."

After this first academic year came to an end, the course of study was shortened to two years to reflect a change in purpose at the school. Rather than turning out fully trained engineers, it was now intended to graduate students ready for two more years of more specialized study in the Schools of Bridges and Roads, of Mines, of Military Engineers, and so on. Students were assigned to a school based on the quality of their work. A graduate with high standing could attend any school. The lower your standing, however, the fewer your choices.

These schools also accepted other students for a four-year course of study. But they did not share the standing of graduates of the Ecole Polytechnique, who were considered to be in state service and received a salary. A new system of entrance exams was established. Candidates were tested in arithmetic, geometry, and algebra, including the solution of polynomial equations of degrees two, three, and four and the theory of infinite series. The following are examples of the types of problems found on these examinations:

1. Prove that a triangle with two equal angle bisectors is isosceles.

2. Divide each side of a triangle into the parts proportional to the squares of their adjacent sides and join the points of division to the opposite vertices. Prove that the lines thus obtained meet in a point, and this point is the center of mass of the triangle formed by the point's projections on the sides of the given triangle.

3. Given angle AOB and point P, find a point M on side AO of the angle such that the circles C and C' drawn through M and P, and tangent to OB, will intersect each other at a given angle.

4. Construct a triangle, given one of its angles, its perimeter, and its area.

5. Let a, b, c, and d be the successive side lengths of a given quadrilateral. Prove that the circumradii of the two quadrilaterals, one of which is formed by the bisectors of the given quadrilateral, and the other one by the bisectors of the exterior angles, are in the ratio

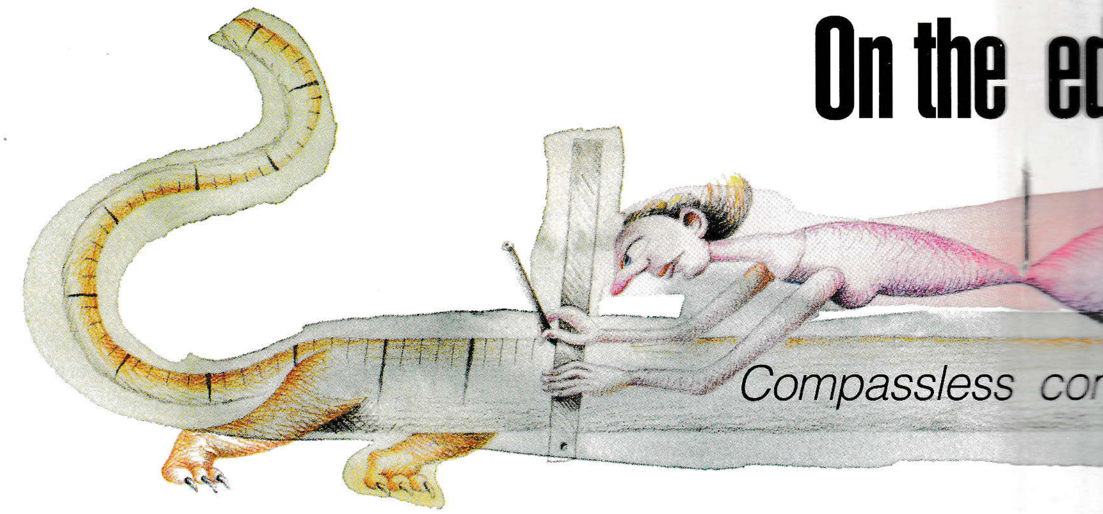
$$\frac{a+c-b-d}{a+c+b+d}.$$

Instruction at the Ecole Polytechnique was provided by professors (who gave lectures), tutors (who explained the lectures and supervised practice sessions), and examiners (who checked students' knowledge with very difficult comprehensive examinations that every student had to pass). Instruction followed a well worked out plan. During the schools' first decade, the mathematical disciplines received the greatest emphasis, taking up about 20 hours each week. These included calculus, synthetic and analytic geometry, mechanics, descriptive geometry, and technical drawing. Experimental physics and chemistry took up a large part of the second year of study.

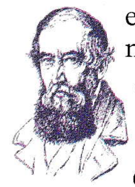
It was required by law to publish all lectures given at the Ecole Poly-

CONTINUED ON PAGE 44

On the ed



Compassless cor



IN GEOMETRY, WE USUALLY do constructions with compass and straightedge, but we can also do constructions using a straightedge alone. Constructions made *sans* compass are called *Steiner constructions,* after the outstanding German geometer of the nineteenth century. Let's study one specific type of Steiner construction: those that begin with two parallel lines on the plane. Many are based on the following properties of a trapezoid.

1. Consider trapezoid $ABCD$ with bases AD and BC. Let P be the point of intersection of its diagonals and Q be the point where the extensions of its lateral sides meet. Then the line PQ passes through the midpoints of the bases AD and BC.

Proof. Let N and M denote the midpoints of AD and BC, respectively (fig. 1). First we prove that points Q, M, and N are collinear. Triangles BCQ and ADQ are similar, so the angles made by corre-

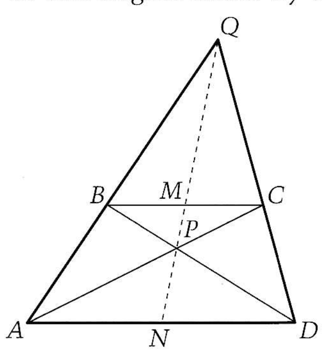


Figure 1

sponding medians QM and QN with corresponding sides CQ and DQ are equal. That is, $\angle BQM = \angle AQN$, and since A, B, and Q are collinear, so are Q, M, and N. In the same way, we can show that points P, M, and N are collinear, using similar triangles BCP and DAP. Thus the four points P, Q, N, and M all lie on the same line. Thus, the four points P, Q, N, and M all lie on the same line.

Now we can easily solve the two following problems:

2. Use a straightedge to divide a given segment into two equal parts if a line parallel to the segment is given. The solution is easy to see if we think of AD (in fig. 1) as the given line segment.

3. Consider two parallel lines and a point Q on the plane. Use a straightedge to draw a line through Q parallel to the two given lines.

We embed the two lines in a copy of figure 1. That is, we draw two lines through point Q, intersecting one of the given parallels at points A and D, and the other at B and C. Now we have trapezoid $ABCD$. Drawing diagonals AC and BD, we find point P, and drawing line QP, we find midpoints M and N. We draw lines AM and CN, and label their intersection point R. Then QR is parallel to AD and BC. Indeed, from pairs of similar triangles we find $RC:RM = MC:AN = BM:AN = AB:BQ$. Since the lines AD, BC, QR cut off proportional segments on transversals QA, RN, these three lines are parallel. The case where point Q is between the two lines is handled analogously and is left for the reader to explore.

Statement 1 also helps solve the following problem:

4. Consider line l with three points A, N, and D marked on it such that $AN = ND$. Let Q be a point that does not belong to l. Using a straightedge alone, draw a line through Q parallel to l. The solution is left to the reader (for example, we can once more reconstruct figure 1).

By applying problem 1 several times, we can divide a given segment into the ratio $1:2^k$ for any positive integer k. If we are given a segment and a line parallel to it, however, we can do more: We can divide the segment into the ratio $1:n$ for any positive integer n. The construction is based on the following problem, which contains problem 1 as a special case.

5. Let the lateral sides AB and CD of trapezoid $ABCD$ meet at the point Q and let K be an arbitrary point of segment BC. Let P be the point where KD and AC meet and let QP intersect AD at L. Then, if $KC = \lambda BC$, we have

$$LD = \frac{\lambda}{\lambda + 1} AD.$$

Proof. Let QL meet BC at F (see fig. 2). Suppose $LD = xAD$. Then, since triangles APD and CPK are similar, we have $KF = xKC$. Thus,

$$FC = (1 - x)KC = (1 - x)\lambda BC.$$

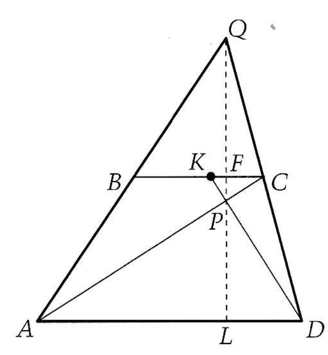
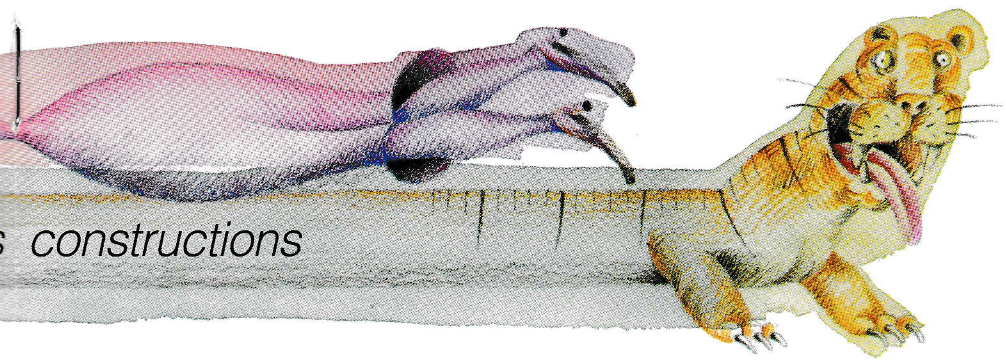


Figure 2

Art by Virginia Smith

edge



But triangles QBC and QAD are also similar. Therefore,

$$(1-x)\lambda = \frac{FC}{BC} = \frac{LD}{AD} = x,$$

That is, $(1-x)\lambda = x$, and

$$x = \frac{\lambda}{\lambda+1}.$$

So the statement is proved. Now suppose the given segment is AD. We also have a line parallel to AD, which we can use to play the role of BC in figure 1. We then know how to divide BC in the ratio 1:2. Letting $\lambda = 1/2$ in the last equation, we find that it allows us to cut AD in the

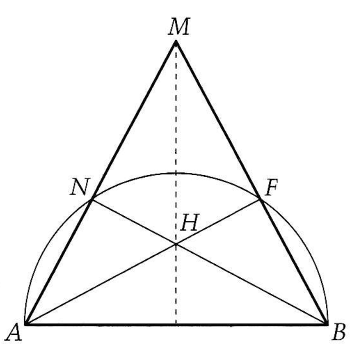


Figure 3

ratio 1:3. Then we can do the same to BC, which allows us to cut AD in the ratio 1:3. Then we can do the same to BC, which allows us to cut AD in the ratio 1:3. Then we can do the same to BC, which allows us to cut AD in the ratio 1:4, and so on.

Now let's consider several other constructions using a straightedge.

6. A semicircle with endpoints A and B is drawn on the plane. Draw a line through a given point M perpendicular to the line AB.

To complete this construction, we first consider the situation

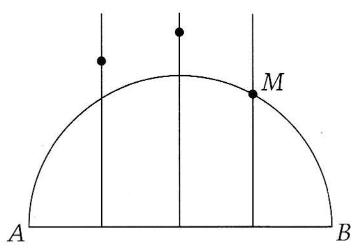


Figure 4

shown in figure 3. Segments AF and BN are altitudes in triangle AMB because the inscribed angles ANB and AFB oppose the diameter AB. Therefore, H is the point where altitudes of triangle AMB meet, and thus MH is perpendicular to AB. So, in the case shown in figure 3, the problem is solved. But what do we do if point M lies on the arc AB (fig. 4) or in some other "inconvenient" place?

In this case we take two arbitrary "convenient" points on the plane and draw perpendiculars through them to the line AB. Then we draw a line through M parallel to these perpendiculars as explained in problem 3.

7. Two intersecting circles on the

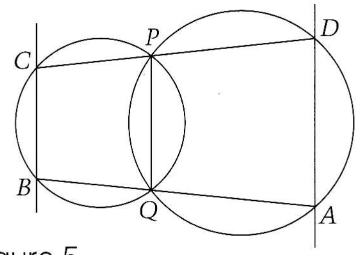


Figure 5

plane are given. Use a straightedge to find their centers.

Here we need another simple fact: If we draw lines through the points where two circles intersect, then the chords that these lines cut from the circles are parallel. (In figure 5 these chords are BC and AD.)

This is not difficult to prove if we remember that opposite angles of a quadrilateral inscribed in a circle are supplementary. Indeed, $\angle CBQ = \angle QPD$ since they are both supplementary to $\angle CPQ$. Then $\angle QPD$ is supplementary to $\angle QAD$, so $\angle CBQ$ supplements $\angle QAD$. This means that BC and AD are parallel.

Now that we have a pair of parallel lines, we can use problems 2 and 3 to find the center of the circle through C, D, and P. What we need is a pair of parallel chords in this circle. We can get this using problem 3, which will tell us how to draw a chord (of this circle) through P parallel to CB. Then, by problem 2, we can bisect each of these chords. The reader can prove that this line must pass through the center of the circle. If we now start over, with two more lines through P and Q, we will get two more parallel chords, one in each of the two original circles. We again use these to get parallel chords in the circle through C, B, and P. The line joining their bisectors is again a centerline, and two centerlines intersect at the center. Of course we can do the same with the other circle.

The last construction is complicated to carry out physically. We sometimes say that geometric constructions are carried out" with the aid of language"; that is, we are merely proving that such a construction is possible, rather than actually carrying it out.

—by Igor Sharygin

Around and around she goes

"Revolutions are celebrated when they are no longer dangerous."
—Pierre Boulez (b. 1925)

by Larry D. Kirkpatrick and Arthur Eisenkraft

MERRY-GO-ROUNDS ARE the most egalitarian ride in that everybody can have a good time. As our personal thrill tolerances increase, we can try such rides as the Loop-the-Loop, the Tilt-A-Whirl, and the Rotor, where the floor is pulled out from under us when the cylinder is spinning fast enough to "pin" us to the wall. The most deceptive ride is the Tea Cups at Disneyland. The Tea Cups subjects us to three simultaneous circular motions. It looks tame enough, but the ride can be a quite dizzying experience depending on how your friends spin the cups and platform.

Circular motions can also be entertaining in the physics class, though we wouldn't suggest there will ever be two-hour lines of people waiting to learn about centripetal forces and angular momentum.

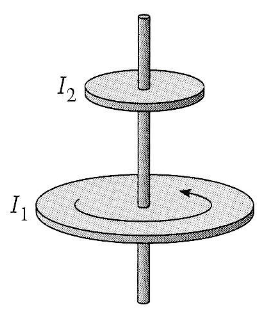


Figure 1

Let's consider some classical examples of circular motion before we move on to some interesting variations. We choose problems that are most easily solved using the conservation laws of linear momentum, angular momentum, and kinetic energy.

Consider a circular disk of radius R_1, mass M_1, and moment of inertia I_1 rotating at a constant angular speed ω_1 about its axis of symmetry. Its angular momentum $L_1 = I_1\omega_1$, its rotational kinetic energy $KE_1 = (1/2)I_1\omega_1^2$, and if it's a uniform disk, its moment of inertia $I_1 = (1/2)M_1R_1^2$, all about the center of mass of the disk.

Now let's assume that we drop a similar disk so that it lands exactly on top of the first disk, face-to-face, as shown in figure 1. Then conservation of angular momentum tells us that

$$L_f = L_1 + L_2 = I_1\omega_1 + I_2\omega_2 = (I_1 + I_2)\omega_f,$$

where the subscript f refers to the final conditions. As a simple example, let $\omega_2 = 0$ and $I_1 = I_2$. Then $\omega_f = (1/2)\omega_1$, as we might expect.

Although angular momentum is conserved, kinetic energy is not conserved. Let's remain with the case $\omega_2 = 0$. Then

$$KE_i = \frac{1}{2}I_1\omega_1^2$$

and

$$KE_f = \frac{1}{2}(I_1 + I_2)\omega_f^2 = \frac{1}{2}\left(\frac{I_1^2}{I_1 + I_2}\right)\omega_1^2$$

$$= KE_i\left(\frac{I_1}{I_1 + I_2}\right).$$

Note that KE_f is always less than KE_i and that the collision is inelastic. If $I_1 = I_2$, half of the original kinetic energy is lost.

As a second example, let's consider a small ball with mass m and speed v colliding with the rim of the circular disk as shown in figure 2.

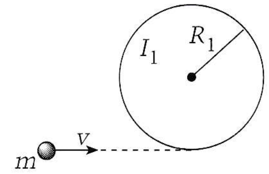


Figure 2

We assume that the disk is initially not rotating, the disk is free to rotate about a fixed axis through its center, and that the ball sticks to the rim. With what angular speed does the wheel rotate?

We use conservation of angular momentum. You might think that there is no initial angular momentum, but even a ball moving in a straight line has angular momentum about all points not on the line of its motion:

$$L_i = mvR_1.$$

Art by Tomas Bunk


TOM BUNK
LEFT - RIGHT
ON
OFF
UP DOWN
MADE IN TAIWAN
MANIFEST

We also know that the final angular momentum is

$$L_f = I_f\omega_f,$$

where I_f is the combined moment of inertia of the disk and the ball. If we assume a uniform disk, we have

$$I_f = I_1 + mR_1^2 = \left(\frac{1}{2}M_1 + m\right)R_1^2.$$

Because $I_f > I_1$,

$$\omega_f < \frac{v}{R_1},$$

and the ball slows down. For the case of equal masses,

$$\omega_f = \frac{2v}{3R_1}.$$

Convince yourself that one-third of the original kinetic energy of the ball is lost in this collision.

The first of our contest problems is based on part of a problem on the preliminary exam that was given nationwide in January to select the U.S. Physics Team that will compete in Iceland this summer.

A. A disk of radius R spins with angular speed ω_0 about its axis, which is held vertically in friction-

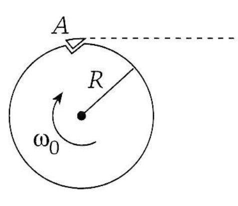


Figure 3

less bearings. The disk's moment of inertia about the spin axis is I_0. At a certain instant, a small chip of mass m breaks off the rim of the disk and flies away moving tangent to the disk as shown in figure 3. What is the angular speed of the disk after the chip breaks off?

B. A ball of mass m and speed v strikes the end of a thin rod of mass M and length a as shown in figure 4. Assume that the rod lies on a frictionless table and that the ball stops after the collision. What are the ve-

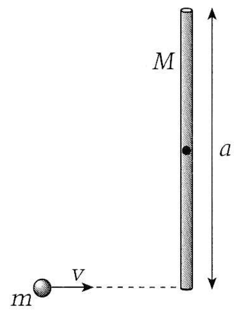


Figure 4

locity of the center of mass of the rod and its rotational speed about its center of mass? For what range of mass m is such a collision possible? (The moment of inertia for the thin rod is $Ma^2/3$ about one end and $Ma^2/12$ about its center of mass.)

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.

Cool vibrations

Our contest problem in the September/October 1997 issue explored some classic problems in oscillations. The first two parts dealt with the change in spring constant (and therefore the oscillation frequency) when springs are added in series or in parallel. Most readers find the problem of two springs in parallel to be straightforward. Two identical springs in parallel will each have to support half the weight of the suspended mass. Each spring will stretch half as much in order to apply half the force. The equivalent spring constant of the pair of two identical springs would therefore be $2k$ since the pair stretches only half as much:

$$F = kx = 2k\frac{x}{2}.$$

A general rule for springs in parallel is

$$k' = k_1 + k_2.$$

In contrast, springs in series must each support the entire weight of the suspended mass. The total stretch of the two identical springs would be twice the stretch of one spring alone. The combined spring constant is $k/2$:

$$F = kx = \frac{k}{2}2x.$$

A general rule for springs in series is

$$\frac{1}{k'} = \frac{1}{k_1} + \frac{1}{k_2}.$$

In part A of the contest problem, a spring is cut in half. With only half the spring, the new spring constant must be $2k$. Two identical half springs of $2k$ and $2k$ in series would have the combined (original) spring constant of k. With k being doubled as a result of cutting the spring in half, the frequency is now

$$\nu = \frac{1}{2\pi\sqrt{\frac{2k}{m}}} = \nu_0\sqrt{2},$$

where ν_0 is the original frequency.

In part B, springs k_2 and k_3 are in parallel. The combined spring constant of k_2 and k_3 is $k_2 + k_3$, following our first rule. Adding k_1 in series

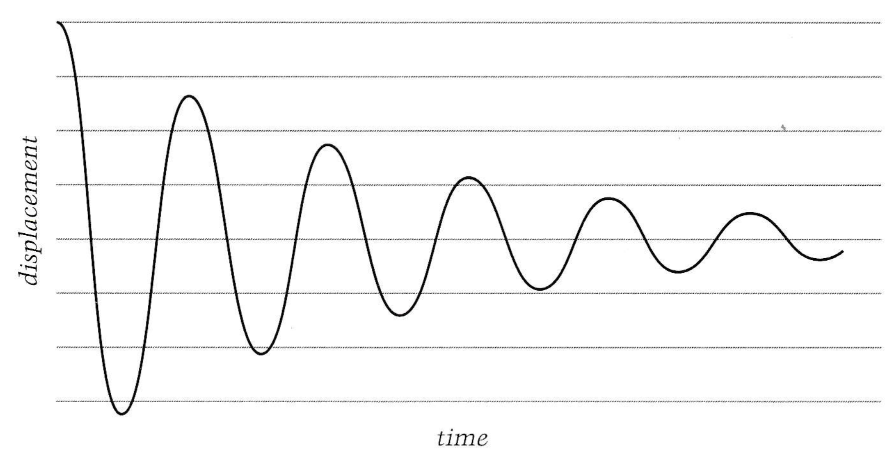


Figure 5

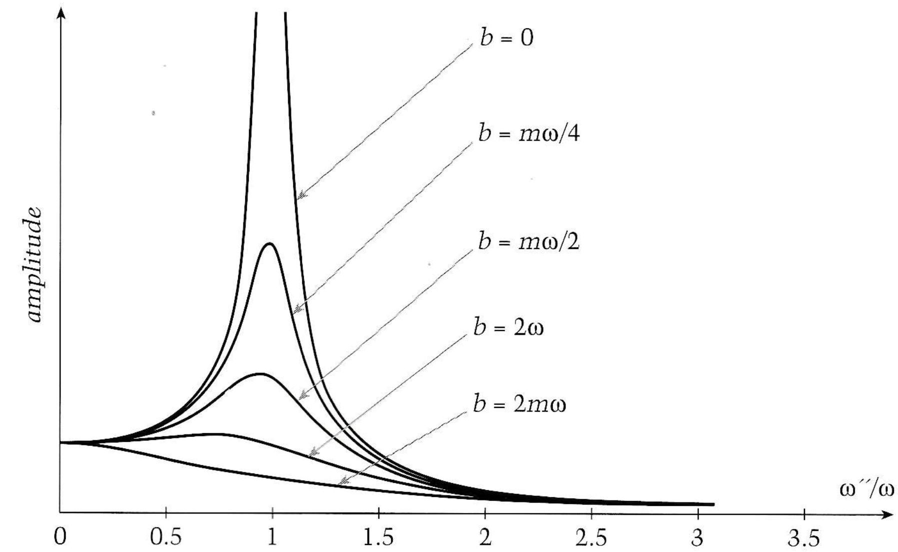


Figure 6

makes the equivalent resistance of the combination

$$\frac{1}{k} = \frac{1}{k_1} + \frac{1}{k_2 + k_3},$$

or

$$k = \frac{k_1(k_2 + k_3)}{k_1 + k_2 + k_3}.$$

The corresponding period T is

$$T = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{m(k_1 + k_2 + k_3)}{k_1(k_2 + k_3)}}.$$

C. Using a spreadsheet and corresponding graphing program, we can generate a graph of the solution for the damped oscillator (fig. 5). The mean lifetime can be found by substituting A/e for the displacement in the equation.

$$\frac{1}{e}A = Ae^{-\frac{bt}{2m}}$$

$$\ln\frac{1}{e} = -\frac{bt}{2m}$$

$$t = \frac{2m}{b}$$

Finding the number of oscillations requires us to first calculate the frequency and the elapsed time when the maximum displacement is $A/4$.

$$\frac{1}{4}A = Ae^{-\frac{bt}{2m}}$$

$$t = (\ln 4)\left(\frac{2m}{b}\right) = 30.8 \text{ s}$$

The ω, using the values given, is 2.236 rad/s, which corresponds to a frequency of 0.356 Hz. Then

$$(0.356 \text{ oscillation/s})(30.8 \text{ s}) = 11 \text{ oscillations.}$$

The corresponding velocity can be found by differentiating the displacement equation:

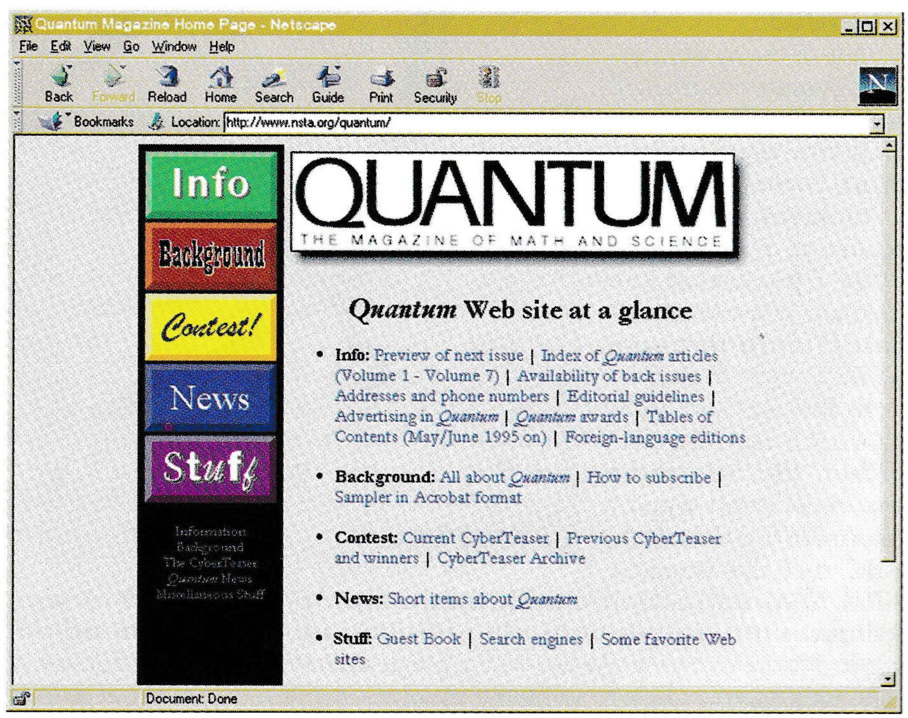
$$v = \frac{dx}{dt} = Ae^{-\frac{bt}{2m}}\left[-\sin(\omega' t + \phi)\right]\omega' + x\left(-\frac{b}{2m}\right)$$

$$v = -Ae^{-\frac{bt}{2m}}\left[\omega'\sin(\omega' t + \phi) + \left(\frac{b}{2m}\right)\cos(\omega' t + \phi)\right].$$

D. Using a spreadsheet and corresponding graphing program, we can generate a graph of the solution for the forced oscillator for different values of the damping coefficient b (fig. 6). One notices that if $b = 0$ and the driving frequency is equal to the natural frequency, we have a resonance effect gone wild, and the amplitude grows without bound. As the damping coefficient b gets larger, we notice that the resonance effects seem to diminish in size.

The corresponding velocity can be found by differentiating the displacement equation:

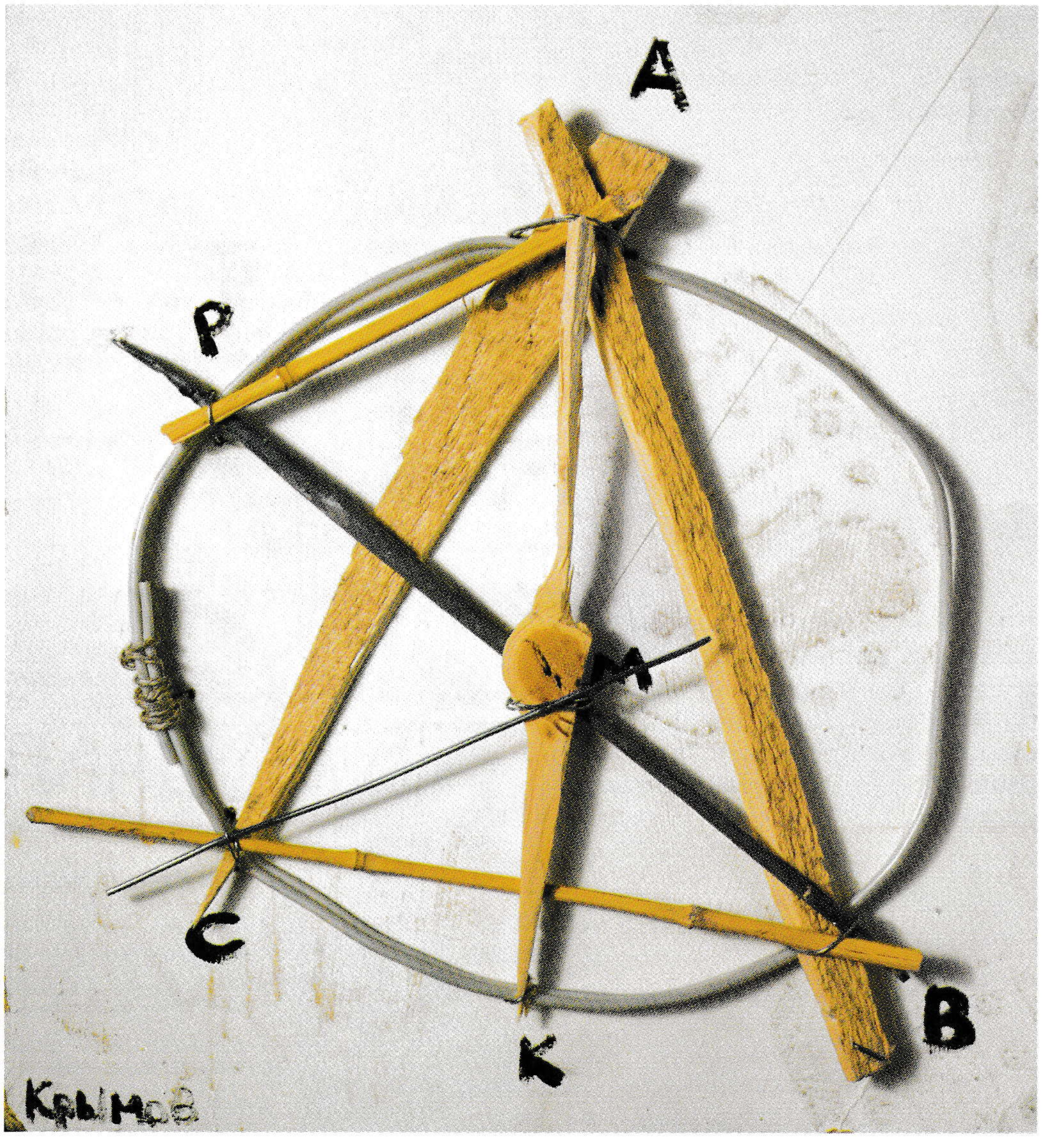
$$v = \frac{dx}{dt} = \frac{F_m}{G}\omega''\cos(\omega'' t - \phi).$$



Points of interest

Unique locations within a triangle

by I. F. Sharygin



Art by Dmitry Krymov

IN GEOMETRY, WE learn about several remarkable points associated with a triangle, such as the centroid, the circumcenter, the incenter, and the orthocenter. In this article, we explore some nontrivial properties of these points. Many of these properties are, in fact, fully equivalent to the definitions of these points. That is, the "remarkable point" is the only one that possesses the property we describe.

The centroid

One of the most interesting points of a triangle is the centroid: the point of intersection of its medians. Let us assume, for a moment, that the reader is not familiar with a proof that the medians of a triangle all pass through the same point.

In fact, let us forget about the centroid altogether. We will show that there exists a point M inside any triangle ABC, such that triangles ABM, BCM, and CAM all have equal areas (figure 1). We will prove that this point exists by actually constructing it. Where can point M lie? The triangles ABM and ABC will share side AB, so (using absolute value for area), if we want $|ABM|$ to be $(1/3)|ABC|$, then the altitude (to side AB) of ABM must be $1/3$ the corresponding altitude of ABC. Thus M must lie on a line parallel to AB, at a distance equal to $1/3$ the altitude to AB in

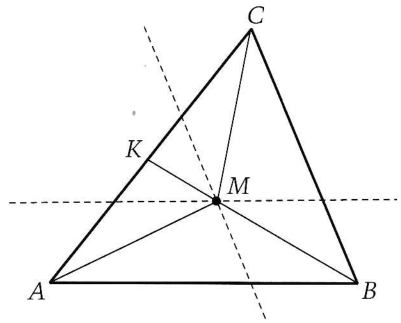


Figure 1

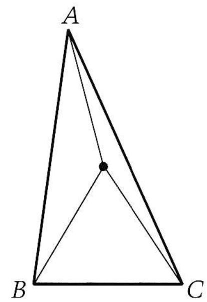


Figure 2

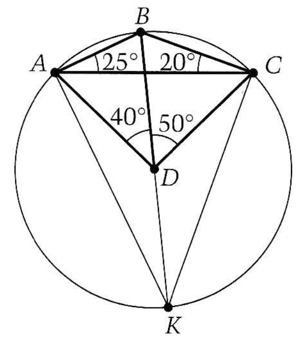


Figure 3

ABC. But M must also have this relationship to side BC; that is, it must lie on a line parallel to BC and at a distance equal to 1/3 the altitude to BC in ABC. The only possible candidate for M is the intersection of these lines (which are clearly not parallel to each other). In fact, this point M works, since if $|ABM| = |BCM| = (1/3)|ABC|$, then $|CAM|$ is also $(1/3)|ABC|$, and our construction shows that this is the only possible such point M.

Now let's return to the question of medians. We can show that each median of ABC passes through M, and that this point divides each median in the ratio 2:1 (starting at the vertex of the triangle). Indeed, let us extend BM to intersect AC at K.

Since $|AMB| = |BMC|$, these triangles have equal altitudes to their common side BM. The reader can prove (for example by drawing these altitudes and considering congruent triangles) that this implies that K is the midpoint of AC, so BK, which passes through M, is a median. Again, $|AMC| = (1/3)|ABC|$, so (as we have already seen) the distance from M to AC is 1/3 the distance from B to AC. Dropping perpendiculars to AC from M and B, and considering the similar triangles they form, we find that this implies that $MK = (1/3)BK$, so M divides median BK in the ratio 2:1, starting from vertex B. Of course, the same results hold for the other two medians of ABC.

Thus we have found the following alternative description of the centroid:

Alternative 1: Point M is the centroid of triangle ABC if and only if triangles ABM, BCM, and CAM have equal areas.

The circumcenter

Another remarkable point in a triangle is the center of its circumscribed circle or *circumcenter*.

Problem 1: In triangle ABC, $\angle A = 30°$ and $\angle B = 80°$. Point K is chosen inside ABC such that triangle BCK is equilateral. Find $\angle KAC$.

Solution: We could apply the law of sines to the situation, but it's better to note that K is the circumcenter of triangle ABC. Indeed, at the circumcenter, side BC subtends an angle equal to twice $\angle BAC$, or 60° (see figure 2). The circumcenter also lies on the perpendicular bisector of side BC. It is not difficult to see that point K is the only one that satisfies both these conditions and so is in fact the circumcenter. Then $\angle CKA = 2\angle CBA = 160°$, and $\angle KAC = 10°$.

Problem 2: In convex quadrilateral $ABCD$, $\angle BAC = 25°$, $\angle BCA = 20°$, $\angle BDC = 50°$, and $\angle BDA = 40°$. Find the acute angle formed by the diagonals of the quadrilateral.

Solution: We reason indirectly and show that D is the center of the circle circumscribed about ABC. Since triangle ABC is obtuse, its circumcenter lies on the opposite side of AC from point B, and from it, sides BA and BC subtend angles of 40° and 50° respectively. The reader is invited to check that there can be only one such point, so it must be point D. It is not difficult now to solve the problem.

However, direct reasoning is usually preferable in geometry. Since we've now guessed the true role of point D, we can try to reason directly. We begin by drawing the circle circumscribing ABC, and extending AD until it meets the circle (fig. 3). In triangle DKA, $\angle DKA = \angle BCA = 20°$, and the exterior angle at vertex D is 40°. Thus, $\angle DAK = 20°$, and $DK = DA$. Similarly, $DK = DC$, and therefore D is the circumcenter of ABC. Now we can easily find the angle between the diagonals: it is 85°.

The incenter

Another remarkable point in a triangle is the center of the inscribed circle, or *incenter*. Let I be the incenter of triangle ABC. We begin by stating two properties of this point that will prove handy when we look for alternative descriptions of the incenter I.

Property I_1: If I is the incenter of triangle ABC, then $\angle AIC = 90° + (1/2)\angle B$.

Property I_2: If I is the incenter of triangle ABC, the line BI passes through the circumcircle of triangle AIC.

The following alternative descriptions of the incenter are based on properties 1 and 2:

Alternative I_1. Let M be a point inside triangle ABC such that $\angle BMC = 90° + (1/2)\angle A$, and line AM passes through the circumcenter of triangle AMC. Then M is the incenter of ABC.

Alternative I_2: Let M be a point inside triangle ABC such that line AM passes through the circumcenter of BMN and line MB passes through the circumcenter of AMC. The M is the incenter of ABC. (The reader can

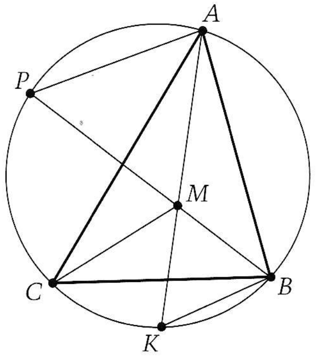


Figure 4

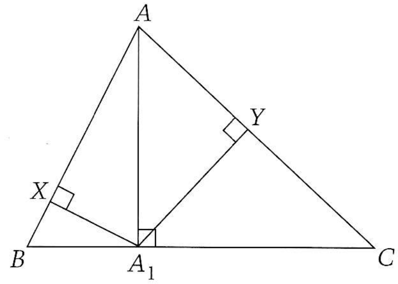


Figure 5

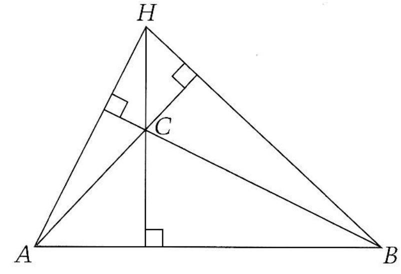


Figure 6

check the converses; that is, if M is the incenter, then M enjoys the two properties described above.)

We'll limit ourselves here to the proof of Alternative I_2. Suppose the circumcircle of ABC meets lines AM and BM for the second time by K and P respectively (fig. 4).

We can show that $\angle MCB = 90° - \angle KMB$. Indeed, the circumcenter O of triangle MCB (not shown in the diagram) lies on AM, and if $\angle MCB = x$, then $\angle MOB = 2x$, and (since triangle OMB is isosceles)

$$\angle KMB = (1/2)(180° - 2x) = 90 - x.$$

This is the result we need.

Similarly, $\angle MCA = 90° - \angle PMA$. Angles KMB and PMA are equal, so $\angle MCA$ and $\angle MCB$ are also equal, and MC bisects $\angle C$. Therefore,

$$\angle KMB = 90^\circ - \angle MCB = 90^\circ - \frac{1}{2}\angle ACB$$
$$= 90^\circ - \frac{1}{2}\angle MKB = \frac{1}{2}(\angle KMB + \angle KBM)$$

That is, $\angle KMB = \angle KBM$, and thus $KM = KB$. So the circle with center K and radius KB passes through point M. It is not hard to see that it must pass through C as well. Indeed, $\angle MCB = 90° - \angle KMB$, and $\angle MKB = 180° - 2\angle KMB$. So $\angle MCB = (1/2)\angle MKB$, which means that C is on the required circle.

But then K is the circumcenter of triangle MKB, so AK bisects angle CAB. Similarly, MB must bisect angle CBA, and M, their intersection, is the incenter of triangle ABC.

Exercises 1–3: Prove properties I_1 and I_2, and alternative I_1.

The following property of the incenter of a triangle is based on geometric vectors. It comes in handy in solving many problems. The most remarkable aspect of this property is that it generalizes to three-dimensional space (and even spaces of higher dimensions).

Alternative I_3: If a, b, and c are the lengths of the sides of the triangle ABC, and I is its incenter, then $a\overrightarrow{IA} + b\overrightarrow{IB} + c\overrightarrow{IC} = \vec{0}$.

Proof. Let I be the center of the inscribed circle, and let AI meet BC at point A_1. Then

$$a\overrightarrow{IA} + b\overrightarrow{IB} + c\overrightarrow{IC} = a\overrightarrow{IA} + b\left(\overrightarrow{IA_1} + \overrightarrow{A_1B}\right) + c\left(\overrightarrow{IA_1} + \overrightarrow{A_1C}\right)$$
$$= \left(a\overrightarrow{IA} + b\overrightarrow{IA_1} + c\overrightarrow{IA_1}\right) + \left(b\overrightarrow{A_1B} + c\overrightarrow{A_1C}\right) = k\overrightarrow{IA}.$$

(In the last transformation we've used the well-known property of the bisector:

$$\frac{A_1B}{A_1C} = \frac{c}{b}.$$

Readers unfamiliar with this fact can find a proof in a standard geometry textbook.)

Thus, the vector sum in the product is a vector collinear with the line AI. Similarly, we can show that this sum is collinear with BI and CI. Thus it must have length 0.

Exercise 4: Show that the equation of alternative I_3 is equivalent to the statement that I is the incenter of triangle ABC.

The orthocenter

One more remarkable point of a triangle is the point where the altitudes meet, or the *orthocenter*. There are many different ways to show that the altitudes of a triangle meet at one point. We will give a proof related to the one we gave for the concurrence of the medians, and leave it for the reader to make conscious the relationship.

We will show that the altitude to BC is the set of points such that the ratio of their distances to AB and AC is equal to cos B/cos A.

Let's first consider an acute triangle ABC. First we look at the distances from A_1 to AB and AC (see figure 5, in which line segments measuring these distances are labeled AX and AY respectively). From right triangle AA_1X, we find $A_1X = AA_1 \sin \angle BAA_1$. But from right triangle ABA_1, we see that $\sin \angle BAA_1 = \cos \angle B$, so we can write $A_1X = AA_1 \cos \angle B$. Similarly, $A_1Y = AA_1 \cos \angle C$, so the ratio $A_1X{:}A_1Y = (\cos \angle B){:}(\cos \angle C)$.

Now we pick any point P at all on altitude AA_1, and note that the ratio of its distances to AB and AC is also equal to $(\cos \angle B){:}(\cos \angle C)$. In fact, altitude AA_1 is the locus of points such that the ratio of their distances to AB and AC is equal to $\cos \angle B/\cos \angle C$. (The proofs of these assertions and of their converses are based on similar triangles and are left to the reader). Of course, corresponding statements hold for the other two altitudes.

Now let H be the intersection of altitudes AA_1 and BB_1. Then the ratio of the distances from H to AB and AC is $\cos B/\cos C$, and the ratio of its distances to AB and BC is $\cos A/\cos C$. Therefore the ratio of the distance from H to AB and BC is

$$\frac{\dfrac{\cos B}{\cos C}}{\dfrac{\cos A}{\cos C}} = \frac{\cos B}{\cos A}.$$

But this means that H lies on altitude CC_1, and the three altitudes are concurrent.

The situation for a right triangle is simple: The orthocenter is just the vertex of the right angle. For an obtuse angle, the reader can construct a proof that is a variation on the one above. Or, we can note that if H is the intersection point of two altitudes AH and BH of obtuse triangle ABC, then AC and BC lie along the altitudes of acute triangle ABH (see figure 6). This means that C is the orthocenter of acute triangle ABH, so the perpendicular from H to AB passes through C. But then the third altitude of ABC lies along this line, and the three altitudes are concurrent at H.

Exercise 5: Prove, in general, that if H is the orthocenter of triangle ABC, then in fact any of the four points A, B, C, and H is the orthocenter of the triangle determined by the other three.

The following property is often useful in solving problems related to the orthocenter of a triangle:

Property O_1: The radius of the circle through two vertices of a triangle and its orthocenter is equal to the triangle's circumradius.

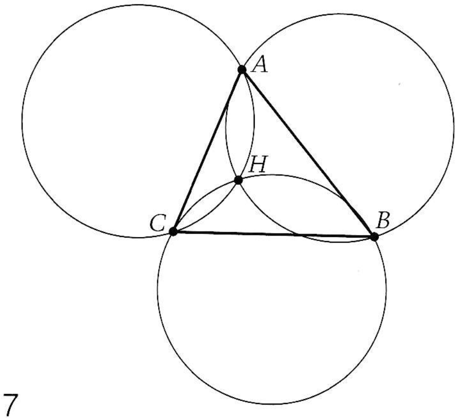


Figure 7

This property is a consequence of a slightly stronger statement that can be made:

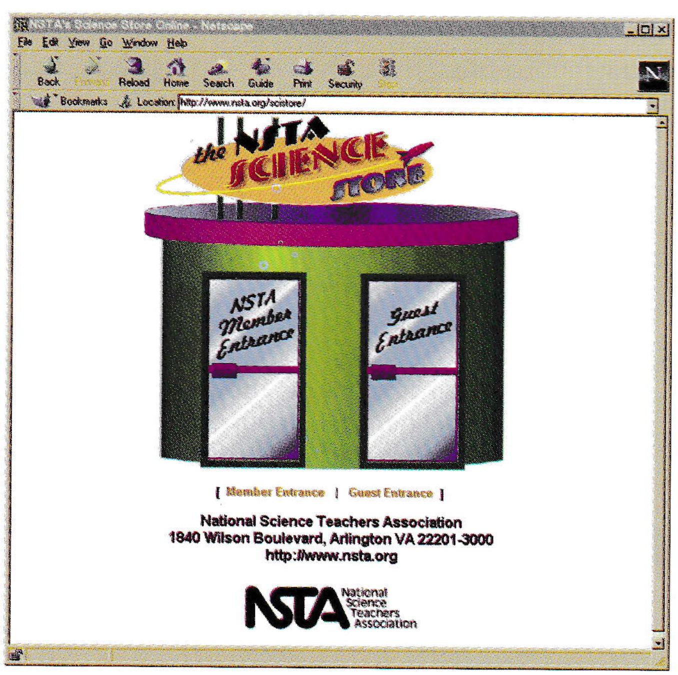
Property O_1': The circle through two vertices of a triangle and its orthocenter is symmetric to the circumcircle of the triangle with respect to the corresponding side of the triangle.

Exercise 6: Prove property O_1 by proving property O_1'.

We conclude our article with one more theorem.

Problem 3: Three equal circles pass through a common point. Show that this point is the orthocenter of the triangle formed by the other intersections (in pairs) of the three circles.

Solution: Let the three circles meet at point H and let the other points of intersection by A, B, and C (figure 7). Since the circle through B, C, and H is symmetrical to the circumcircle of ABH (with respect to BH), it must contain the orthocenter of triangle ABH. Similarly, the circle through A, C, and H contains the orthocenter of triangle ABH. Thus, the orthocenter of ABH must be point C. By exercise 5 above, this means that H is the orthocenter of ABC, and we are done.



Visit the NSTA Science Store Online!

Open 24 hours a day, 7 days a week

Over 300 items specially selected for science teachers, students, and parents

Books, posters, software, CD-ROMs, and more

Browse at your leisure, order at your convenience:

- online
- by phone
- by fax
- by mail

Our Web address:

www.nsta.org/scistore

Homemade pendulums

Getting into the swing of things

by G. L. Kotkin

FEW DEVICES IN PHYSICS are as simple and reliable as pendulums. Nevertheless, experiments with them can illustrate many interesting features of oscillatory processes, which are extremely important not only in mechanics but also in electrical engineering.

Try to perform the experiments that follow and explain the observed phenomena. Although the experiments themselves are simple, it isn't easy to explain all the details of a pendulum's motion. To do so you should know the basic laws of harmonic oscillation and have some practice with trigonometric functions. Advanced students should understand this paper, and younger students can do the experiments and explain some of their observations.

Y-suspension pendulum

Make a pendulum with two points of support as shown in figure 1. The length of the pendulum's string l should be much larger than the length of the suspension string in the upper, forked part of the pendulum.

Push the pendulum at some angle to the figure's plane. You will see that at first the pendulum swings in the direction of the initial impulse. However, after a while its motion gradually transforms into revolutions about the vertical axis. Then the pendulum again oscillates in a plane, but this time it is not in the plane of the initial oscillations. What will occur next? Again revolutions—but in the opposite direction! Gradually the motion becomes oscillations in the plane of the initial impulse. This succession of events repeats itself many times. Can you guess the reason for the strange behavior of a Y-suspension pendulum?

Coupled pendulums

(a) Suspend two identical pendulums, each made of a nut fixed to a string, from a not-too-taut string AB (fig. 2). Pull one of the pendulums, let it go, and observe the motion of the system. Change the mass of the bobs, the length of the pendulums, and the tension of the upper string and observe the effects on the system's motion.

(b) Suspend two identical pendulums from a rigid frame. Tie them together with a horizontal thread at some height (fig. 3). Pull one of the pendulums, let it go, and observe the motion of both pendulums.

Double pendulum

Construct a double pendulum as shown in fig. 4. Both strings have the same length, but the upper bob is much heavier than the lower bob. Push the lower bob and observe the resulting motion.

Now that we have played with these pendulums, let's try to explain their oscillating motion.

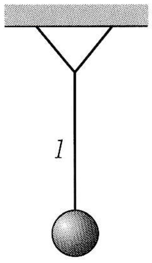


Figure 1

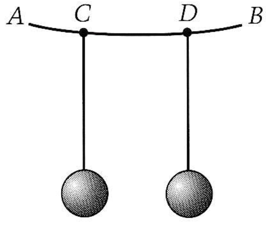


Figure 2

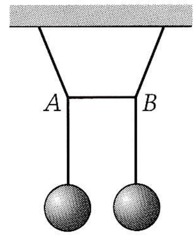


Figure 3

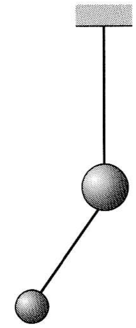


Figure 4

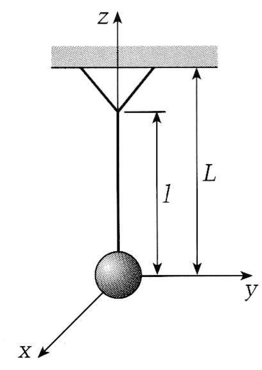


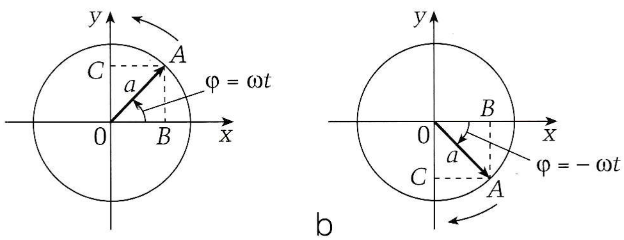
Figure 5

1. The rather complicated motion of a Y-suspension pendulum is composed of two simple oscillations: one parallel to the plane zy and another perpendicular to it (fig. 5). These oscillations operate independently of each other (provided the pendulum doesn't deviate from the vertical line too much). The respective periods are

$$T_1 = 2\pi\sqrt{\frac{l}{g}}$$

Art by V. Ivanyuk


ТИХО!
ИДУТ ОПЫТЫ



a b

Figure 6

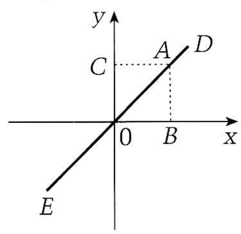


Figure 7

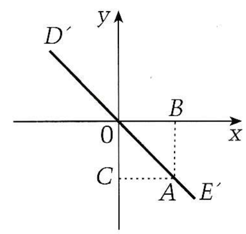


Figure 8

and

$$T_2 = 2\pi\sqrt{\frac{L}{g}}.$$

First we study how to describe the motion of the bob in our coordinate system.[1] Let point A perform circular motion with radius a and angular velocity ω either in the counterclockwise (fig. 6a) or clockwise direction (fig. 6b). The projection B of point A onto the x-axis performs a harmonic oscillation with amplitude a and angular velocity ω equal to $2\pi/T$.

The projection C of the same point A onto the y-axis performs a similar harmonic oscillation but reaches the point of maximum displacement from the equilibrium position O a quarter of a period later (that is, at the moment radius OA makes 1/4 turn around point O). You can easily observe this by watching the movements of point C and B during the rotation of radius OA. At the instant when the projection of point A onto the x-axis assumes maximum displacement from point O, its projection onto the y-axis passes point O and vice versa.

This verbal description of the complicated motion can be made much shorter in the language of mathematics. Indeed, from fig. 6a we have

$$\begin{cases} x = OB = a\cos\phi = a\cos\omega t, \\ y = OC = a\sin\phi = a\sin\omega t \\ = a\cos\left(\omega t - \dfrac{\pi}{2}\right) = a\cos\omega\left(t - \dfrac{T}{4}\right). \end{cases} \tag{1}$$

[1] L. Aslamazov and I. Kikoyin, "Wave watching," *Quantum*, January/February 1991, pp. 12–16.

These formulas are correct for *any* angle (not just for acute angles). This follows from the definitions of the sine and cosine functions.

When the motion of point A proceeds clockwise (fig. 6b), we can see that the oscillations of point C lead the oscillations of point B by a quarter period. On the other hand, we can say that the oscillations of point C lag behind those of point B by three quarters of a period. In this case the angle ωt is negative, so

$$\begin{cases} x = -a\cos(-\omega t) = a\cos\omega t, \\ y = a\sin(-\omega t) = -a\sin\omega t \\ = a\cos\left(\omega t + \dfrac{\pi}{2}\right) = a\cos\left(\omega t - \dfrac{3\pi}{2}\right) \\ = a\cos\omega\left(t - \dfrac{3T}{4}\right). \end{cases} \tag{2}$$

Formulas (1) and (2) show that the coordinates x and y vary periodically with time. In other words, points B and C perform oscillatory motion.

Now let point A perform harmonic oscillations along the line EOD (fig. 7) with frequency ω and amplitude $b = OD = OE$. In this case projections B and C simultaneously reach the most positive or negative displacements along the x- and y-axes and simultaneously pass point O. Thus we would say both points oscillate "in phase." Assume for simplicity's sake that $\angle DOx = 2\pi/4$. In this case the amplitudes of oscillation of points B and C are identical and equal to

$$b\cos\frac{\pi}{4} = b\sin\frac{\pi}{4} = \frac{b}{\sqrt{2}}.$$

Since $OB = b\cos\omega t$, we get

$$x = y = \frac{b}{\sqrt{2}}\cos\omega t. \tag{3}$$

When point A oscillates along the line $E'D'$ (fig. 8), then at the moment of point B's maximum positive displacement along the x-axis, point C reaches its maximum negative displacement along the y-axis. In this case points B and C oscillate completely out-of-phase. In other words, the oscillations of point C lag behind those of point B by half a period (or, equivalently, leads them by the same value).

Then

$$\begin{cases} x = \dfrac{b}{\sqrt{2}}\cos\omega t, \\ y = -\dfrac{b}{\sqrt{2}}\cos\omega t = \dfrac{b}{\sqrt{2}}\cos(\omega t - \pi) \\ = \dfrac{b}{\sqrt{2}}\cos\omega\left(t - \dfrac{T}{2}\right). \end{cases} \tag{4}$$

Let's recall the Y-suspension pendulum. If its bob is displaced by the distance a in the direction OD (fig. 7) and then set free, its coordinates will change according to the following formulas:

$$\begin{cases} x = a\cos\omega_1 t, \\ y = a\cos\omega_2 t, \end{cases} \tag{5}$$

where

$$\omega_1 = \frac{2\pi}{T_1} = \sqrt{\frac{g}{l}},$$

$$\omega_2 = \frac{2\pi}{T_2} = \sqrt{\frac{g}{L}},$$

$$a = \frac{b}{\sqrt{2}}.$$

Remember, the lengths l and L are almost equal, so the frequencies ω_1 and ω_2 are the same to a first approximation. Therefore, at the very beginning, the oscillations along the x- and y-axes are almost in phase.

Since $\omega_2 > \omega_2$, the y-oscillation increasingly lags behind the x-oscillation over time. We reasoned above that in the circular counterclockwise motion of a point, its projections onto the x- and y-axes oscillate with a quarter-period phase difference. Thus, when the phase difference of the x- and y-oscillations of the bob reach a quarter of the period, the bob will just move along a circle in the counterclockwise direction. When the phase lag is half of the period, the pendulum oscillates along the line $E'D'$. The phase lag steadily increases, so there will be a moment when the y-oscillation lags behind the x-oscillation by three-quarters of the period, which corresponds to the clockwise circular motion of the bob.

Finally, the phase difference will equal the period itself—in this case the bob will oscillate along the line ED. This is what we have observed in the experiment with the Y-suspension pendulum.

Our story can be retold in the language of formulas. Equations (5) can be rewritten in the form of

$$\begin{cases} x = a\cos\omega_1 t, \\ y = a\cos(\omega_1 t - \phi), \end{cases} \tag{6}$$

where $\phi = (\omega_1 - \omega_2)t$. The phase difference ϕ will not change markedly during a few periods, but since it steadily increases, it will be noticeable after many periods. Inserting the phase values $\phi = 0$, $\pi/2$, π, and $3\pi/2$ into (6) results in formulas (3), (1), (4), and (2), respectively. When $\phi = 2\pi$, the bob again oscillates along the line ED.

The theory we have described can be experimentally checked. Indeed, we can calculate what time is necessary for the bob to pass through the entire cycle. If this time is T_0, then $\phi_0 = (\omega_1 - \omega_2)T_0 = 2\pi$, from which we get $2\pi/T_0 = \omega_1 - \omega_2$, or

$$\frac{1}{T_0} = \frac{1}{T_1} - \frac{1}{T_2}. \tag{7}$$

Now we should measure the periods T_1 and T_2 of the pendulum's oscillations, first forcing it to swing in the plane yz and then in the plane xz. The period T_0 is measured independently (it doesn't depend on the way we initiated the oscillations). Check formula (7) by measuring T_2, T_1, and T_0. To determine the value of T_1 (or T_2), measure the time period necessary for, say, 10 oscillations. In performing these routine measurements, note the factors that cause the experimental results to deviate from the theoretical results. It seems that the primary reason is the decrease in amplitude caused by air resistance. Thus, the bobs in our pendulums should not be too light.

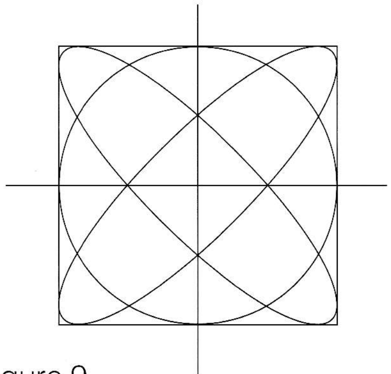


Figure 9

In reality, the movement of a Y-suspension pendulum proceeds not along a circle or a line but along a rather complicated trajectory (fig. 9), which more or less uniformly "fills" an entire square.

If the initial displacement of the pendulum forms an angle with the plane of symmetry that differs from $\pi/4$, the characteristic "instantaneous" trajectories of the bob are the diagonals of a rectangle and the ellipses inscribed in it. The fact that the motion of a Y-suspension pendulum is composed of independent oscillations is referred to by saying that its oscillations obey the principle of superposition.[2]

2. Now let's explain the oscillations of the coupled pendulums (figs. 2 and 3). If one of them is initially displaced and then let go, the other pendulum will begin to oscillate with gradually increasing amplitude. This is a result of the displacement of the string AB caused by the oscillations of the first pendulum. When the string AB is displaced, the elastic force affects the second pendulum and imparts acceleration to it. Thus, energy is transferred from the first pendulum to the second. As a result, the oscillations will decrease in one pendulum and increase in the other. Finally, the first pendulum will stop. At this instant the amplitude of oscillation of the second pendulum reaches a maximum. Then the first pendulum will swing with increasing amplitude, and the second one will stop, and so on....

The complicated oscillatory motion of the system composed of two pendulums can be described as a result of the summation (superposition) of two oscillations with frequencies ω_1 and ω_2. The first of these is the frequency for symmetric oscillation, which occurs when both pen-

[2]Wave interference was considered in contest problem by Arthur Eisenkraft and Larry D. Kirkpatrick, "Rising Star," *Quantum*, September/October 1994, pp. 44–47.

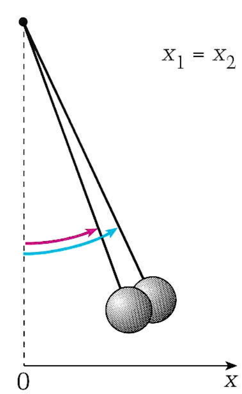


Figure 10

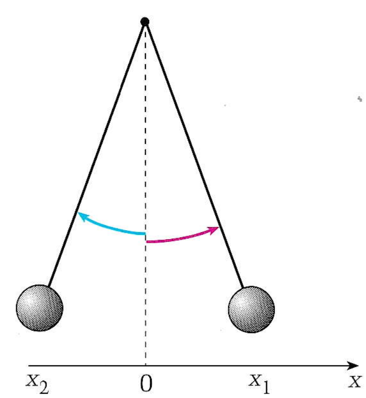


Figure 11

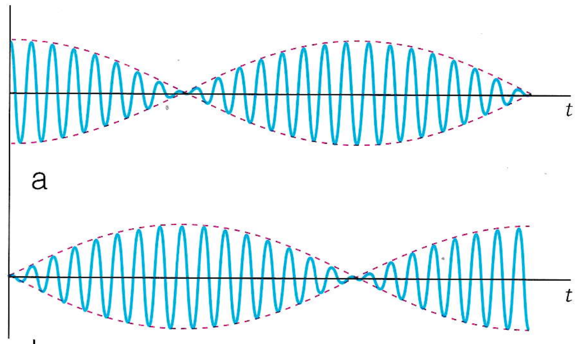


Figure 12

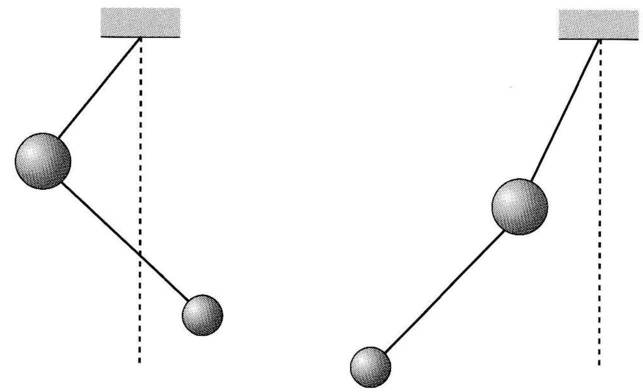


Figure 13

dulums are displaced to the same side and simultaneously let go (fig. 10). The second frequency describes the asymmetric oscillation, which occurs when the pendulums are displaced in opposite directions and let go simultaneously (fig. 11).

While watching the coupled pendulums, note that the angular frequency ω_2 is larger than ω_1. These two kinds of oscillation are known as the normal modes of coupled pendulums. Let's write down the systems of equations describing the normal modes. The first is

$$\begin{cases} x_1' = a\cos\omega_1 t, \\ x_2' = -a\cos\omega_1 t, \end{cases}$$

and the second is

$$\begin{cases} x_1'' = a\cos\omega_2 t, \\ x_2'' = a\cos\omega_2 t, \end{cases}$$

Let's add the respective displacements x_1' and x_1'' as well as x_2' and x_2'':

$$\begin{cases} x_1 = x_1' + x_1'' = a\cos\omega_1 t + a\cos\omega_2 t, \\ x_2 = x_2' + x_2'' = -a\cos\omega_1 t + a\cos\omega_2 t \end{cases}.$$

With the help of the cosine addition formula, we get

$$\begin{cases} x_1 = 2a\cos\left(\dfrac{\omega_1 - \omega_2}{2}t\right)\cos\left(\dfrac{\omega_1 + \omega_2}{2}t\right), \\ x_2 = 2a\sin\left(\dfrac{\omega_1 - \omega_2}{2}t\right)\sin\left(\dfrac{\omega_1 + \omega_2}{2}t\right) \end{cases}.$$

The formula for x_1 can be rearranged as

$$x_1 = A\cos\Omega t,$$

where

$$A = 2a\cos\left(\frac{\omega_1 - \omega_2}{2}t\right)$$

and

$$\Omega = \frac{\omega_1 + \omega_2}{2}.$$

Since ω_1 is almost equal to ω_2, the expression

$$\cos\left(\frac{\omega_1 - \omega_2}{2}t\right)$$

varies very slowly. Initially it is close to 1 (when $\omega_1 = \omega_2$, this expression always equals 1). Therefore, the factor

$$2a\cos\left(\frac{\omega_1 - \omega_2}{2}t\right)$$

in the formula for x_1 can be considered to be a slowly varying amplitude (the red dashed line in fig. 12a). Similar reasoning is valid for x_2. Figures 12a,b show the graphs of the oscillations described by these formulas. Such motions are known as beating (or modulated) oscillations. The period of beating equals half the period of the sinusoid (the dashed curve) and is determined by the formula

$$\frac{\omega_1 - \omega_2}{2}T_0 = \pi$$

or

$$\frac{1}{T_0} = \frac{1}{T_2} - \frac{1}{T_1},$$

where T_1 and T_2 are the periods of the normal modes. The frequency of beating is equal to the difference of the frequencies of the normal modes. The closer T_1 and T_2 are to each other, the larger the period.

3. Beating also occurs in the motion of a double pendulum. The light pendulum either swings with large amplitude or almost stops. The movement of the large pendulum can go almost unnoticed. Such beatings are explained by additions of the two normal modes shown in figure 13. Explain this motion on your own.

Problems

1. Derive a formula for the period of the oscillation of the mathematical pendulum, considering its motion as a rotation about the vertical axis.

2. Experimentally find the length of the Y-suspension pendulum necessary to make the bob draw a figure eight.

3. Adjust the lengths of two independent pendulums in such a way that one could measure the periods of 10 s, 20 s, and so on without counting every swing. Such pendulums can be used to measure a person's heart rate after physical exercise. (Galileo Galilei himself measured the period of oscillation of a chandelier in a church by using his pulse as a clock.)

Symmetry in algebra

Getting started with group theory

by Mark Saul and Titu Andreescu

SYMMETRY IS A FUNDAMENTAL mathematical concept. The study of symmetry, which is called group theory, has been a productive area of mathematical research for two centuries, and its treasury of uses and results shows no sign of being depleted.

In geometry, the symmetry in certain figures strikes the eye immediately, and the difficulty lies in harnessing it to achieve certain results. This is rarely the case in algebra. Algebraic symmetry appeals to the mind, not the eye, and reveals itself only slowly as one works through a series of problems.

Example: Solve the following system of equations:

$$\begin{cases} x+5y=9 \\ 5x+y=15. \end{cases}$$

Following the usual textbook solution, one would multiply one of the equations by 5, then subtract. This will of course get us the answer, and the method generalizes to any pair of simultaneous linear equations, and to simultaneous equations with more variables.

Or, we could solve one equation for x and substitute into the other equation. This method also generalizes for any pairs of simultaneous linear equations (although it gets difficult when we involve more variables).

But here's a more subtle way to solve this system, a way that generalizes in another direction: We have $x + 5y = 9$ and $5x + y = 15$. Adding, we find that $6x + 6y = 24$, so $x + y = 4$. Then subtract this from the first equation to get $4y = 5$, or $y = 5/4$. Finally, we subtract this from the second equation to get $4x = 11$, so $x = 11/4$.

Why does this method work? Because the left-hand side of the two equations are symmetric in x and y: the two variables play the same roles. (Of course, if the variables are given values, and the arithmetic operations are carried out, the results will be different. This is why the right-hand sides of the two equations are not the same.) In higher mathematics, this concept of algebraic symmetry is made even more precise.

The following problems can be thought of as a generalization of the preceding one. In general, if we perceive algebraic symmetry in a system of equations, we should act on them to preserve this symmetry.

1. Solve simultaneously:

$$\begin{cases} x+2y+z=14 \\ 2x+y+z=12 \\ x+y+2z=18. \end{cases}$$

2. Solve simultaneously:

$$\begin{cases} x+y=7 \\ y+z=-2 \\ z+x=9. \end{cases}$$

3. Solve simultaneously:

$$\begin{cases} xy=6 \\ yz=-2 \\ zx=10. \end{cases}$$

4. Solve simultaneously:

$$\begin{cases} (x+1)(y+1)=24 \\ (y+1)(z+1)=30 \\ (z+1)(x+1)=20. \end{cases}$$

5. Solve simultaneously:

$$\begin{cases} xy-x-y=11 \\ yz-y-z=14 \\ zx-x-z=19. \end{cases}$$

6. Solve simultaneously:

$$\begin{cases} x(x+y+z)=4 \\ y(x+y+z)=6 \\ z(x+y+z)=54. \end{cases}$$

7. Solve the system:

$$\begin{cases} x+[y]+\{z\}=1.1 \\ \{x\}+y+[z]=2.2 \\ [x]+\{y\}+z=3.3. \end{cases}$$

Here, the notation $[x]$ means "the greatest integer not exceeding x," and $\{x\}$ means "the fractional part of x," that is, $\{x\} = x - [x]$.

8. If a is a given positive real number, solve simultaneously:

$$\begin{cases} x^2 - xy = a \\ y^2 - xy = a(a-1). \end{cases}$$

9. Solve the following system of n equations in n variables (where n is some integer greater than 2).

$$\begin{cases} x_1 + x_2 + x_3 + \ldots + x_n = 1 \\ x_1 + x_3 + x_4 + \ldots + x_n = 2 \\ x_1 + x_2 + x_4 + \ldots + x_n = 3 \\ \ldots \\ x_1 + x_2 + x_3 + \ldots + x_{n-1} = n. \end{cases}$$

10. A triangle has sides of lengths 13, 14, and 15. Its inscribed circle divides each side into two segments, making six segments in all. Find the length of each segment.

11. The three altitudes of acute triangle ABC (with sides a, b, and c) determine six segments along the triangle's sides (see figure 1). If we let $x = \cos A$, $y = \cos B$, and $z = \cos C$,

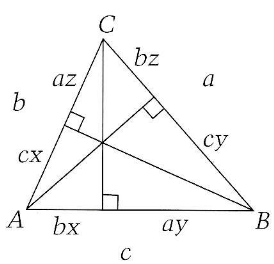


Figure 1

then the trigonometry of the right triangle lets us represent the six segments as shown in figure 1. Clearly, we have

$$\begin{cases} ay + bx = c \\ cx + az = b \\ bz + cy = a. \end{cases}$$

Solve this system of simultaneous equations for x, y, and z in terms of a, b, and c.

ANSWERS, HINTS & SOLUTIONS ON PAGE 52

CONTINUED FROM PAGE 27

technique. Copies were printed and distributed to each student, and often passed far beyond the school walls. The influence of these lectures was great: The majority of French mathematics textbooks in the first half of the nineteenth century were based on them. The school's leadership did all it could to kindle the intellectual curiosity and scientific creativity in its charges.

On the 18 Brumaire of the year VII (9 November 1799), a 30-year-old commander of the troops of the Paris military district, General Napoleon Bonaparte, carried out a coup d'état and became the First Consul of the Republic. Napoleon was a talented military leader, but also a member of the Institute de France[1] in the division of mechanics. He was interested in mathematics, had studied astronomy, and was the author of an article on ballistics. It was important for him to obtain the support of the scientists and students of the Ecole Polytechnique. Soon after he came to power, he conferred the title of Senator on three of the school's professors—Berthollet, Monge, and Laplace. A new school charter was adopted, which persisted until 1853.

In May 1804, by decision of the Senate, Napoleon was proclaimed "Emperor of the French by God's mercy and the establishment of the Republic," and France became an empire. Napoleon reorganized the Ecole Polytechnique, subordinating it to the military. In fact, even before 1804, the students' status hadn't differed much from that of cadets in military schools. Established during the most difficult years of the revolution, the school was created by military order, and now, on July 16, 1804, its military status was made official. The students were formed into a battalion of five companies under the command of a general. They were considered to be in military service, and their dormitories were considered barracks. They were paid the salary of an artillery sergeant. The school was moved from the Bourbon Palace to the reconstructed buildings of two older schools: the College of Navarre and the College de Boncoeur.

Napoleon plunged France into a more or less permanent state of war, which preoccupied students and teachers and led to a frequent reduction in examination requirements, advanced courses, and so on. Being well aware of the school's military value, Napoleon exempted the students from onerous military duties. Thus, the school was able to continue its growth and development.

On presenting the draft of the law establishing the school to the Convention in 1794, Fourcroy, a member of the Committee on Public Safety, declared, "Without hesitation I predict that the new school will bring glory to France." These words turned out to be prophetic. The Ecole Polytechnique became one of the most important factors in scientific progress in the nineteenth century. Its prestige was so great that its alumni, regardless of whatever other position they had achieved, would often sign themselves "former student of the Ecole Polytechnique." The school created a scientific elite based on personal ability and talent rather than social class.

A complete list of all the outstanding figures of science, military science, and technology among its alumni is impossible to give. Among them we find the French mathematicians Cauchy, Hermite, Jordan, and Poincare; physicists Arago and Fresnel; the entire Becquerel dynasty; the astronomer Le Verrier; the chemist Guy-Lussac; the philosophers Comte and Sorel; and field marshals Joffre and Foch. To this day, this unique educational establishment continues to train France's elite science and engineering students. For over 200 years, the school has lived up to its motto, "For the fatherland, sciences, and fame."

[1]The Institute de France, founded in 1795, incorporated five national academies, including the Academy of Science.

AT THE BLACKBOARD II

The horrors of resonance

Are you in for a rough landing?

by A. Stasenko

ENGINEERING STUDENTS are often introduced to the subject of resonance with the surprising fact that soldiers marching in step across a bridge can cause the bridge to collapse. Thus, officers generally order their troops to break step when crossing bridges.

A related problem can afflict airliners rolling on a runway. However hard the builders may work, they will not produce an absolutely even runway. Therefore, an airplane may "jump" on it when landing too quickly. Let's consider this situation.

Let a plane of mass m move with a constant velocity v and touch the runway with its wheels (fig. 1). Each wheel is supplied with a spring that has a spring constant k. The springs have length H when relaxed. If the height of the airplane's center of mass is y at

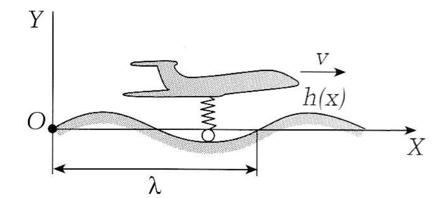


Figure 1

any given moment and the size of the runway's unevenness is h, the deformation of a spring will be $\Delta y = y - H - h$. Therefore, the spring develops the elastic force

$$F = -k\Delta y = -k(y - H - h).$$

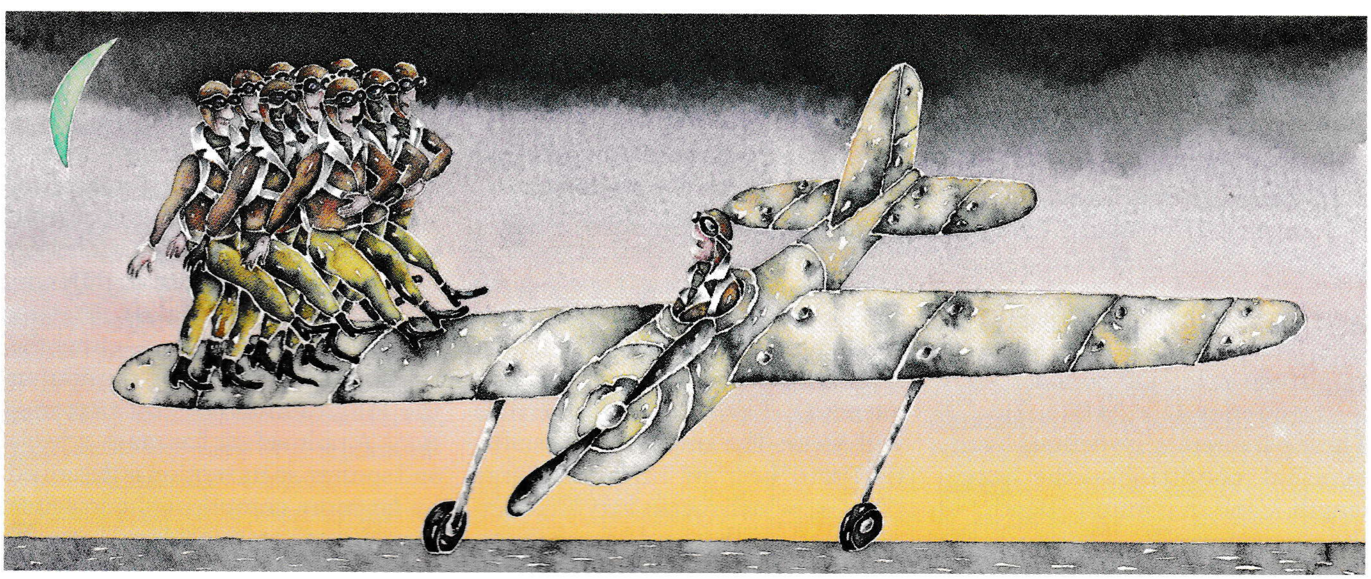
The minus sign indicates that the direction of the elastic force acting on the plane is opposite to the spring's deformation Δy. When the spring is stretched, the force is directed downward, but when it is compressed, the force is directed upward. Therefore, this elastic force tries to restore the equilibrium position (and thus is called the *restoring force*).

Now we write Newton's second law describing the vertical motion of the airplane as $ma_y = -mg + 2F$. The 2 appears on the right because an airplane has two sets of wheels. Inserting the expression for the elastic force into this equation results in

$$ma_y = -mg - 2k(y - H - h).$$

We can immediately see a special case of equilibrium when the plane rests motionless on the runway (assume, for simplicity's sake, that $h = 0$ at this point). Then the plane's acceleration is $a_y = 0$, and the last equation yields

$$y_0 - H = -\frac{mg}{2k}.$$



Art by Sergey Ivanov

for the static deformation of the spring. This value is negative because the spring is compressed under the plane's load.

Next we measure the vertical displacement of the plane's center of mass relative to the equilibrium position: $Y = y - y_0$. We further simplify the equation by dividing both sides by the mass m:

$$a_y = -\frac{2k}{m}(Y - h).$$

Now we proceed with the transformation of this equation. First, we recall that acceleration is the second derivative of the displacement with respect to time: $a_y = Y''$. Next we assume that the runway unevenness is a harmonic function with wavelength λ along the x-axis and amplitude h_0:

$$h = h_0 \sin\left(2\pi\frac{x}{\lambda}\right).$$

The constant horizontal velocity corresponds to the equation $x = vt$, and we denote the combination of positive values as

$$\omega_0^2 = \frac{2k}{m}.$$

Now we can rewrite the equation in the form

$$Y'' + \omega_0^2 Y = \omega_0^2 h_0 \sin\left(2\pi\frac{v}{\lambda}t\right).$$

If the right side of this equation were zero, most readers would recognize it as the equation for harmonic oscillation with natural frequency

$$\omega_0 = \sqrt{\frac{2k}{m}}.$$

However, we have a nonzero right side: The harmonic function with amplitude $\omega_0{}^2 h_0$ and period $T = \lambda/v$ (or frequency $\Omega = 2\pi/T = 2\pi v/\lambda$). These parameters are defined by the external conditions—the wavelength λ and maximum "depth" of the unevenness h_0. Therefore, the

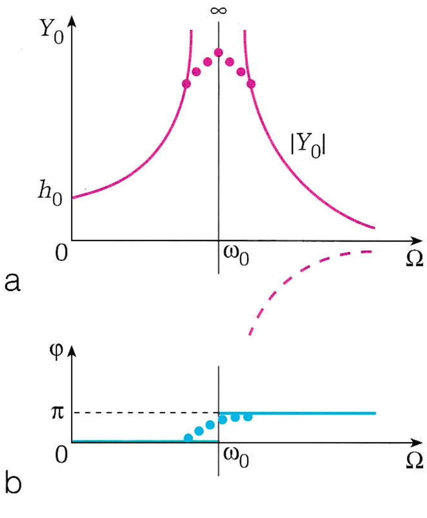


Figure 2

resulting oscillations are called *forced oscillations.*

Let's find out how this oscillatory system (an airplane with two springs) responds to the runway's unevenness. We look for a solution in the form of harmonic oscillations occurring with the frequency of the externally applied force Ω:

$$Y = Y_0 \sin \Omega t.$$

After taking the derivative of this equation two times (to obtain $Y'' = -\Omega^2 Y_0 \sin \Omega t$), inserting it into the motion equation, and canceling out $\sin \Omega t$, we get an equation for the amplitude Y_0:

$$Y_0(-\Omega^2 + \omega_0^2) = \omega_0^2 h_0.$$

Figure 2a qualitatively shows the dependence of the oscillation amplitude Y_0 upon the frequency of the external excitation Ω. When Ω tends to zero—when the plane's velocity is small or the runway is flat ($\lambda \to \infty$)—Y_0 tends to h_0. So when the plane's velocity is small or the wavelength of the unevenness is very large, the plane moves steadily along the runway without much vertical motion. However, if the runway's wavelength and the plane's velocity are such that the frequency of the forced oscillation approximately equals the natural frequency $\Omega = 2\pi v/\lambda$, something terrible will occur: The amplitude of oscillation will become infinitely large ($Y_0 \to \infty$), which means very large vertical motions that could cause the pilot to lose control of the airplane. This is a case of *resonance.*

At large frequencies ($\Omega > \omega$), the value of Y_0 becomes negative (the dashed line in fig. 2a), but we can "hide" the minus sign in the argument of the sine function

$$-|Y_0|\sin \Omega t = |Y_0|\sin(\Omega t + \pi).$$

In other words, the phase of oscillation ϕ changes by π when the external force oscillates in the vicinity of the natural frequency of the plane's oscillations—that is, when $\Omega = \omega_0$ (fig. 2b).

Of course, scientists and engineers do their best to avoid the amplitude resonance catastrophe (excluding the possibility of $|Y_0| \to \infty$). One way to do so is to employ frictional vibration damping. For example, an oil shock absorber, a cylinder containing oil and a piston, could be connected in series with the spring. This helps when the natural sink of energy, friction, is too small to damp the oscillations. In this case we insert into the motion equation a dissipative force (resulting in dissipation of mechanical energy by transformation into heat), and now $|Y_0|$ will not tend to infinity (see the dotted curve in fig. 2).

Also, there are two real-world reasons why airliners avoid the resonance amplitude catastrophe. First, the bumpy surface of the runway will probably not be a strictly periodic function (with constant wavelength λ). Second, planes don't land or take off with a constant velocity; they accelerate quickly before takeoff and brake rapidly after landing.

However, for land vehicles (say, a raiload car), both λ (the rail's length) and v (a train's velocity) are rather stable, so sometimes you can observe the resonance. The raiload car begins to "jump" and oscillate in different ways: strictly vertical or rotating about the horizontal axes—either the transverse one (pitching motion) or the longitudinal one (rocking motion). A railroad car has many wheels and springs, so its motion is described by far more complicated equations than the one we considered here. ◘

IN THE OPEN AIR

Why is the sky blue?

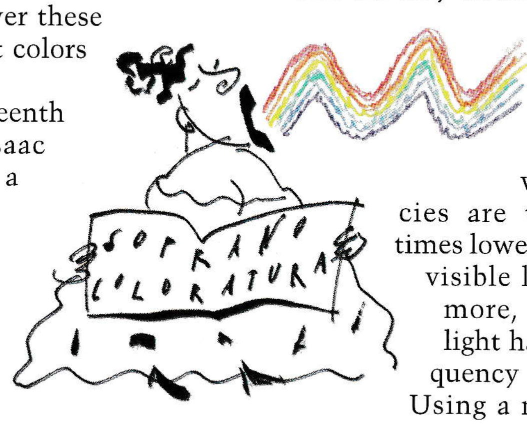
Is it just for the atmosphere?

by Alexander Buzdin and Sergei Krotov

HAVE YOU EVER WONDERED why the sky is blue? For many years inquisitive people have asked this question and many others, such as How many colors exist in the world? and Why are they different? Thanks to the work of scientists, we now know enough to answer these questions about colors and light.

In the seventeenth century Sir Isaac Newton made a great contribution to the development of the study of colors. He observed that solar (white) light separated into many colors when it passed through a glass prism. The resulting spectrum exactly corresponded to the colors of a rainbow. Thus Newton produced a laboratory-made rainbow with the colors **red**, **orange**, **yellow**, **green**, **blue**, **indigo**, and **violet**. (An easy way of remembering the order of the colors is to use the name Roy G. Biv as a mnemonic.) The experiments of Newton helped us realize that white light is a mixture of its component colors. In particular, Newton demonstrated that a mixture of the rays of a separated white beam is also white.

Much more time went by before scientists discovered that light is the oscillations of tightly connected electric and magnetic fields spreading in space—in other words, electromagnetic waves. Another common example of an electromagnetic wave is a radio wave. The nature of light and radio waves is the same, and the only distinction between them is the frequency of electromagnetic oscillation: Radio wave frequencies are thousands of times lower than those of visible light.[1] What's more, each color of light has its own frequency of oscillation. Using a musical analogy, red corresponds to deep (bass) tones and violet light to high tones.



Art by Michael Jones

The speed of light, c, equals $3.0 \cdot 10^8$ m/s. It's many thousands of times larger than the velocity of sound waves, so a person listening to the radio in Seattle actually hears a musician playing in Moscow earlier than a Muscovite concertgoer. This refers to its speed in a vacuum. When light passes through a transparent medium, its speed is slower. Of particular importance is the fact that different colors have different speeds in a transparent medium. This phenomenon, called dispersion, makes it possible to separate white light into its component colors by passing it through a prism.

Any color of the rainbow can be obtained from white light. But why do the objects have the colors they do? If we were to take a yellow ball into a dark room, we would not see any color; the ball does not radiate colored light. To see the ball's color, we must illuminate the ball. When the ball is illuminated, the incident light is partially absorbed and partially reflected, and we see only the reflected part. Due to their individual molecular structures, different bodies absorb or reflect light from various spectral ranges differently.

Let's take a tomato, for instance. At different stages of ripening it will predominantly reflect either green or red light rays. This happens because of the molecular rearrangements in the tomato as it ripens. (It's no coincidence that chemistry considers color an important characteristic of a substance.)

Colors on command

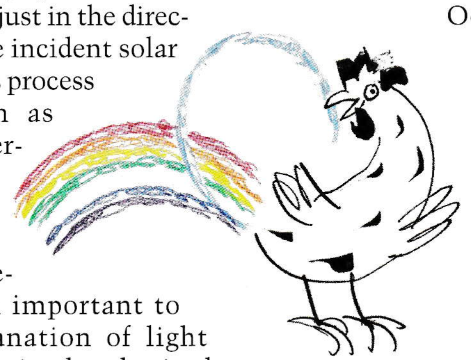
"The sky above is azure-blue," wrote the Georgian poet Nico Baratashvily. But is this line correct from a physical point of view? Usually the sky's blue color is explained by light scattering in the atmosphere. Why, then, isn't the night

[1]See also A. Leonovich "Surfing the electromagnetic spectrum," pp. 32–33 in the January/February 1995 issue of *Quantum*.

sky blue when the Moon is full? And why do different parts of the sky have different tints of blue—some bright and others dull? And what happens to the sky at sunset?

An attentive person will see that at sunset the western part of the sky becomes slightly tinted with yellow and orange; then, when the Sun is fiery red, the sky changes from yellow-orange to bright red; and finally the sky is painted in magenta up to the angular altitude of about 25°. Since the reason for this dramatic performance is sunlight scattering in the atmosphere, let's look at light scattering in more detail.

The explanation of the celestial colors was given by the English physicist J. W. Rayleigh. In simplified form it looks like this: The color of the sky is determined by the fact that different frequencies of light scatter differently. The electromagnetic wave "rocks" the electrons in the air molecules. Rays at the violet end of the spectrum have the strongest influence upon the electrons. Thus, the electrons of the air molecules "capture" the oscillation energy of the blue part of the spectrum from the incident solar wave. This generates extra motion for the electrons—the so-called forced oscillations. But oscillating electrons themselves radiate electromagnetic waves. However, this secondary radiation propagates in all directions, not just in the direction of the incident solar light. This process is known as the scattering of light.



Another phenomenon important to the explanation of light scattering in the sky is the heterogeneous distribution of air molecules in the atmosphere, as seen in the persistent fluctuations of air density. Indeed, were air molecules distributed homogeneously in the sky, the scattering would be quite different, and the sky would be jet black.

So air molecules scatter most of the blue part of the light spectrum, and these regions of the sky are perceived as blue or light blue. (Here a question may arise—why blue and not violet? The reason is twofold: first, the human eye is not very sensitive to violet light, and second, solar light has "fewer" violet rays than blue ones.)

The longer the atmospheric path of solar light, the fewer blue rays in it. This explains why the setting Sun is reddish orange. In the evening the Sun's rays travel much farther through the atmosphere than at midday, when they come from directly above.

Particle persuasion

It's clear that smoke, dust, and other tiny particles suspended in the air notably affect the light scattering in the atmosphere. After powerful volcano eruptions, sunrises and sunsets display wonderful colors—the Sun and Moon can even be blue! Here's a description of the effects produced by the catastrophic eruption of Krakatau in 1883, given by the Russian scientist V. A. Obruchev: "The fine ashes screened the Sun in Japan and other places at distances of more than 3,000 km. These ashes floated in the atmosphere for a long time and caused the bluish tint of solar and lunar discs seen from Africa and the Pacific Ocean islands, as well as the splendid red dawns observed everywhere on the Earth at the end of 1883 and in early 1884."

In this case the blue color of the Sun and Moon resulted from light scattering on the atmospheric aerosol composed of particles ranging from 0.4 to 0.9 μm and thus comparable in size to the wavelength of visible light. Because of their relatively large size, they scatter light at the red end of the spectrum more strongly than at the violet end. The Sun and Moon observed through such an aerosol are seen as bluish discs because after the scattering of the red constituent of white light, only the blue rays reach the human eye.

Sometimes a beautiful, mysterious, bluish haze hovers over a green, open space not too spoiled by industrial activity. The Blue Mountains in Australia and the Blue Ridge Mountains in the eastern United States both are named for and famous for their bluish haze. Their color results from light scattering on tiny particles much smaller than the wavelength of visible light. These particles can be organic macromolecules emitted by the verdant surroundings or tiny fragments torn off the pointed parts of plants by atmospheric electric fields. The extra scattering at the blue end of the spectrum arises only where such particles accumulate in the atmosphere.

So the colors that paint the celestial sphere are caused by the combination of Rayleigh scattering and light scattering on small, suspended particles. It's comforting to know that the "azure-blue" sky does have a sound physical basis behind it. Here are some related topics to consider:

1. Why does a Christmas tree decorated with different-colored lights look red from a distance in the evening?

2. The next time you're sitting near a bonfire, ask your friends why the smoke looks blue against the trees (near the ground) but once high above them turns yellow against the sky?

3. Let a few drops of milk fall into a glass of water and watch a lamp through it—it looks reddish orange. However, the light emitted from the side of the glass is light blue. Explain the difference in colors.

ANSWERS, HINTS & SOLUTIONS

Math

M226

For every square of the board we can point out its "predecessor" squares—the squares from which the king can move directly to the square. Clearly, the number of different paths that lead to a square is the sum of the number of different paths that lead to each of its predecessors. Now we can write "1" in the lower left square of the board and then fill in the rest of the squares (fig. 1).

1	11	61	231	681	1683
1	9	41	129	321	681
1	7	25	63	129	231
1	5	13	25	41	61
1	3	5	7	9	11
1	1	1	1	1	1

Figure 1

M227

Suppose first that there are infinitely many gangsters, all of whom have a contract out on some gangster Al. Then the gangsters who are out for Al solve our problem.

Now suppose that there is no such Al—that is, each gangster is being sought only by a finite number of others. We construct the required subset by induction.

We choose one gangster, Bill, and kill all his enemies. Bill will be the first member of our subset. Of the surviving gangsters, we select another. Since we have killed finitely many gangsters (Bill's enemies are finite), and there are infinitely many left, we can choose one, say Charlene, who is not Bill's target. Charlene also cannot be Bill's enemy because these are already dead. Now we kill all Charlene's enemies. Then we choose a third gangster Dean, who is not sought by either Charlene or Bill, and kill his enemies, and so on.

It is not hard to see that this process can be continued indefinitely. Indeed, suppose n gangsters are already chosen. Then we've killed off only finitely many gangsters (we have supposed that no gangster has infinitely many enemies), so there are infinitely many others left, and one can choose an $(n + 1)$st gangster who is not the target of any of the preceding n. The set {Bill, Charlene, Dean ... } meets the requirements of the problem.

M228

The function on the left side of this inequality is defined in the square $-1 \le x \le 1$, $-1 \le y \le 1$. Let $M = (x_0, y_0)$ be a point of this square at which the expression on the left side is strictly positive. That is,

$$(y_0^3 - \arcsin x_0)(x_0^3 + \arcsin y_0) > 0.$$

Then, for the point $M' = (y_0, -x_0)$ we'll have

$$(-x_0^3 - \arcsin y_0)(y_0^3 - \arcsin x_0)$$
$$= -(y_0^3 - \arcsin x_0)(x_0^3 + \arcsin y_0) < 0.$$

In other words, if point M belongs to the figure determined by the inequality, then its image M' under a clockwise rotation by 90° about the origin does not belong to the figure. Conversely, a counterclockwise rotation by 90° about the origin maps points that do not belong to the figure into points that do belong to the figure. Thus, the area of the figure equals the area of its complement in the aforementioned square (in fact, these two sets are congruent), and this area is equal to

$$\frac{1}{2} \cdot 4 = 2.$$

(Here we've assumed that the "area" of the region where

$$(y^3 - \arcsin x)(x^3 + \arcsin y) = 0$$

is equal to zero, which is certainly true.)

M229

Consider five vectors:

$$\mathbf{a}\left(x, \sqrt{1-x^2}\right), \mathbf{b}\left(y, \sqrt{4-y^2}\right),$$
$$\mathbf{c}\left(z, \sqrt{9-z^2}\right), \mathbf{d}\left(t, \sqrt{16-t^2}\right), \mathbf{e}(6, 8).$$

The system is equivalent to the following vector equality: $\mathbf{a} + \mathbf{b} + \mathbf{c} + \mathbf{d} = \mathbf{e}$. But $|\mathbf{a}| = 1$, $|\mathbf{b}| = 2$, $|\mathbf{c}| = 3$, $|\mathbf{d}| = 4$, and $|\mathbf{d}| = 10$. That is, $|\mathbf{a} + \mathbf{b} + \mathbf{c} + \mathbf{d}| = |\mathbf{a}| + |\mathbf{b}| + |\mathbf{c}| + |\mathbf{d}|$. This can happen if and only if all the vectors **a, b, c, d,** and **e** are collinear. Now we can easily find the answer. For example, let's find x. Let $\mathbf{a} = \lambda\mathbf{e}$, so that $x = \lambda 6$, and $\sqrt{1-x^2} = \lambda 8$. Then $\lambda = 0.1$, and $x = 0.6$. Similarly, $y = 1.2$, $z = 1.8$, and $t = 2.4$.

M230

First we'll show, in this case, that D lies on ray MA (and not on ray ME, which could happen if $\angle ABC$ is obtuse). Suppose we draw a semicircle with diameter AC. Then, if

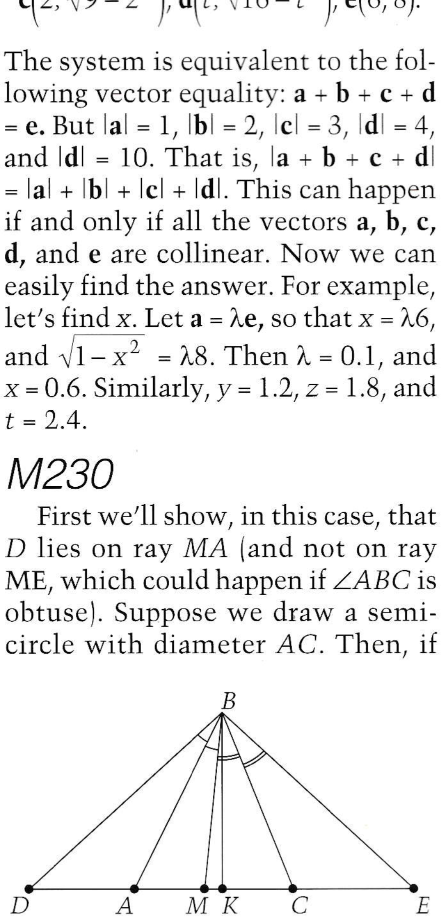


Figure 2

$\angle ABC$ is acute, point B will lie outside the semicircle. This means that $BM > AM$, so in triangle ABC, we have $\angle BAM > \angle ABM$, so D lies on ray MA. Similarly, E lies on ray MC.

We use the fact that the bisector of an angle in a triangle divides the side to which it is drawn in the ratio of the triangle's other two sides. Applying this theorem to triangles DBM and BEM, we obtain $DB/BM = DA/AM$, $BE/BM = CE/CM$. Dividing the first of these equalities by the second, we get

$$\frac{DB}{BE} = \frac{DA}{CE}. \tag{1}$$

Let the bisector of $\angle DBE$ intersect DE at the point K. Then

$$\frac{DB}{BE} = \frac{DK}{KE}. \tag{2}$$

Now if $a/b = c/d$, then

$$\frac{a}{b} = \frac{a+c}{b+c}$$

as well (the reader can prove this, for instance, by examining cross products). This implies that

$$\frac{DB}{BE} = \frac{DK - DA}{KE - CE} = \frac{AK}{KC}.$$

Since BK bisects $\angle DBE$, we can write

$$\begin{aligned}\angle DBK &= 2\angle ABM + \angle MBK \\ &= \angle EBK = 2\angle CBM - \angle MBK,\end{aligned}$$

so $2\angle ABM = 2\angle CBM - 2\angle MBK$. Dividing by 2 and simplifying, we find $\angle ABM = \angle CBK$. Let $\angle ABM = \alpha$ and $\angle MBC = \beta$. (Thus, $\angle CBK = \alpha$, $\angle ABK = \beta$).

Now we let the area of triangle XYZ be S_{XYZ}. We can write

$$\begin{aligned}\frac{AK}{KC} &= \frac{S_{ABK}}{S_{CBK}} = \frac{\frac{1}{2}AB \cdot BK\sin\beta}{\frac{1}{2}CB \cdot BK\sin\alpha} \\ &= \frac{AB^2}{CB^2} \cdot \frac{\frac{1}{2}CB \cdot BM\sin\beta}{\frac{1}{2}AB \cdot BM\sin\alpha} \\ &= k^2 \cdot \frac{S_{CBM}}{S_{ABM}} = k^2.\end{aligned}$$

Thus, $DB/BE = k^2$. The case when $\angle ABC$ is obtuse is left to the reader.

Physics

P226

Let us consider a small time interval Δt, during which the first beetle travels a distance $\vec{s}_1 = \vec{v}_1 \Delta t$, and the second one $\vec{s}_2 = \vec{v}_2 \Delta t$.

What must the displacement of the cockroach be to preserve the equilateral shape of the triangle that connects all the insects? Assume the first beetle to be motionless while the second beetle is making a step $\vec{s}_2$ (fig. 3). To preserve the equilateral shape of the connected triangle $B_1B_2'C$, the cockroach must shift its position by $\overrightarrow{CC'}$ such that

$$\left|\overrightarrow{CC'}\right| = \left|\vec{s}_2\right| = v_2 \Delta t.$$

Now let the second beetle rest while the first one changes its position by $\vec{s}_1$. In this case the cockroach must make a step $C'C''$, which is characterized by

$$\left|\overrightarrow{C'C''}\right| = \left|\vec{s}_1\right| = v_1 \Delta t.$$

The cockroach should make both steps CC' and $C'C''$ if the beetles are displaced by $\vec{s}_1$ and $\vec{s}_2$:

$$\overrightarrow{CC'} + \overrightarrow{C'C''} = \overrightarrow{CC''}.$$

The vector diagram $CC'C''$ (fig. 4) yields

$$\left|\overrightarrow{CC''}\right| \le \left|\overrightarrow{CC'}\right| + \left|\overrightarrow{C'C''}\right| = (v_1 + v_2)\Delta t.$$

The respective velocity of the cockroach is thus

$$v_c = \frac{\left|\overrightarrow{CC''}\right|}{\Delta t} \le v_1 + v_2 \le 2v.$$

The equality holds when the cockroach's displacements $\overrightarrow{CC'}$ and $\overrightarrow{CC''}$ have the same directions.

P227

By the statement of the problem, the ball rolls without slipping, so at any moment the velocities of points A and B (fig. 5), which contact the chute, are equal to zero. We assume the ball to be absolutely rigid, which means that distance between any two points of the ball doesn't vary. Therefore, all the points of segment AB are motionless at any moment in time. Thus, the motion of the ball is a rotation about the axis AB, which travels with linear velocity v.

The instantaneous velocity of any point of the ball is $\omega\rho$, where ω is the angular velocity of rotation and ρ is the distance from the point to the axis AB. The velocity of the ball's center (point O in fig. 5) is v, and the distance of point O to AB is

$$\rho_O = |OC| = R\frac{\sqrt{3}}{2}.$$

Therefore,

$$\omega = \frac{v}{\rho_O} = \frac{2v}{R\sqrt{3}}.$$

So, the points with the maximum velocity are located farthest from the axis AB. From a geometrical viewpoint it is clear that there is only one point located farthest from AB: point d in fig. 5. The respective distance from the axis AB is

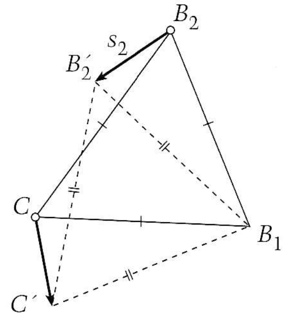


Figure 3

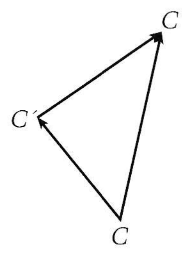


Figure 4

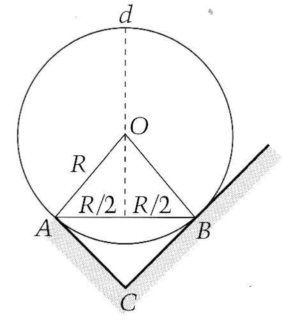


Figure 5

$$\rho_d = \rho_O + R = R\left(1 + \frac{\sqrt{3}}{2}\right),$$

so the velocity of point d is

$$v_d = v_{\max} = \omega\rho_d$$
$$= \frac{2v}{R\sqrt{3}} R\left(1 + \frac{\sqrt{3}}{2}\right) = v\frac{2+\sqrt{3}}{\sqrt{3}}.$$

P228

By the conditions of the problem, the amount of heat transferred during a small period Δt to the surface area $4\pi r^2$ of the spherical drop is

$$\Delta Q = \frac{\alpha(T_a - T_d)}{r} 4\pi r^2 \Delta t,$$

where r is the drop's radius, α is a constant coefficient depending upon the thermal conductivity of the surrounding medium, and T_a and T_d are the temperatures of air and the drop $(T_a > T_d)$. This heat is spent to evaporate some mass of liquid matter Δm. Denoting the latent heat of evaporation by L, one can write conservation of energy as

$$L\Delta m = -\Delta Q = -\alpha(T_a - T_d)4\pi r\Delta t.$$

The minus sign is included on the right-hand side to indicate that the mass of the drop decreases with time. Since $m = (4/3)\pi r^2\rho_0$, where ρ_0 is the drop's density,

$$\Delta m = 4\pi r^2 \rho_0 \Delta r.$$

Plugging this formula into the energy equation, we get

$$r\Delta r = -\frac{\alpha(T_a - T_d)}{L\rho_0}\Delta t,$$

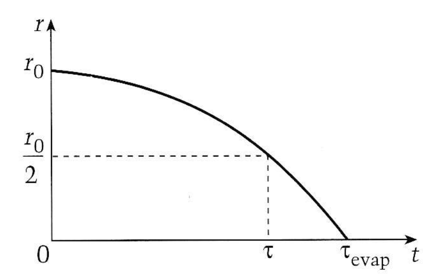


Figure 6

or

$$(r^2)' = -\beta,$$

where $\beta = 2\alpha(T_a - T_d)/L\rho_0$ is a positive constant.

Therefore, the square of the drop's radius decreases linearly with time:

$$r^2 = r_0^{\,2} - \beta t,$$

where r_0 is the drop's radius at the initial moment $t = 0$. Thus, the radius decreases (fig. 6) by the formula

$$r = \sqrt{r_0^2 - \beta t}.$$

By the conditions of the problem, the radius decreases by half during time $\tau = 10$ min = 600 s. That is, it becomes $r_0/2$, so

$$\frac{r_0^2}{4} = r_0^2 - \beta\tau,$$

$$\beta = \frac{3}{4}\frac{r_0^2}{\tau}.$$

Let's designate the duration of the drop's life (period of complete evaporation) by τ_{evap}, so $r = 0$ when $t = \tau_{evap}$, from which we get

$$\tau_{evap} = \frac{r_0^2}{\beta} = \frac{4}{3}\tau = 800\text{ s}.$$

P229

First, a note of caution: The electric resistance of an incandescent lamp is not a constant. Indeed, it varies with the voltage drop across it: The higher the voltage, the higher the temperature of the lamp's filament and thus its resistance. By the conditions of the problem, the brightness of the filament is the same for both positions of the switch, so the voltage drop V_l across the lamp and its resistance R are also the same in these cases.

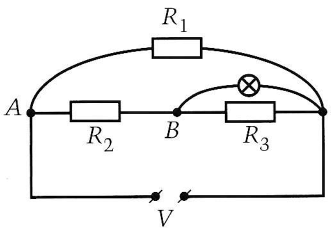


Figure 7

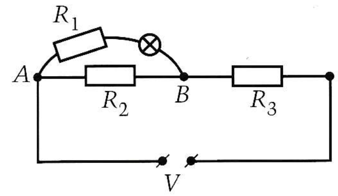


Figure 8

The equivalent circuit diagrams for closed and open positions of the switch are shown in figures 7 and 8. The resistances of subcircuits BC and AC for the closed switch are

$$R_{BC} = \frac{RR_3}{R+R_3}, R_{AC} = R_2 + R_{BC}. \quad (1)$$

Inserting the values of R_2 and R_3 (in ohms) into (1) yields

$$R_{BC} = \frac{90R}{R+90}, R_{AC} = \frac{270(R+60)}{R+90}. \quad (2)$$

The voltage drop across the lamp for the closed switch is

$$V_l = \frac{V}{R_{AC}} R_{BC}.$$

Taking into account (2) and the given value $V = 54$V, we have

$$V_l = \frac{18R}{R+60}. \quad (3)$$

A similar calculation for the open switch yields

$$V_l = \frac{36R}{R+150}. \quad (4)$$

Finally, from (3) and (4) we obtain the sought value of voltage drop across the lamp: $V_l = 6$ V.

P230

To solve this problem we shall use conservation of energy and momentum for an isolated system. Initially, before radiating a photon, the system was a motionless hydrogen atom in the excited state, which means the orbital electron occupied not the lowest energy level E_1, but some level with higher energy E_n. An atom can be excited by some external event, such as a collision with other atoms or a free electron or the absorption of a photon (a quantum of light). In the idealized case, when there is no recoil, the frequency ν_0 of the emitted light is described by the equation $E_n - E_1 = h\nu_0$. In a motionless atom the total energy is equal to the rest energy of the nucleus (proton) m_pc^2 and the energy of the selectron E_n, and the total momentum of the atom is zero.

After radiating a photon with energy $h\nu$, the isolated system includes both the photon and hydrogen atom, which acquired a certain velocity v due to recoil. In this case the total energy of the system is

$$m_pc^2 + E_1 + \frac{m_pv^2}{2} + h\nu,$$

and the total momentum of the system is

$$\frac{h\nu}{c} - m_pv.$$

According to conservation of energy and momentum, we get

$$m_pc^2 + E_n = m_pc^2 + E_1 + \frac{m_pv^2}{2} + h\nu,$$
$$0 = \frac{h\nu}{c} - m_pv.$$

Since $E_n - E_1 = h\nu_0$, we obtain

$$h\Delta\nu = h\nu - h\nu_0 = -\frac{m_pv^2}{2} = -\frac{(h\nu)^2}{2m_pc^2},$$

or

$$\Delta\nu = -\frac{h\nu^2}{2m_pc^2} = -\frac{h}{2m_p\lambda^2}.$$

For relatively small frequency deviations ($\Delta\nu << \nu$) we can write

$$\Delta\nu = \Delta\left(\frac{c}{\lambda}\right) = -\frac{c}{\lambda^2}\Delta\lambda.$$

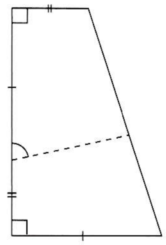


Figure 9

Finally,

$$\Delta\lambda = \frac{h}{2m_pc} = 6.7 \cdot 10^{-16} \text{ m}.$$

Brainteasers

B226

Let t be the number of doughnuts one can buy for \$1. Then $25 = t^2$, so $t = 5$. Thus, one doughnut is 20¢.

B227

Consider three groups of children: girls, boys that ride behind girls, and boys that ride behind boys. Since every girl rides behind a boy, we conclude that there is a boy after each girl (otherwise, there would be a girl after another girl). Thus, there are as many boys riding behind girls as the total number of girls. But we know that the number of boys riding behind girls is equal to the number of boys riding behind boys. So, all these groups consist of equal numbers of children. Therefore, there are 10 girls and 20 boys.

B228

Two trips to the shop took Boris $3 \times 55 - 35 - 50 = 80$ minutes. Thus, it takes him 40 minutes to buy and eat an ice cream cone and come back to school, so he has quite enough time to go to the shop again.

B229

See figure 9.

B230

Air is a poor heat conductor. By fluffing their feathers, the birds enlarge the air layer between their bodies and the atmosphere.

Gradus ad Parnassum

1. Add the three given equations to get

$$4x + 4y + 4z = 44,$$

or

$$x + y + z = 11.$$

If we subtract this in turn from each of the given equations, we find very quickly that $y = 3$, $x = 1$, and $z = 7$. The reader who has tried substitution will appreciate how much easier this solution is.

2. Adding the three equations, we find

$$2x + 2y + 2z = 14,$$

or

$$x + y + z = 7.$$

Subtracting each of the given equations in turn from this one, we find $z = 0$, $x = 9$, and $y = -2$.

3. Taking a hint from problem 2, we multiply the three equations together, to find $x^2y^2z^2 = 6 \cdot 15 \cdot 10$, so $xyz = \pm 30$. Then we divide this equation by each of the given equations to find $(x, y, z) = \pm(2, 3, 5)$.

4. Let's be quick about this. Let $p = x + 1$, $q = y + 1$, and $r = z + 1$. Then we have $pq = 24$, $qr = 30$, and $rp = 20$, and we have the same kind of equations as in problem 3. We find that $(p, q, r) = (4, 6, 5)$ or $(-4, -6, -5)$. The corresponding values for (x, y, z) are $(3, 5, 4)$ and $(-5, -7, -6)$.

5. We can make this problem resemble problem 4 by the trick of adding 1 to each side of each equation. For example, the first equation becomes

$$xy - x - y + 1 = 12,$$

or

$$(x - 1)(y - 1) = 12.$$

We then let $p = x - 1$, $q = y - 1$, $r = z - 1$, and proceed as before. We find that $(x, y, z) = (5, 4, 6)$ or $(-3, -2, -4)$

6. Adding all three equations, and factoring the left side, we find that

$(x+y+z)^2 = 64$, so $x+y+z = \pm 8$. Then we divide each of the given equations by this relation, to find

$$(x,y,z) = \left(\frac{1}{2}, \frac{3}{4}, \frac{27}{4}\right) \text{ or } \left(-\frac{1}{2}, -\frac{3}{4}, -\frac{27}{4}\right).$$

7. Add the given equations to get

$$2x + 2y + 2z = 6.6,$$

or

$$x + y + z = 3.3.$$

Subtracting the first equation from this new relation, we get $\{y\} + [z] = 2.2$. But this means that $[z] = 2$ and $\{y\} = .2$. Subtracting the new relation in turn from the other two given, we find that $[x] + \{z\} = 1.1$, so $[x] = 1$ and $\{z\} = .1$, and $\{x\} + [y] = 0$, so $\{x\} = 0$ and $[y] = 0$. Finally, we can paste the values of x, y, and z together to find $x = 1$, $y = .2$, and $z = 2.1$.

8. Adding the equations, we find

$$x^2 - 2xy + y^2 = (x-y)^2 = a^2,$$

so $x - y = \pm a$. We write the given equations as

$$x(x-y) = a, \; -y(x-y) = a(a-1),$$

and divide each in turn by $x - y$. We find that $(x, y) = (1, 1-a)$ or $(-1, a-1)$.

9. Rather than adding the equations, we take advantage of their symmetric relation to the sum

$$x_1 + x_2 + x_3 + \ldots + x_n.$$

Denoting this sum by S (forgetting for a moment that we know that it is equally numerically to 1), we can write the system as

$$\begin{cases} S - x_2 = 2 \\ S - x_3 = 3 \\ \ldots \\ S - x_{n-1} = n-1 \\ S - x_n = n. \end{cases}$$

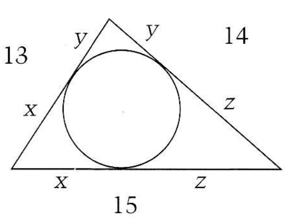


Figure 10

Then we recall that in fact $S = 1$, and see immediately that

$$x_2 = -1, x_3 = -2, x_4 = -3, \ldots, x_n = (-n+1).$$

Then, from a well-known formula,

$$x_2 + x_3 + \ldots + x_n =$$
$$-\left[1 + 2 + \ldots + (n-1)\right] = -\frac{n(n-1)}{2}.$$

Finally,

$$\begin{aligned} x_1 &= 1 - (x_2 + x_3 + \ldots + x_n) \\ &= 1 + \frac{n(n-1)}{2}. \end{aligned}$$

10. Since tangents to a circle from a point outside are equal, the segments of the sides are equal in pairs. Let their lengths be x, y, z (see fig. 10). Then we have

$$\begin{aligned} x + y &= 13 \\ y + z &= 14 \\ z + x &= 15. \end{aligned}$$

We can solve these using the method of problem 2. We find that $x = 7$, $y = 6$, and $z = 8$.

11. We cannot use the method of problem 2 (or problem 3), because the coefficients of x, y, and z do not follow the same patterns as in those problems. But suppose we divide the equations by ab, ac, and bc, respectively. We obtain

$$\begin{aligned} \frac{y}{b} + \frac{x}{a} &= \frac{c}{ab} \\ \frac{x}{a} + \frac{z}{c} &= \frac{b}{ac} \\ \frac{z}{c} + \frac{y}{b} &= \frac{a}{bc}, \end{aligned}$$

and now we can let $A = x/a$, $B = y/b$, $C = z/c$, and apply the method of problem (2). We find that

$$\begin{aligned} A &= \left(\frac{c^2 + b^2 + a^2}{2abc}\right) - \frac{a}{bc} \\ &= \frac{c^2 + b^2 - a^2}{2abc}, \end{aligned}$$

with corresponding expressions for B and C. Then we can easily find that:

$$\begin{aligned} x &= \frac{b^2 + c^2 - a^2}{2bc} \\ y &= \frac{a^2 + c^2 - b^2}{2ac} \\ z &= \frac{a^2 + b^2 - c^2}{2ab}. \end{aligned}$$

But if you remembered the law of cosines, you already knew this!

Index of Advertisers

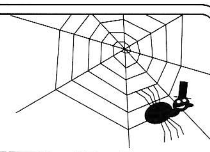
Metrologic Instruments cover 3

Talk back to us

drop us a line at *Quantum*,
1840 Wilson Blvd.,
Arlington VA 22201

or

zip us an electron or two at
quantum@nsta.org



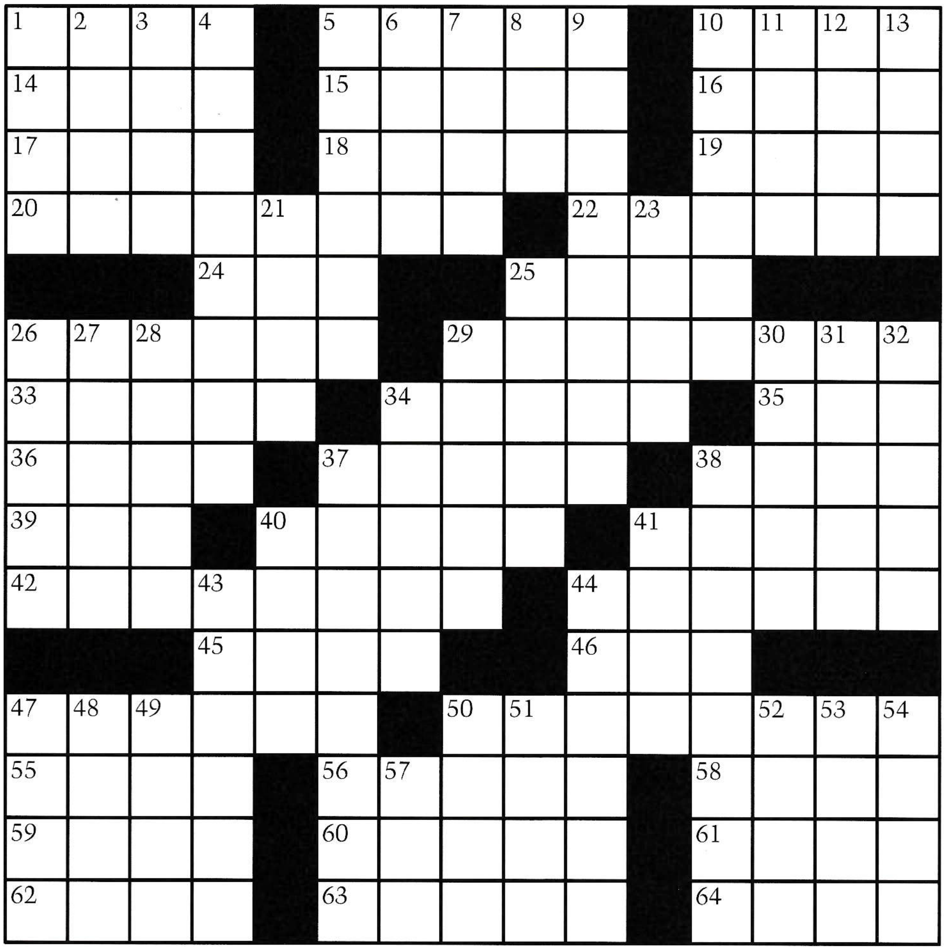
Visit QUANTUM on the Web!

We've opened up a site on the World Wide Web, so stop by if you can. As with most Web sites, we're still growing. But already you'll find an **index** of *Quantum* articles, a **directory** of personnel and services, **background** information on *Quantum* and its sister magazine *Kvant*, and more. Recent visitors will have found a **preview** of this issue, plus a Brainteaser **contest**.

Point your Web browser to
http://www.nsta.org/quantum

Criss × cross science

by David R. Martin



Across

1 Ireland
5 Noble gas
10 Feudal estate boss
14 Contract
15 ___ oil (caraway oil)
16 French historian ___ Halevy
17 Horses' walk
18 Australian statesman Herbert ___
19 365 days
20 Containing H_2O
22 Element 88
24 Hebrew judge
25 Fir or hemlock
26 Taxi driver
29 Abscissa's partner
33 Smell in Brit.
34 Diptera
35 Fast particle's energy
36 Group of rays
37 Newsmen
38 Arrive
39 "no ___, ands, or buts..."
40 Intestine section
41 Eurasian finch
42 Element 89
44 Chemist James ___ (1893–1978)
45 Walked upon
46 Cap
47 3.258 light-years
50 Synthetic rubber
55 Pain
56 Gram-negative aerobic bacteria
58 Reverse curve
59 Celt
60 Belgium province
61 ___ Letterman
62 60,079 (in base 16)
63 Organic compounds
64 60,906 (in base 16)

Down

1 Alphabet run
2 Now ___ me down ...
3 Geophysicist Harry ___ (1859–1944)
4 Max. or min.
5 It's used in explosives
6 Talk wildly
7 Common differential operator
8 In and ___
9 Mg_3N_2 and Mn_4N
10 ___ jar (capacitor)
11 Oil: comb. form
12 Indonesian islands
13 Skin: suff.
21 Winglike
23 French liqueur
25 Light refracter
26 Warm sea fish
27 712,444 (in base 16)
28 Bragg
29 Disulphuric(vi) acid
30 Sweater type
31 Virologist Howard ___
32 Occurrence
34 Psychoanalysis founder
37 Geologic epoch
38 Center's path
40 Concerning
41 Cleansing agent
43 In and of ___
44 Singing groups
47 Book leaf
48 44,202 (in base 16)
49 Actress ___ Perlman
50 Usage: comb. form
51 Jewish month
52 Mild oath
53 Compacted snow
54 61,162 (in base 16)
57 Container

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE JANUARY/FEBRUARY PUZZLE

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

Milk routes

Count the whey

by Dr. Mu

WELCOME BACK TO COWCULATIONS, THE column devoted to problems best solved with a computer algorithm. The winter snow has finally hit the Midwest. The muck is knee deep in the barnyard and covered with a foot of clean white powder, which sure brightens up the place on a clear sunny day. Town folk come out just to gaze upon our countrified setting. There's the barn and silo blanketed in fresh snow, with smoke gently rising from the farm house chimney. And there's farmer Paul, hitching up the old milk wagon, getting ready to make the daily milk run. The wagon works better than the truck in these conditions, when you're bound to run into slick spots and deep snow. I'm doing the pulling 'cuz I need the exercise badly. On these wintry days, we get lazy all bedded down in the barn on our comfy farm mats. Without a daily walk, I get cranky.

All of our customers live between our farm, in the southwestern part of Cream County, Wisconsin, and the town of Paris, in the northeastern section. Not only are

Art by Mark Brenneman

they between the farm and town, they are on the route to town. To be precise, once we leave the farm, the route we follow passes by every customer while always moving in a direction that leads closer to town. We never have to go out of our way as we travel the milk route into town. In this weather, I'm not happy with any other milk route.

We're accustomed to our regular route, which runs on automatic. Once farmer Paul makes a delivery and returns to the wagon, I'm off to the next stop, down two blocks, up a couple more, then stopping right at the next milk stop. But today's weather conditions have put a crimp in my routine by blocking some roads. I need to do some cowculations to see if I can still find a route to town. Of course, it must go by every customer while always moving closer to town.

The roads in the county are laid out in an $M \times M$ rectangular array. Our farm is on {1, 1} and the town is on {M, M}. The k customers are located on the roads {{x1,y1}, {x2, y2}, ..., {xk, yk}} and can be arranged so they are all reached on a route that moves ever closer from farm to town. Snowdrifts are located on roads {{sx1, sy1}, {sx2,sy2}, ..., {sxj,syj}} all within the $M \times M$ array of roads.

Cream County map

I'll draw a map of the Cream County road system to show you where the customers and snowdrifts are located using *Mathematica.* The customers and snowdrifts are located on the following roads given by {x, y} coordinates.

```
customers = {{2, 3}, {5, 6}, {9, 9}};
snowdrifts = {{3, 5}, {4, 7}, {1, 9}, {7,
4}, {7, 7}, {6, 1}};
```

There are 100 roads in the county laid out in a 10 × 10 array. I place a 0 in the array positions to represent the farm house and the town. All customers' roads are given a value of 1, and snowdrifts are assigned the value of 3. All remaining roads, which are clear, are identified with 2. This is carried out in the following sequence of *Mathematica* expressions.

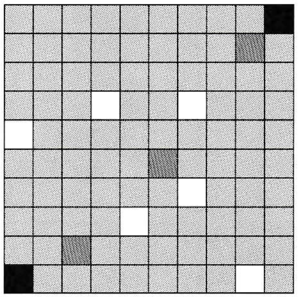
```
n = 10;
road[1, 1] = 0;
road[n, n] = 0;
road[x_, y_] := 1 /;
MemberQ[customers, {x, y}]

(* This expression should be read;
assign road[x, y] to 1 provided {x, y} is
an element of the customer list*)

road[x_, y_] := 3 /; MemberQ[snowdrifts,
{x, y}]
road[x_, y_] = 2;
CreamCounty = Array[road, {n, n}];
```

I can view Cream County with ListDensityPlot.

```
CCMap = ListDensityPlot[CreamCounty,
Frame -> False]
```



Road key: black = farm and town, white = snowdrifts, dark gray = customers, light gray = all the rest.

Milk route

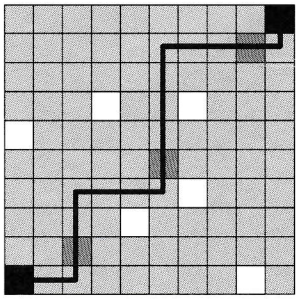
A milk route is a list of roads from the farm {1,1} to the town {10,10}. Consider the following milk route:

```
milkRoute = {{1, 1}, {3, 1}, {3, 4}, {6, 4},
{6, 9}, {10, 9}, {10, 10}};
```

Subtract .5 from all road positions,

```
milkRoute - .5
```

```
{{0.5, 0.5}, {2.5, 0.5}, {2.5, 3.5}, {5.5,
3.5}, {5.5, 8.5}, {9.5, 8.5}, {9.5, 9.5}}
```



and draw a line down the center of the roads on the milk route, with just a little thickness added.

```
Show[CCMap, Graphics[{Thickness[.02],
Line[milkRoute - .5]}]]
```

This suggests a problem, which, you guessed it, is your Challenge Outta Wisconsin.

COW 9

Write a program that cowculates how many routes are available from the farm to the town that always moves closer to town, delivers the milk to each customer, and never goes through a snowdrift. Two routes are different if they differ on at least one road. Of course, your program should be able to handle any county road system and any set of customers and snowdrifts.

Snowdrifts block the usual way,
But families need their milk today.
To make it through,
I'm asking you,
To lick this COW, and count the whey.

—Dr. Mu

Solution to COW 7

COW 7 was posed two issues back. Given a sequence $\{x_1, x_2, \ldots, x_n\}$ of integers of length n, write an efficient algorithm (of order n) to find a subsequence $\{x_L, x_{L+1}, \ldots, x_H\}$ of consecutive terms with the largest sum. Return the beginning and ending indices L, H, and the maximum sum $\sum_{i=L}^{H} x_i$.

A cremeDeLaCreme solution was sent in by Richard Rice (rrice@aw.sgi.com). His solution is based on a surprisingly simple algorithm of order n. Start summing consecutive terms from first positive term (say at j), and, as long as the subsum remains positive, continue adding consecutive terms. Keep a MaxSoFar of the largest subsum. Once the subsum goes negative (say at term k), reset the starting left position to the first positive term after k. Continue in the same manner, comparing subsums with the MaxSoFar. It is easy to prove that it is not necessary to consider any starting term between $j+1$ and k. (A little chore for you.) Here is Rice's solution in *Mathematica:*

```
cremeDeLaCreme[x_] := Module[
  {SubSum = 0, MaxSoFar = 0, n = Length[x],
    L = 0, low = 0, high = 0, H, term},
  Do[term = x[[H]];
    If[SubSum <= 0, SubSum = term; L = H,
      SubSum += term];
    If[SubSum > MaxSoFar,
      MaxSoFar = SubSum; low = L; high =
        H], {H, 1, n}];
  {MaxSoFar, {low, high}}]
```

Let's test it out with a list of winnings consisting of random numbers between –50 and 50 from length 100 to 100,000 in jumps of powers of 10.

```
Table[winnings = Table[Random[Integer, {(-
  50), 50}], {10^i}];
  {10^i, Timing[cremeDeLaCreme[winnings]]},
  {i, 2, 6}] // MatrixForm
```

```
    100          {0.05 Second, {290, {54, 91}}}
   1000         {0.11 Second, {1117, {635, 905}}}
  10000        {0.99 Second, {3049, {5174, 6498}}}
 100000       {9.56 Second, {6335, {40966,59940}}}
1000000 {95.84 Second, {15795, {313633, 367146}}}
```

Not bad. It took about 1/10 of a second to solve the problem for $n = 1{,}000$. Recall that the cubic algorithm was estimated to take 4.5 hours to get this solution in *Mathematica*.

Also notice that as n increases by a factor of 10, the time increases by a factor of 10, showing that the algorithm is linear. For $n = 1{,}000{,}000$, the linear model takes about a minute and a half, while the cubic (worst case) algorithm would require over 50 years on a Cray supercomputer. Good algorithms, like good feed, can make a big difference.

The problem was also solved by Shawn Kuo (skuo@WPI.EDU), a high school junior at the Massachusetts Academy of Math and Science at Worcester.

And finally...

Send in your solution to me at drmu@cs.uwp.edu. Past solutions are available at http://usaco.uwp.edu/cowculations.

If competitive computer programming is your passion, stop by the USA Computing Olympiad web site at http://usaco.uwp.edu. The 1997 USA team recently returned from the Ninth International Computing Olympiad in Cape Town, South Africa, with a bronze, a silver, and a gold medal. This is an all-expenses-paid trip for the top four high school computer programmers. Check it out and maybe you'll want to sign up for the 1998 USA National Championship.

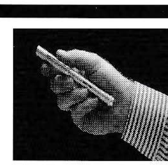
QUANTUM

makes a perfect gift!

Use the response card in this issue to order *Quantum* for your child, grandchild, niece, nephew, mother, father, friend . . . Or call **1 800 SPRINGER (777-4643)**. Give them six colorful, challenging, entertaining issues of *Quantum*—a year's worth of reading pleasure!

METROLOGIC® EDUCATIONAL LASERS

Helium-Neon Lasers
ranging from .05 to 5.0mW
NOW AVAILABLE IN
GREEN (.05mW only)



Laser Pointer 3mW
Only $49.95
Optics Kit $20.00

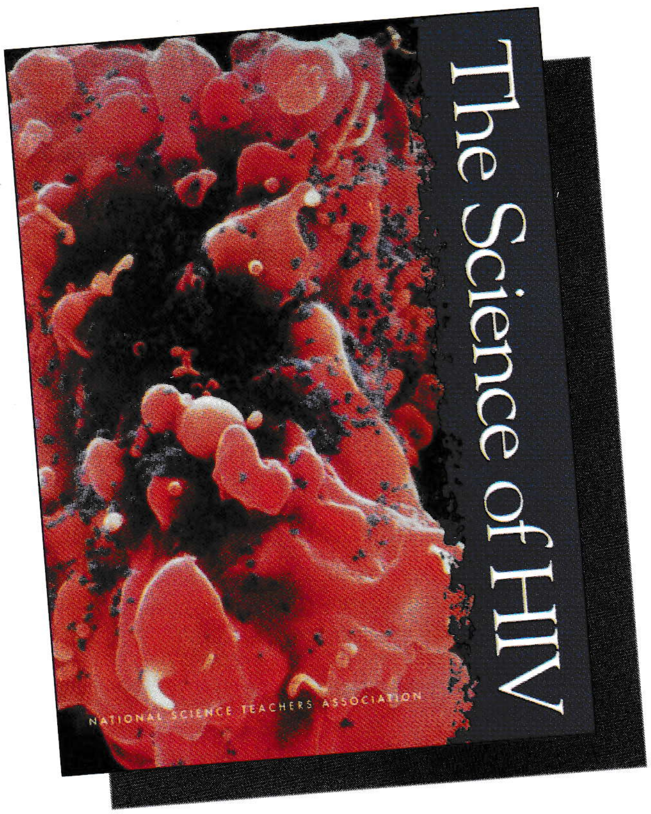
Various Kits Also Available
* speed of light
* build a laser kit
* laser optics lab
* physical optics lab
* power meters

1-800-667-8400
In New Jersey 609-228-8100
90 COLES ROAD
BLACKWOOD, NJ 08012
EDUCATIONAL DISCOUNTS AVAILABLE

Circle No. 1 on Reader Service Card

A Community Resource
To Understand and Prevent AIDS

The Science of HIV Curriculum Package



Developed by the National Science Teachers Association with funding from Abbott Laboratories. Written by Michael DiSpezio. Video by Summer Productions.

NSTA's new science-based resource guide is different from most "AIDS books"—its activities and readings focus on biological concepts relating to HIV. Activities cover the following subjects:

- **selected topics in cell biology**
- **basic virology**
- **HIV structure, replication, and genetics**
- **immune system function and HIV infection**
- **drug therapeutics**
- **prevention strategies**
- **a global perspective on the AIDS pandemic**

This curriculum package can be used as a community educational resource or to expand upon a high school biology or health curriculum. Reproducible student pages make lesson plans flexible; educator pages provide background and presentation strategies. Material appropriate for anyone at the high school level and above.

The text is coordinated with an original video made for this project. Animations of complex concepts are interwoven with scientist interviews and compelling stories of adolescents who are living with HIV. The video has won numerous awards, including:

- **Best Achievement for Children's Programming**
  1997 International Monitor Awards
- **Silver for Children's Programming**
  1997 Houston International Film Festival
- **Gold Circle Award**
  American Society of Association Executives

Grades 9–College, 1997, 184 pp, 30-minute video

#PB136X **$45.00**

To Order Call 800-722-NSTA

Visit Our Web Site www.nsta.org