QUANTUM

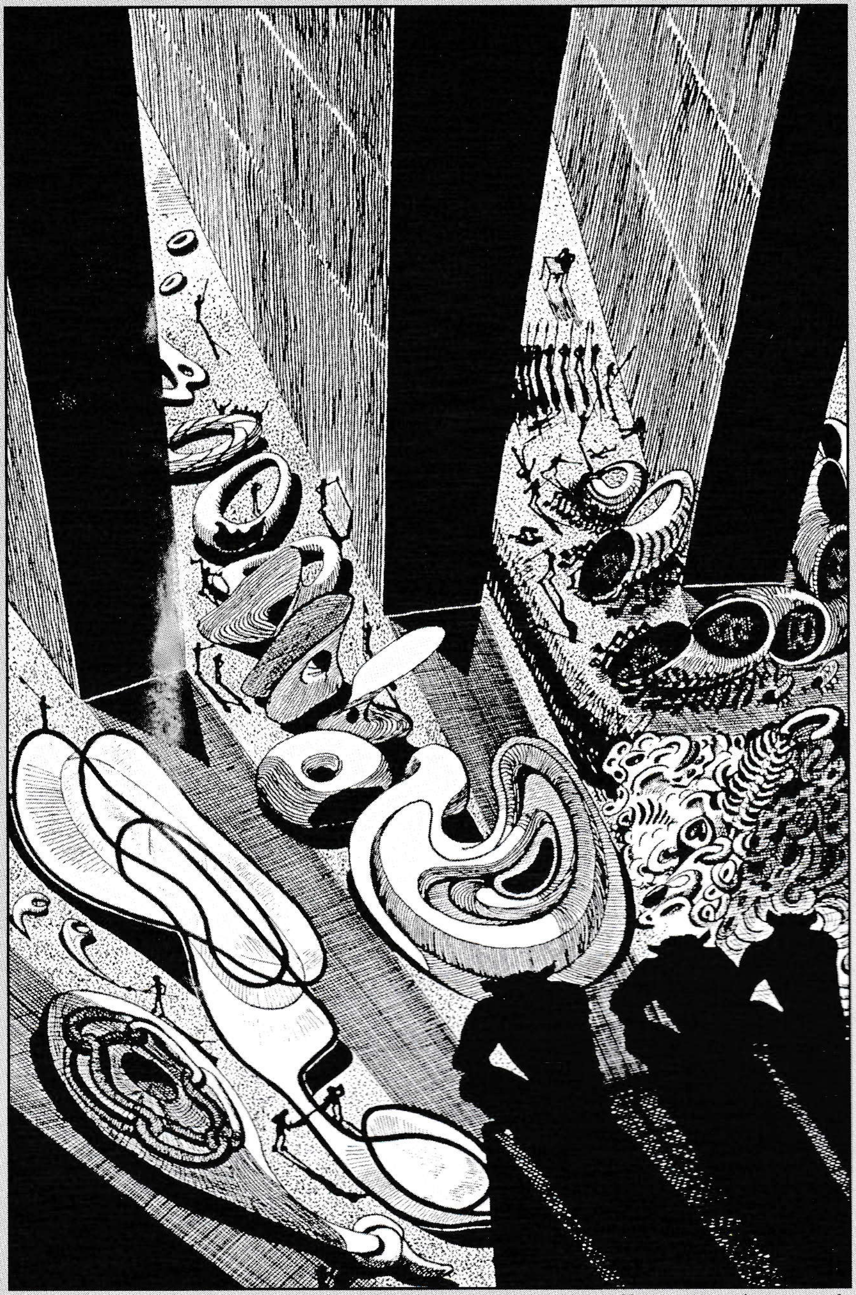
MAY/JUNE 2000

US $7.50/CAN $10.50

Springer

NSTA

GALLERY Q



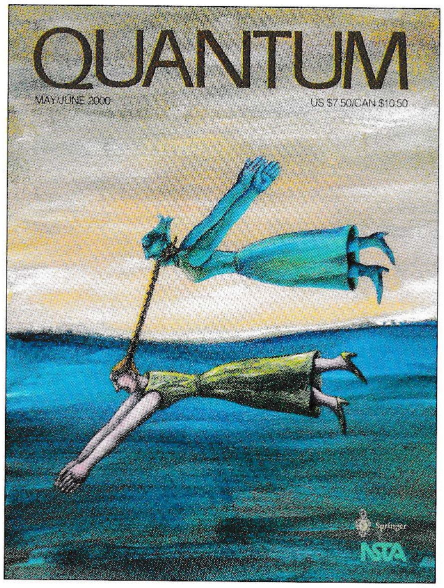
India ink on paper, 21 × 32 cm, reprinted by permission of A. T. Fomenko

Topological Zoo (1967) by Anatoly Fomenko

AMONG THE MANY DIFFERENT PERTURBATIONS of physical space depicted above, you will find a soapy film stretched across a circular wire. Upon closer observation, you will discover that the shape of the wire frame combines both a Mobius strip and triple Mobius strip. The soapy film, however, has no trouble working within this wiring configuration to establish the minimal surface needed to cover the frame. To learn more about the fascinating ability of soap films to reveal elegant iridescent solutions to complex mathematical challenges, turn to page 4. Afterword, you'll never look at a soap bubble the same way again.

QUANTUM

MAY/JUNE 2000
VOLUME 10, NUMBER 5



Cover art by Jose Garcia

The displaced person hovering over our submerged cover girl hints at a solution to problems of buoyancy that can often make you feel that you are in way over your head. With a little insight, however, you'll soon be swimming with the sharks without fear and tackling even the weightiest issues. Dive in by turning to page 34.

FEATURES

DEPARTMENTS

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.

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 10 (6 issues) will be published in 1999–2000. *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.

Subscription Information:
North America: Student rate: $18; Personal rate (nonstudent): $25. This rate is available to individual subscribers for personal use only from Springer-Verlag New York, Inc., when paid by personal check or charge. Subscriptions are entered with prepayment only. Institutional rate: $49. 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:
Representatives: (Washington) Paul Kuntzler (703) 243-7100; (New York) Jay Feinman (212) 460-1682; and G. Probst, Springer-Verlag GmbH & Co. KG, D-14191 Berlin, Germany, telephone 49 (0) 30-827 87-0, telex 185 411.

Printed on acid-free paper.

Visit us on the internet at www.nsta.org/quantum/.

QUANTUM

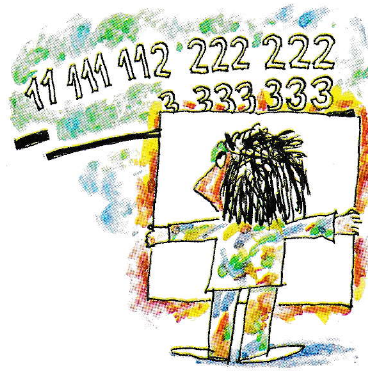
THE MAGAZINE OF MATH AND SCIENCE

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.

BRAINTEASERS

Just for the fun of it!



B291

All square? The number 11 111 112 222 222 – 3 333 333 is a perfect square. Find its square root.

B292

The right approach. Line segment MN is the projection of a circle inscribed in a right triangle ABC onto its hypotenuse AB. Prove that angle MCN is 45°.

B293

The hose knows. A man is filling two tanks with water using two hoses. The first hose delivers water at the rate of 2.9 liters per minute, the second at a rate of 8.7 liters per minute. When the smaller tank is half full, he switches hoses. He keeps filling the tanks, and they both fill up completely at the same moment. What is the volume of the larger tank if the volume of the smaller tank is 12.6 liters?

B294

The culprit in the caves. Sixteen caves are arranged in a row. The sheriff of a nearby town, whom the residents respectfully call Big Brow, knows that a robber, called Elusive Joe, is hiding in one of those caves. The sheriff also knows that, on advice from his friends, Elusive Joe moves from one cave to the next every night. The sheriff with his deputies can search only one cave a day. Can the sheriff catch the criminal before the end of May if he starts searching the caves on the first of May?

B295

A watched pot. A small pot with water is placed into a large pot, which is also filled with water. The large pot is placed on a gas-stove burner, and the water in it is brought to a boil. Will the water boil also in the small pot?

ANSWERS, HINTS & SOLUTIONS ON PAGE 52

Art by Pavel Chernusky

Minimal surfaces

Embarking on a plane discussion

by A. Fomenko

When the 19th century Belgian physicist Plateau began experimenting on the shape of soap films, he could hardly imagine that this work would initiate a new direction in scientific research, whose vigorous development would continue to the present, and which is now known as Plateau's problem. Most of us have been familiar with Plateau's experiments since childhood: all children like blowing soap bubbles or making soap films with a wire ring.

Here is some advice for those who want to obtain beautiful soap films. You need some thin, flexible wire, some dishwashing detergent, a glass or bowl of warm water, and some glycerin (one can do without it, but if the mixture contains glycerin, the films will be more stable. Mix the soap into the water, then add the glycerin. Make a ring with a handle from a piece of wire, dip the ring into the solution, then carefully take it out. Changing the shape of the ring will change the shape of the film.

When we remove the wire ring from the soapy water, a striking iridescent soap film forms inside the ring. The size of the film can be rather large. However, the larger it is, the more quickly it will burst under the force of gravity. On the other hand, if the ring is small, gravity may often be ignored in the study of soap films. We shall use this fact below.

How is a soap film formed?

Let us consider how the properties of the surface layer of the liquid change as soap is added to it. Figure 1a shows schematically the boundary between two media—water and air. The arrows indicate the attractive forces between the molecules of water, which are *polar*. Polar molecules are those that contain an asymmetric distribution of electric charges. The forces shown are responsible for the surface tension at the boundary between the two media. Unlike the water molecules, the molecules of soap consist of long, thin, non-polar hydrocarbon chains, with a polar oxygen radical at one end of the chain. When soap molecules are added to water, they tend to come to the surface and cover it

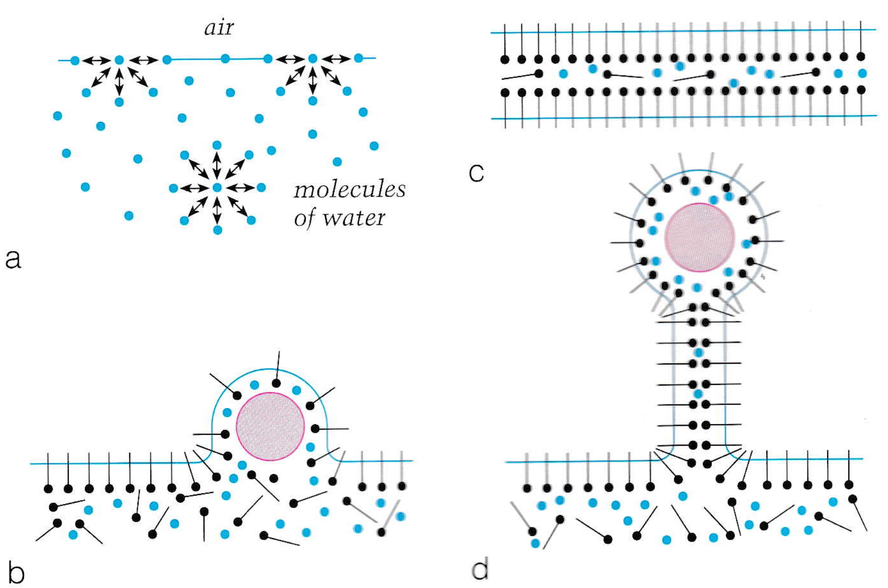


Figure 1

with a uniform layer, with the nonpolar end of the molecule oriented outward (figure 1b). Pushing the water molecules down, the soap molecules reduce the capillary tension. This circumstance renders the surface film more elastic, and it is this elasticity that allows the soap film to form when we dip and remove the wire ring.

A cross section of the wire is show in figure 1. When the wire reaches the surface, the surface swells and covers the wire. This occurs because the number of molecules near the wire decreases temporarily (figure 1b). Therefore the surface tension, which depends on the number of water molecules in the surface layer, increases. This makes the region of the surface near the wire more flexible, which in turn gives rise to the formation of a soap film when the ring is removed from the water.

It is also clear that the thickness of the soap film enclosed by the ring cannot be less than the sum of the lengths of two soap molecules (figure 1c). The formation of a soap film is illustrated in figure 1d. As the wire ring is lifted out of the water, the flexible film covers the ring and is dragged along by it. The force of gravity restricts the size of the surface, so that when the ring is far enough from the liquid, the film breaks.

Minimal surfaces as a model for soap films

The physical principle underlying the formation of soap films is very simple. A physical system will not change its shape unless it can change easily to another shape with less energy. The energy of the surface of a soap film is often described in terms of the surface tension of the liquid. It depends on the attractive forces between the molecules and on the fact that these forces are unbalanced at the boundary of the surface. The presence of unbalanced forces gives rise to the following interesting effect: the liquid film turns into an elastic surface that tends to minimize its area and thus minimizes the surface energy per unit area. Here we disregard the force of gravity and the air pressure.

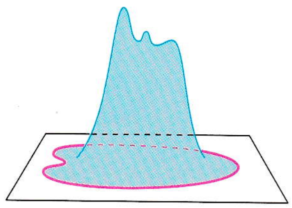


Figure 2

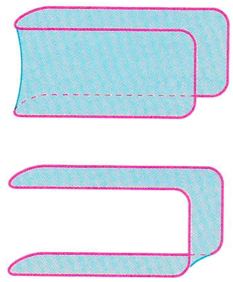


Figure 3

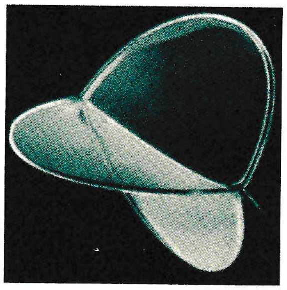


Figure 4

For these reasons, we can model a soap film with a smooth surface of minimal area that spans the contour formed by the wire. In mathematics, this is called a *minimal surface*. The classical theory of such surfaces is part of the *calculus of variation*. The branch of mathematical analysis dates back to the eighteenth century. Today, this theory also uses the more modern methods of topology and differential geometry. While the theory cannot be explained in detail using only the tools of elementary mathematics, many of its results can be illustrated by example and verified experimentally. In this process, some elegant geometric problems arise that can be solved using simple mathematics.

Simple contours

Let us first consider the case in which the contour of the wire is warped only moderately, so that no point is "on top of" another point. More specifically, we mean that we can find a plane such that the projection of the contour onto this plane is convex, and any two different points of the contour project onto two different points of the plane. In these cases, an advanced theorem guarantees that there exists a unique minimal surface spanned by this contour. If the contour is planar, the existence and uniqueness of a minimal surface seems clear: the contour bounds a region of the plane, and it is clear (see figure 2) that any other surface spanning it has a larger area than the plane surface.

Different films spanned by the same contour

If we allow the wire to assume contours more complicated than those described above, the uniqueness theorem will not hold. Figure 3 shows a simple example of this, which can be verified experimentally. Notice that the contour in figure 3 cannot be projected "nicely" onto any plane: in any projection (onto any plane), there are pairs of points, or even entire segments, which project onto a single point.

Problem 1. Design other contours that can be spanned by different minimal surfaces. Is it possible that more than two such surfaces exist?

If the contour formed by the wire meanders wildly (for example, if it ties itself in knots), the uniqueness of the minimal surface breaks down, and the structure of a minimal surface can become very complicated. For example, it can develop *singularities*, points near where the surface cannot be obtained by warping a simple disk, but contains branches. Using soap films, such a surface can be modeled by starting with a wire contour that itself contains branches.

Branching contours

A simple contour of this type is shown in figure 4. The minimal surface spanned by this contour has a whole segment consisting of singular points. Branching from this segment, the surface forms three plane sheets with an angle of 120° between them. The factors responsible for this effect can be clarified by solving the following problem.

Problem 2. (i) Connect three

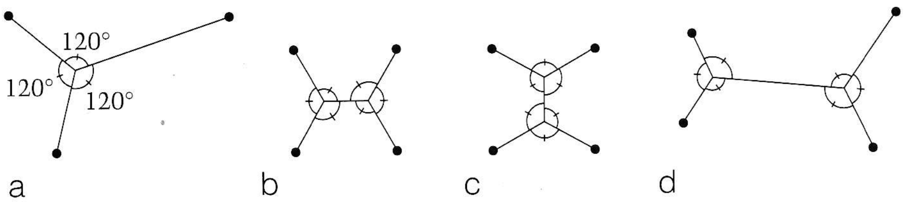


Figure 5

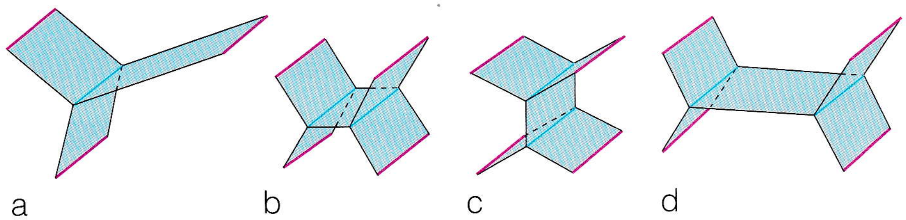


Figure 6

points by segments such that the sum of their lengths is minimal. Solve the same problem for (ii) the vertices of a square, and (iii) the vertices of an arbitrary quadrilateral.

The answers to this problem are given in figure 5, and the parts of the corresponding minimal surfaces are shown in figure 6. Notice that all the angles here are equal to 120° and that there are two possible answers to problem (ii) (see figures 5a and 5b).

It is interesting that there is no minimal film in which more than three planar regions intersect along a single segment. This fact can be explained using simple geometric considerations. Suppose four planar regions that are part of a minimal surface intersect along some segment, and project the surface onto a plane perpendicular to this segment (figure 7). We now want to be sure that the sum of the segments that are the projections of the planar regions is minimal. But in fact we can replace the point where all four segments intersect with two points in which three segments intersect (figure 7), and the result of problem 2 shows that the sum of the lengths will be smaller.

The three-dimensional situation corresponding to this argument is shown in figure 8. Let us look a bit deeper into this situation. It is not hard to see that splitting the intersection of four planar regions into two intersections of three planar regions will reduce the total area. The minimal surface is in stable equilibrium: small perturbations of the film can only increase its area. This means that the sum of the forces acting on every point of the film is zero. At the singular points, where three sheets converge, this observation means that the sum of the three forces acting in the directions of these sheets must be zero. This implies that the angle between the convergent sheets is 120°.

Problem 3. Draw a polygonal line of minimal length connecting five vertices of a regular pentagon. How many such lines exist?

Other beautiful branching films can be obtained if the film is spanned by a frame of a tetrahedron or cube (figure 9). Notice the singular points at which several singular edges converge.

Problem 4. (i) Find all the angles between the planes of the minimal surface spanned by the edges of a regular tetrahedron (figures 9a and 9b). Find the length x of the side of the central square for the minimal surface that is spanned by the edges of a cube (figure 9b).

Problem 5*. Is it possible that three singular curves converge at a singular point of a minimal surface? What can be said about five singular curves?

Open contours

The interaction of a surface with its boundary contour is a very important characteristic of the surface. Various special cases of this interaction can be studied by topological methods. It follows from the physical properties of a minimal surface that it cannot contain holes. If it did, surface tension would force the hole to expand until the entire film, or part of it, collapsed at the boundary line. It is easy to test this property experimentally: just puncture a soap film quickly with a thick piece of wire or a rod. In advanced work, a mathematical definition of a soap film's boundary is based on this property.

This is, of course, a mathematical model, which simplifies reality. In

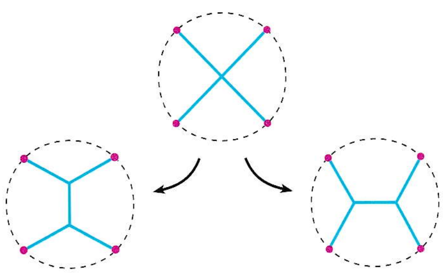


Figure 7

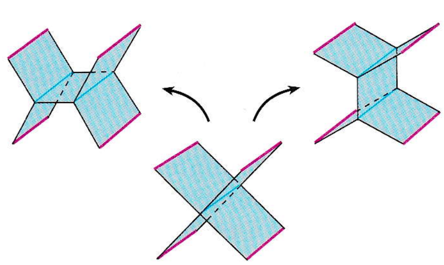


Figure 8

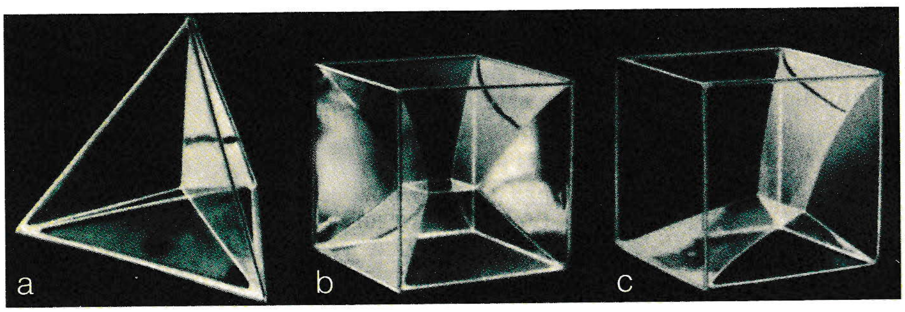


Figure 9

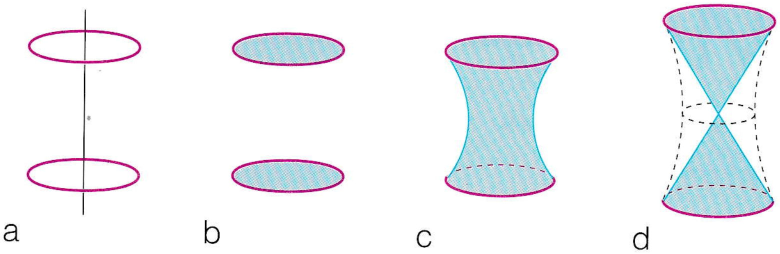


Figure 10

fact there exist stable minimal surfaces that contain holes, in the sense that they "hang" on a wire loop but leave some parts of it free (see figure 11a). This phenomenon is explained by the fact that a real wire has a finite thickness, which can, in certain cases, stabilize a mathematically unstable construction. For this to happen, the wire must be sufficiently thick compared to the size of the minimal surface.

Another example is shown in figure 11b. In this case, the contour is an open curve with two ends that can be straightened into a flat segment. Mathematically, there cannot be a minimal surface that spans a segment embedded in three dimensions and that does not intersect itself. This anomaly can be explained by the fact that the segment does not form a loop.

Several contours

Let us now consider the properties of minimal surfaces whose boundaries consist of, say, two circles. Let's take the two circles in two parallel planes, both perpendicular to a vertical axis and centered on this axis (figure 10a). If the circles are far apart from each other, the minimal film coincides with two plane disks spanned by these circles (figure 10b). However, if the circles are moved closer to each other, a different minimal surface begins to appear. This surface, called a *catenoid*, spans both the circles (figure 10c).

This surface possesses many interesting properties. For example, it can be obtained by rotating a *catenary* curve about its axis of symmetry. To construct this curve, consider two points in a vertical plane, and suppose that one endpoint of a heavy chain is attached to each point. The chain hangs freely under the force of gravity, and the shape it assumes is called a *catenary*. Rotating this curve about its axis of symmetry, we obtain a catenoid, and arranging it vertically we obtain the minimal surface shown in figure 10d.

Note that the minimal surface is *not* a cone with its vertex at the origin. Figure 10d shows how such a cone would change its shape to assume a position with minimal area (provided that the boundary circles are fixed). The cone ultimately becomes a catenoid. This phenomenon is similar to the situation in which a fourfold singular point splits into two threefold singular points. Indeed, we can think of figure 7 as showing the transformation of a "one-dimensional cone" consisting of two intersecting segments into a more complex curve with two threefold singularities. In principle, a similar phenomenon occurs in the case of a two-dimensional cone. In this case, however, the vertex is transformed into a circle, the narrowest cross-section of the catenoid. In both cases, deforming the film (or one-dimensional curve) decreases its area (or length) and makes the surface (or curve) assume a position giving the minimal area (or length).

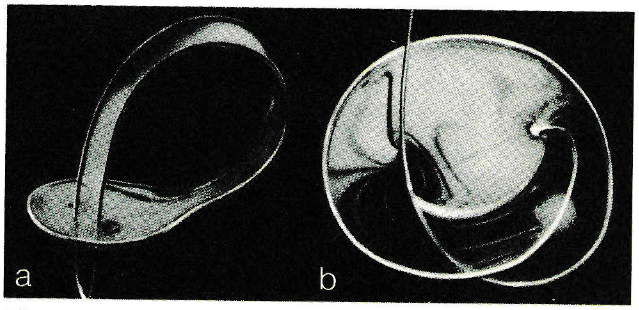


Figure 11

Minimal surfaces in nature

It turns out that minimal surfaces are widespread in nature, since they are the most economical way to form skeletons of living organisms. The most striking example is provided by the skeletons of *radiolaria*. These are microscopic sea organisms with various peculiar shapes. The English scientist D'Arcy Wentworth Thompson in his book *On Growth and Shape* was perhaps to first to notice that capillary action plays an important role in defining the shape of these organisms. Radiolaria consist of small bunches of protoplasm confined in foam-like forms similar to soap bubbles or films. These organisms are rather complex in shape, and the minimal surfaces they exhibit have, in general, many branch points and edges, which occur where the bulk of the organism's liquid mass is concentrated. We can easily see a similar concentration of liquid in soap films: the liquid flows along the film until it meets an edge where three sheets coincide. Here the liquid thickens the film so that the singular edges of the minimal surface are highlighted. A similar process occurs in radiolaria. As the liquid is concentrated along the branching edges, solid fractions of the seawater separate out and settle along the edges, gradually forming a solid skeleton. The geometry of this skeleton can be visualized by looking at the branching edges of a soap foam; that is, the edges shared by soap bubbles in the foam. These edges form a complex network that is the 'liquid skeleton' of the foam. When an organism with such a skeleton dies, the soft tissues gradually vanish, and a solid skeleton, evolved as described above, remains.

Thus the skeleton is just a representation of the system of branching edges and singular points of a

CONTINUED ON PAGE 13

Molecular interactions up close

Distant relations

by G. Myakishev

IF THERE WERE NO ATTRACTIVE forces between molecules, all matter would be in a gaseous state under all conditions. The molecules are held together and can form liquids and solids only because of the attractive forces.

However, the attractive forces alone cannot produce stable atomic and molecular structures. Stability is based on the balance of forces, which is ensured by the existence of repulsive forces that act at extremely small distances between atoms and molecules.

There is no doubt about the existence of intermolecular forces. To determine the value of these forces and their dependence on the distance between molecules is, however, a very difficult task.

Van der Waals forces

Although there are no methods of measuring directly the intermolecular forces at any distance, we now know a great deal about them, but not everything.

A Dutch physicist, Johannes Diderik van der Waals (1837–1922), was the first to introduce this concept and prove the key role of intermolecular forces in the description of real gases. He did not attempt to determine the exact dependence of these forces on distance, but simply proposed that the repulsive forces act at small distances, while at large distances they are replaced by the attractive forces, which slowly decrease with distance.

On the basis of these very simple postulates and his own logic and intuition, van der Waals obtained an equation describing the state of real fluids (gases and liquids). Van der Waals' real gas model describes not only gases and liquids but also melting and evaporation. The intermolecular forces are often called the van der Waals forces.

Electromagnetic nature of the intermolecular forces

Until the beginning of the twentieth century, it was impossible to theoretically analyze the intermolecular forces. The simple and well-known gravitational force clearly could not be a significant factor in the interaction of objects with such small masses as molecules. Therefore, the only way to explain the origin of molecular interactions was to propose that the intermolecular forces were electromagnetic, because other forces were unknown in that time.

Atoms, and especially molecules, are very complex systems consisting of a large number of charged particles. The structure of such systems was unknown. The forces acting between the molecules clearly depended on their structure. Of course, only very simple cases were considered in the beginning.

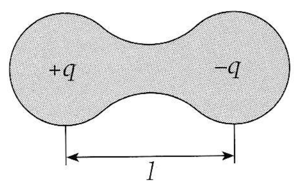


Figure 1

Orientational forces

In many molecules (for instance, water molecules) the positive and negative charges are distributed in such a way that the average position of the centers of positive and negative charges do not coincide. As a first approximation, such a molecule can be modeled as an electric dipole, that is, a system of two point charges, $+q$ and $-q$, separated by a small distance l (figure 1). The electrical properties of such a molecule are characterized by the dipole moment

$$p = ql.$$

Art by Leonid Tishkov


Interaction of
moleculman and gardman
Тишков 99

At first, nobody knew how to calculate the dipole moments of these molecules. Such calculations were impossible before the advent of quantum mechanics.

If, on the other hand, it is assumed that the dipole moments of two molecules, p_1 and p_2, are known, then the dependence of the interaction force on the distance between the molecules can be calculated with the help of Coulomb's law. The attractive force between two dipoles is maximal when they are located along the same line (figure 2). This force appears because

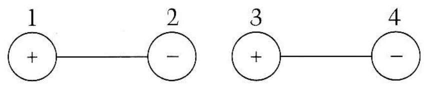


Figure 2

the distance between the unlike charges 2 and 3 is slightly less than the distance between the like charges 1 and 3 or 2 and 4.

Since the force of the dipole-dipole interaction depends on the mutual orientation of the dipoles, it is called the orientational force. The stochastic thermal motion continuously changes the orientation of the molecular dipoles. Therefore, the force of interaction between the dipoles should be calculated as the mean value of all possible orientations. The calculations showed that the attractive force is proportional to the product of the dipole moments p_1 and p_2 of the molecules and inversely proportional to the intermolecular distance raised to the seventh power:

$$F_{\mathrm{or}} \sim \frac{p_1 p_2}{r^7}.$$

Compared with the Coulomb force that acts between two point charges, and is proportional to $1/r^2$, the orientational force decreases very rapidly.

Inductive (polarization) forces

Another type of relatively simple molecular interaction arises between two molecules, only one of which has a dipole moment. The dipole molecule generates an electric field that polarizes another molecule in which the electric charges are initially uniformly distributed. As a result, the positive charges are displaced in the direction of the electric field, while the negative charges are shifted in the opposite direction. The nonpolar molecule is therefore slightly stretched; it acquires a dipole moment and becomes polarized (figure 3).

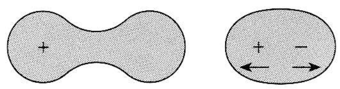


Figure 3

The attractive force in this case can also be calculated. It is proportional to the dipole moment p of the polar molecule with a coefficient α, which characterizes the ability of a nonpolar molecule to be polarized. This force is also inversely proportional to the seventh power of the intermolecular distance:

$$F_{\mathrm{i}} \sim \frac{p\alpha}{r^7}.$$

This attractive force is called the inductive or polarization force, because it is generated by the molecular polarization that is induced by the electrostatic induction.

Dispersion forces

It is widely known that attractive forces arise not only between polar molecules but also between nonpolar molecules. For example, although the atoms of inert gases have no dipole moment, they nevertheless interact with each other. The origin of these forces was determined only after the creation of quantum mechanics.

Qualitatively and crudely, the generation of such forces can be explained as follows. In atoms and molecules the electrons perform an intricate motion around the positively charged nuclei. Although the average atomic (molecular) dipole moment of the nonpolar structures is zero, the electrons may assume an asymmetric "position" at any particular moment of time and impart the molecule with a nonzero instantaneous dipole moment. Such a short-term dipole generates an electric field that polarizes the neighboring nonpolar atoms (molecules). Therefore, the former nonpolar molecules turn into stochastically induced instantaneous dipoles that interact with each other. The total force of interaction between nonpolar molecules results from the average interaction of all possible instantaneous dipoles with the dipole moments generated by the mutual induction of the adjacent molecules.

Quantum mechanics says that in this case the attractive force is proportional to the polarizability of the two molecules α_1 and α_2 and inversely proportional to the seventh power of the intermolecular distance:

$$F_{\mathrm{d}} \sim \frac{\alpha_1 \alpha_2}{r^7}.$$

These forces are called "dispersion" forces, because optical dispersion, that is, the dependence of the refractive index on the frequency (color) of the light, is determined by the interaction between the nonpolar molecules.

The dispersion forces act between all atoms and molecules because the nature of this force does not depend on whether a molecule has a dipole moment. Usually, the dispersion forces are larger than the orientational or inductive forces. However, if polarized molecules (such as water molecules) take part in the molecular interaction, the orientational force can be larger than the dispersion force (by a factor of 3 for water molecules). When the interacting polar molecules have large dipole moments (CO, HCl, etc.), the orientational forces are dozens or even hundreds of times larger than all other forces.

The principal feature of all three types of force is the $1/r^7$ law, which describes their attenuation with distance. However, at distances greater than the molecular size another effect comes into play: the limited velocity of propagation of the electromagnetic interaction. Therefore,

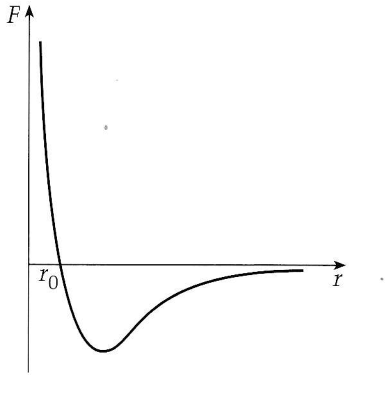


Figure 4

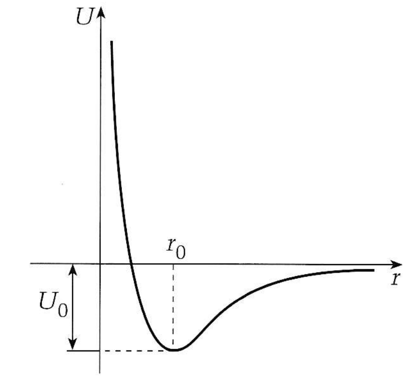


Figure 5

at distances of about 10^{-5} cm the attractive forces decrease more rapidly, proportional to $1/r^8$.

Repulsive forces

Let's consider the repulsive forces that exist between molecules at very small distances. This problem is less complicated, on the one hand, and more complicated, on the other. Since the repulsive forces grow extremely rapidly when the molecules approach each other, it is not necessary to know the precise law governing this rise in the analysis of many molecular events.

However, in contrast with the attractive forces, the repulsive forces are much more "individual," that is, they depend strongly on the structure of the interacting molecules. Even if we know how molecule *A* repels molecules *B* and *C*, we cannot say how molecules *B* and *C* repel each other. When molecules come in direct contact with each other, their individual character is felt more strongly than at large distances.

A rather good agreement between experimental data and theoretical calculations is obtained by assuming that the repulsive forces change as

$$F_{rep} \sim \frac{1}{r^{13}}.$$

Since the attractive and repulsive forces are proportional to $1/r^7$ and $1/r^{13}$, respectively, we may plot an approximate dependence of the total intermolecular force on the distance between molecules or atoms. In figure 4 the repulsive forces are assumed to be positive and the attractive forces are negative. The total force is zero at a distance r_0, which is approximately the sum of the radii of the molecules.

In studying a large number of atoms and molecules, it is more convenient to use the potential energy instead of the interaction force. The aim is to obtain the average characteristics of the molecular system. The mean potential energy determines, as we shall see very soon, many features of the structure and properties of matter.

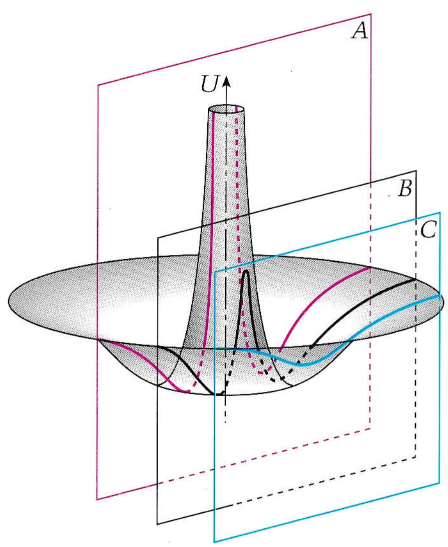


Figure 6

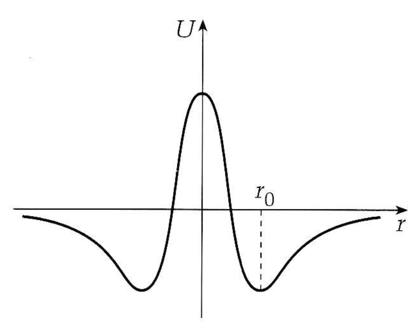


Figure 7

Since the change in potential energy is equal to the work performed by the force, we can find the dependence of the potential energy on distance from the relationship between the force and the distance. There is a rule that if a force varies with distance as $1/r^n$, the corresponding potential energy will vary as $U \sim 1/r^{n-1}$. This rule agrees with dimensional analysis (energy is force times distance).

The potential energy, as we know, can be determined with respect to an arbitrary level. Usually it is assumed that $U \to 0$ as $r \to \infty$, which yields the potential energy plot shown in figure 5, where U_0 is the depth of the potential well.

However, the potential curve (that is, the dependence of the potential energy on distance; figure 5) will have this shape only when the molecules approach each other along the line connecting their centers (for example, if the molecules approach each other in the plane *A* in figure 6). In other cases the potential curve looks like that shown in figure 7 (the molecules move in the plane *B*) or in figure 8 (they move in the plane *C*).

The basic problem

Many properties of a substance can be explained if the character of the interaction of its molecules is known. We will discuss here only one very general problem: how the dependence of the potential energy on intermolecular distance can be used to quantitatively determine the difference between a gas, a liquid, and a solid on the basis of kinetic

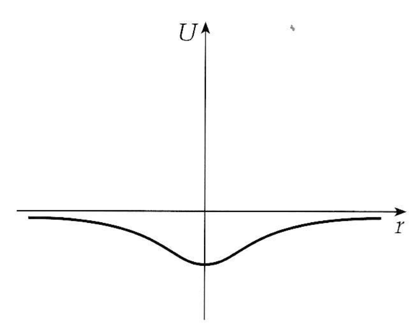


Figure 8

theory. As a first step, however, we will consider the energy of molecular motion.

If we know the dependence of the potential energy on distance, we can determine the character of the motion using only energy conservation. Let's assume that one molecule is at rest and the other is in motion. The motion of the molecule depends on its total energy. According to energy conservation, the total energy of the molecule is constant:

$$E = E_k + U = \text{const},$$

where E_k is the kinetic energy and U is the potential energy.

First, consider the case where $E = E_1 > 0$ (figure 9). The total energy can be plotted as a straight line parallel to the r axis, because it has the same value at any value of r. When the molecule moves along the r axis, its kinetic and potential energies vary: the higher the potential energy, the smaller the kinetic energy, and vice versa (don't forget the negative sign of the potential energy!). If the molecule moves from the right to the left, its kinetic energy grows to a maximum at the point $r = r_0$, where the potential energy assumes the smallest value. The kinetic energy will then decrease and vanish at $r = r_1$, where the total energy equals the potential energy. The molecule cannot enter the region where $r < r_1$. If it could, then the potential energy of the molecule would be greater than the total energy, which means that the kinetic energy would be negative. The kinetic energy, however, is always positive.

At the point $r = r_1$ the molecule stops and starts to move in the opposite direction as a result of the repulsive force. This is the so-called turning point of the molecular trajectory. The molecule subsequently moves in the positive r-direction and goes to infinity.

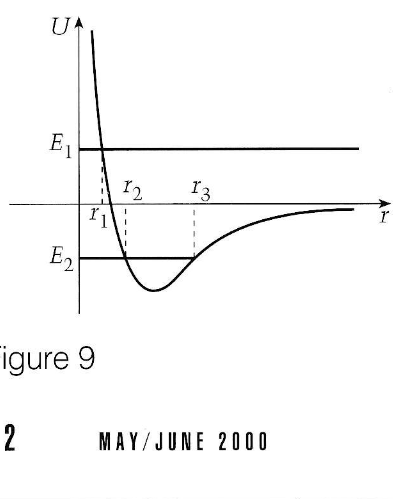


Figure 9

Quite a different scenario occurs when $E = E_2 < 0$ (figure 9). In this case the molecule is situated in the potential well and cannot escape from it. This is the bound state of molecules, in which they oscillate near the equilibrium position. Separation of the system into two independent particles is impossible without increasing its total energy to $E > 0$.

Energy of molecular interaction in solids, liquids, and gases

Let's now determine the quantitative criterion for distinguishing gases, liquids, and solids on the basis of kinetic theory.

Gases. More information on the state of a real gas can be obtained by plotting the potential energy of a molecule as a function of the distance to the nearest neighbors (figure 10). The potential energy of this molecule is zero along most of its trajectory, because the mean intermolecular distance in gases is much greater than the molecular size. The molecule's nearest neighbors are located at points 1 and 2. The molecule travels at an appreciable distance from neighbor 1 and closer to neighbor 2.

The mean potential energy of the molecule is negative and very small. Its value is equal to the area delineated by the potential curve between points 1 and 2 and the r axis divided by the length of the interval 1-2. The total mean energy is necessarily positive; otherwise, the molecule would be bound to its neighbors. This is possible only if the mean kinetic energy of a molecule in the gas is larger than the mean value of the absolute value of its potential energy:

$$\overline{E_k} = \left|\overline{U}\right|.$$

In fact,

$$\overline{E} = \overline{E_k} + \overline{U},$$

where $\overline{U}$ is a negative value.

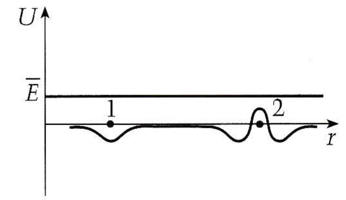


Figure 10

Liquids. In liquids and solids the molecules are situated at small distances from each other. Therefore, every molecule interacts with several nearest neighbors. Let's consider how two nearest molecules located at a distance of about $2r_0$ from each other affect the given (central) molecule.

The potential curve in question can be obtained by superimposing the curve plotted in figure 7 (two-body interaction) on the curve obtained by shifting the first curve slightly farther than $2r_0$. Since the values of the potential energy are added, the depth of the potential well increases almost twofold, while the energy peaks decrease (figure 11). When the interactions with other molecules are taken into account, the potential curve looks like that plotted in figure 12.

In order for a molecule to remain in the liquid, its mean energy must be negative ($\overline{E} < 0$). This is the prerequisite condition for a molecule to remain in the potential well that is formed by its neighbors. If $\overline{E} > 0$, the molecule will escape from the well and leave the liquid.

Since $\overline{E} = \overline{E_k} + \overline{U}$ and $\overline{U} < 0$, the mean kinetic energy in the liquid is less than the absolute value of the mean potential energy: $\overline{E_k} < |\overline{U}|$. Since this inequality is not very strong, $\overline{E_k}$ is only slightly lower

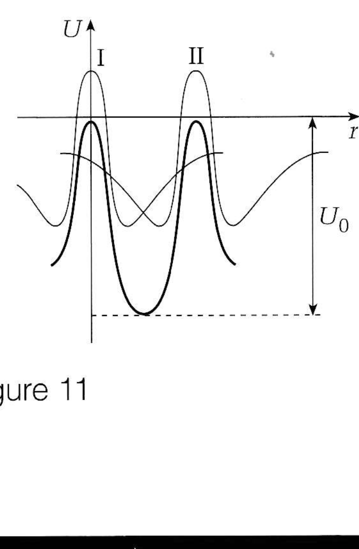


Figure 11

than the absolute value of the potential energy: $\overline{E_k} \le |\overline{U}|$ and $|\overline{E}| \le |U_0|$ (the minimal value of the potential energy). For this reason a molecule cannot remain in a well very long. The stochastic character of molecular motion causes the energy of the molecules to change continuously, sometimes reaching a value higher than the average energy, sometimes lower than it.

When the energy of a molecule exceeds the height of the potential barrier that separates one well from another, the molecule jumps from one equilibrium position to another. This is the principal feature that determines the character of the thermal motion and the fluidity in liquids. The mean energy of a molecule increases with increasing temperature, causing the frequency of jumps between the potential wells to increase.

Solids. In solids the potential energy of interaction of a molecule with its nearest neighbors resembles the potential energy of interaction in liquids (figure 12). However, the potential wells in solids are slightly

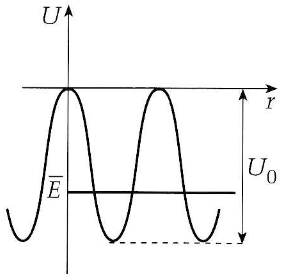


Figure 12

deeper than in liquids, because the molecules in solids are closer to each other. As in the case of liquids, the condition $\overline{E_k} < |\overline{U}|$ is satisfied in solids. However, the kinetic energy of the molecules in solids is considerably lower than in liquids. Solids, as we know, are formed by cooling a liquid. In solids the mean kinetic energy of molecules, therefore, is markedly lower than the absolute value of the potential energy:

$$\overline{E_k} \ll |\overline{U}|.$$

In figure 12 the mean energy of the molecule inside a potential well is represented by the straight line labeled $\overline{E}$. The molecule oscillates at the bottom of the potential well. The barriers separating the adjacent wells are so high that the molecules are confined in the wells and escape from their cells very rarely. To change one equilibrium position for another (that is, to jump from one well to another), a molecule must acquire an energy that is much larger than the average energy. Such an event occurs very rarely. For this reason solids, in contrast with liquids, retain their shape. ◘

Quantum on molecular interaction:

A. Eisenkraft and L. D. Kirkpatrick, "Cloud Formation," January/February 1995, pp. 36–38.

B. Yavelov, "Van der Waals and His Equation," November/December 1997, pp. 36–37.

A. Leonovich, "The Nature of Ideal Gas," May/June 1998, pp. 32–33.

V. Meshcheryakov, "Planetary Building Blocks," July/August 1998, pp. 4–10.

A. Dozorov, "Electric Multipoles," September/October 1999, pp. 4–8.

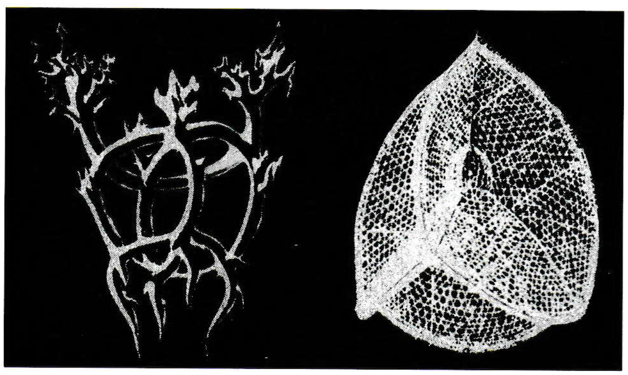


Figure 12

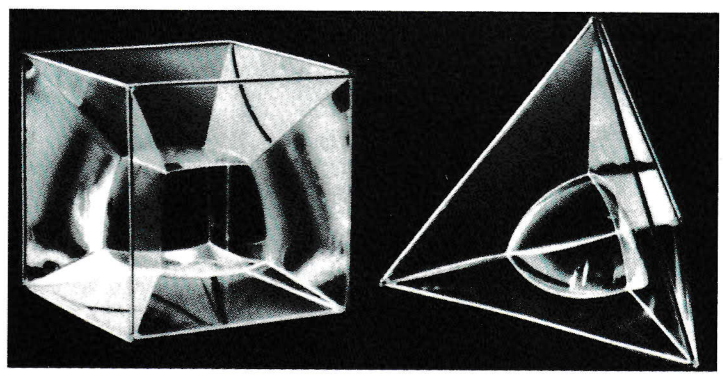


Figure 13

CONTINUED FROM PAGE 7

complex minimal surface. The cell membranes play a role in the formation of the solid skeleton of radiolaria, because they accumulate salt from the seawater. Figure 12 shows two radiolaria skeletons.

Figure 13 shows two minimal surfaces formed by soapbubbles in cubical and tetrahedral frames. Its similarity to figure 12 is amazing. The radiolaria skeletons accurately reproduce the structure of the branching edges of the minimal surfaces.

Minimal surfaces play an important role in chemistry, where the interaction at the boundaries of different media is responsible for the nature and rate of many chemical reactions. Well-known membranes, such as an ear-drum, membranes that surround cells in living organisms, or membranes that separate different organs from each other, are examples of minimal surfaces.

This interesting subject attracts experts in many fields such as biologists, chemists, and physicists. As we have seen, many interesting mathematical problems involve minimal surfaces. The correct statement and analysis of these problems would provide a better understanding of this wonderful natural phenomenon. ◘

Fermat's little theorem

Proving its value to mathematicians

by V. Senderov and A. Spivak

THE USUAL SCHOOL CURriculum does not include very much number theory. However, anyone interested in mathematics will eventually want to study this beautiful part of the subject and will want to know about Fermat's little theorem. This theorem is not just beautiful. It is also useful. The computer scientists R. Graham, D. Knuth, and O. Patashnik, the authors of the textbook entitled *Concrete Mathematics* for students of computer science, included this theorem in their book.

The theorem we present here was discovered by Pierre Fermat (a lawyer in Toulouse, France who lived from 1601 to 1665) in 1640. It can be stated very simply: *if p is a prime and a is an integer, then* $a^p - a$ *is a multiple of p.* It may not be obvious at first why this modest proposition is so important. But in fact it does deserve very close attention.

We start with some concepts that can be easily understood by a seventh grader and end with some recent discoveries in cryptography.

Particular cases

Of any two consecutive integers a and $a + 1$, one is even, and the other is odd. Therefore, the product $a(a + 1) = a^2 + a$ is even for any integer a.

We can prove that $a^2 + a$ is divisible by 2 in another way by considering two cases:

If a is even, then a^2 is also even, and the sum of the two even numbers a and a^2 is even;

If a is odd, then a^2 is also odd, and the sum of the two odd numbers a and a^2 is even.

This is a remarkable property of the polynomial $a^2 + a$. A special case of Fermat's little theorem concerns similar results about another polynomial, $a^2 - a = (a - 1)a$. All the integer values of this polynomial are also even.

Exercise 1. Prove this fact.

Let us consider the polynomial $a^3 - a$. It can be easily factored:

$$a^3 - a = a(a^2 - 1) = a(a - 1)(a + 1).$$

This is the product of three consecutive integers, so, as we have already seen, it must be even. But we can say more. Since one of the three consecutive integers is divisible by 3, their product $(a - 1)a(a + 1) = a^3 - a$ is a multiple of 3 (and thus it is a multiple of 6).

Exercise 2. Prove that for any integer a the sum $a^3 + 5a$ is a multiple of 6.

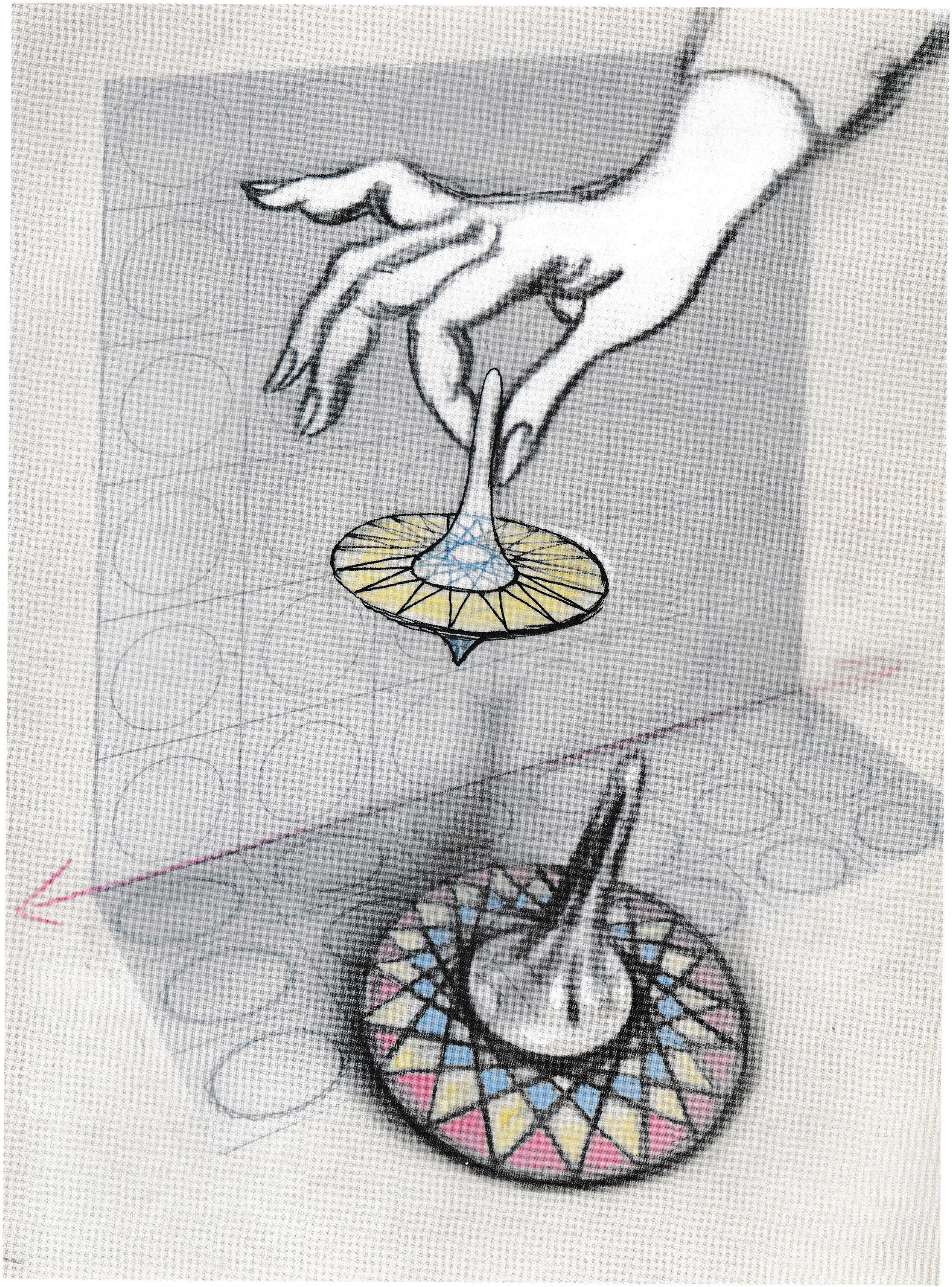
Let us continue examining integer values of the polynomial $a^n - a$ for various exponents n. What about the value of $n = 4$? For $a = 2$ and $a = 3$, the polynomial $a^4 - a$ takes the values of $2^4 - 2 = 14$ and $3^4 - 3 = 78$. These values are even and have no common divisor other than 2 (and 1). No luck! This bad luck has to do with the fact that 4 is a composite number. Fermat's little theorem says that the values of the polynomial $a^p - a$, for integer values of a, all have a common factor if p is prime. It says nothing about what happens when p is not prime.

Let $p = 5$. Let us calculate several values of the polynomial $a^5 - a$. For $a = \pm 1$ and for $a = 0$, we obtain zero. Further calculations yield $2^5 - 2 = 30$, $3^5 - 3 = 240$, $4^5 - 4 = 1020$, $5^5 - 5 = 3120$, $6^5 - 6 = 7770$, ... All these values are multiples of 30.

Let us prove that this pattern continues. Since $30 = 2 \cdot 3 \cdot 5$, the proof of the divisibility by 30 breaks down into three parts: first, we must prove that $a^5 - a$ is a multiple of 2, then that it is a multiple of 3, and, lastly, that it is a multiple of 5.

The first part is not difficult: a^5 and a have the same parity: either they are both even, or both odd. The second part is also straightforward:

Art by Yuri Vaschenko



$$a^5 - a = a(a^4 - 1) = a(a^2 - 1)(a^2 + 1) = a(a - 1)(a + 1)(a^2 + 1),$$

and the product of three consecutive integers is divisible by 3.

The third part of the proof is slightly more difficult. It is clear that one of any five consecutive integers is divisible by 5. Therefore, the product $(a - 2)(a - 1)a(a + 1)(a + 2)$ is a multiple of 5. Unfortunately, this is not the factorization of the polynomial $a^5 - a$, which factors as $(a - 1)a(a + 1)(a^2 + 1)$.

What can we do? A straightforward method is to try all remainders obtained upon division by 5: any integer gives a remainder of 0, 1, 2, 3, or 4. If the remainder is 0, then the second factor in $(a - 1)a(a + 1)(a^2 + 1)$ is divisible by 5. If the remainder is 1 or 4, then the first or third factor is divisible by 5. If, on the other hand, the remainder is 2 or 3, then we will have a fourth factor. For students who are unused to working with remainders, we give the following explanation: if $a = 5b + 2$, i.e., if a gives a remainder of 2 obtained upon division by 5, then $a^2 + 1 = (5b^2 + 2)^2 + 1 = 5(5b^2 + 4b + 1)$. The case $a = 5b + 3$ can be treated analogously.

Another approach to the same result is the following. We can write:

$$a^2 + 1 = (a - 2)(a + 2) + 5.$$

Therefore, if we are interested only in the remainders obtained upon division by 5, then $a^2 + 1$ can be replaced by $(a - 2)(a + 2)$. We can use this notation to express this idea:

$$a^2 + 1 \equiv (a - 2)(a + 2) \pmod 5.$$

The symbol "≡" is read "is congruent to," and "mod" is read "modulo." The notation was suggested by Gauss in 1801. By definition, a is congruent to b modulo n if $a - b$ is a multiple of n, i.e., if $a - b = kn$, where k is an integer.

The notation

$$a \equiv b \pmod n$$

is useful because the properties of congruences are similar to those of equalities. For example, congruences may be added: if $a \equiv b \pmod n$ and $c \equiv d \pmod n$, then $a + c \equiv b + d \pmod n$.

Let us prove this. By definition, $a = b + kn$ and $c = d + ln$, where k and l are integers. Therefore,

$$a + c = (b + kn) + (d + ln) = b + d + (k + l)n.$$

Q.E.D.

By analogy, the formulas

$$a - c = (b + kn) - (d + ln) = b - d + (k - l)n,$$

$$ac = (b + kn)(d + ln) = bd + knd + bln + kln^2 = bd + (kd + bl + kln)n$$

show that congruences can be subtracted and multiplied. Since they can be multiplied, they can also be raised to an integral power: if $a \equiv b \pmod n$, then for any natural number m the congruence $a^m \equiv b^m \pmod n$ is valid.

However, one can divide congruences only with some caution:

$$6 \equiv 36 \pmod{10},$$

but

$$1 \not\equiv 6 \pmod{10}.$$

Exercise 3. Solve the congruence $3x \equiv 11 \pmod{101}$.

Exercise 4. Which integers satisfy the congruence $14x \equiv 0 \pmod{12}$?

Exercise 5. Let $k \neq 0$. Prove that

(i) if $ka \equiv kb \pmod{kn}$, then $a \equiv b \pmod n$;

(ii) if $ka \equiv kb \pmod n$ and the numbers k and n are coprimes, then $a \equiv b \pmod n$.

Let us look further at the polynomials $a^p - a$. We shall prove that, for any integer a, $a^7 - a$ is divisible by 7. As before, we can consider all seven remainders obtained upon division by 7: $0^7 - 0 = 0$, $1^7 - 1 = 0$, $2^7 - 2 = 128 = 7 \cdot 18, \dots, 6^7 - 6 = 279{,}930 = 7 \cdot 39{,}990$. We can even be slightly economical: Since any integer can be represented as $a = 7b$, $7b \pm 1$, $7b \pm 2$, or $7b \pm 3$, we can consider only the cases $a = 0$, 1, 2, and 3 when checking Fermat's little theorem.

Mechanical checking, however, is not very useful. Factoring the polynomial is more revealing:

$$a^7 - a = a(a^6 - a) = a(a^3 - 1)(a^3 + 1) = a(a - 1)(a^2 + a + 1)(a + 1)(a^2 - a + 1).$$

Since

$$a^2 + a + 1 = (a^2 + a - 6) + 7 \equiv a^2 + a - 6 = (a - 2)(a + 3) \pmod 7$$

and since

$$a^2 - a + 1 \equiv a^2 - a - 6 = (a + 2)(a - 3) \pmod 7,$$

we have

$$a^7 - a \equiv a(a - 1)(a - 2)(a + 3)(a + 1)(a + 2)(a - 3) \pmod 7.$$

The product of seven consecutive integers is divisible by 7, so $a^7 - a$ must be divisible by 7 for any integer a. This proves Fermat's little theorem for the case $p = 7$.

Exercise 6. Prove that

(i) the greatest common divisor of the numbers $a^7 - a$ is 42;

(ii) the greatest common divisor of the numbers $a^9 - a$ is 30. (Note that 30 is not divisible by 9. This relates to the fact that 9 is not a prime number.)

Now consider the case $p = 11$. It is clear that

$$a^{11} - a = a(a^{10} - 1) = a(a^5 - 1)(a^5 + 1) = a(a - 1)(a^4 + a^3 + a^2 + a + 1)(a + 1)(a^4 - a^3 + a^2 - a + 1).$$

It is not so easy to guess, this time, what to do next. An exhaustive search through all eleven remainders is still possible. Such a search will show that the values of the polynomial $a^4 + a^3 + a^2 + a + 1$ are divisible by 11 for $a \equiv 3$, 4, 5, and 9 (mod 11), and the values of $a^4 - a^3 + a^2 - a + 1$ are divisible by 11 for $a \equiv 2$, 6, 7, and 8.

If we remove the parentheses in the product $(a - 3)(a - 4)(a - 5)(a - 9)$, we obtain

$$(a^2 - 7a + 12)(a^2 - 14a + 45) \equiv (a^2 + 4a + 1)(a^2 - 3a + 1) = a^4 + a^3 - 10a^2 + a + 1 \equiv a^4 + a^3 + a^2 + a + 1 \pmod{11}.$$

Similarly, we can verify that

$$(a - 2)(a - 6)(a - 7)(a - 8) \equiv a^4 - a^3 + a^2 - a + 1 \pmod{11}.$$

What do we do next? For $p = 13$, our method requires that we raise the numbers 1 through 12 to the power of twelve and remove parentheses in the product of thirteen factors $a - 6, a - 5, \dots, a + 5, a + 6$. This is very tedious, even if we use only

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$4k$	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64	68	72
$4k$ mod 19	4	8	12	16	1	5	9	13	17	2	6	10	14	18	3	7	11	15

Table 1

the numbers 1, 2, 3, 4, 5, and 6 and multiply "only" six parentheses $(a^2 - 1)(a^2 - 4)(a^2 - 9)(a^2 - 16)(a^2 - 25)(a^2 - 36)$.

As p increases, the number of cases to be looked at increases. To prove Fermat's little theorem in general, we must use more powerful arguments.

Exercises.

7. (i) Prove that the product of any four consecutive integers is divisible by 24.

(ii) Prove that the product of any five consecutive integers is divisible by 120.

(iii) Prove that $a^5 - 5a^3 + 4a$ is divisible by 120 for any integer a.

8. Prove that, for any natural a, a^5 ends with the same digit as a (in decimal notation).

9. Prove that $m^5n - mn^5$ is divisible by 30 for any integer m and n.

10. Prove that if k is not divisible by any of the numbers 2, 3, and 5, then $k^4 - 1$ is divisible by 240.

11. (i) Prove that $2222^{5555} + 5555^{2222}$ is divisible by 7.

(ii) Find the remainder upon division by 7 of $(13^{14} + 15^{16})^{17} + 18^{19^{20}}$.

12. Prove that the number $11^{10} - 1$ ends in two zeros (i.e., it is divisible by 100).

13. (i) Find all integers a such that $a^{10} + 1$ ends in zero.

(ii) Prove that the number $a^{100} + 1$ cannot end in zero for any integer a.

14. Let n be an even number. Find the greatest common divisor of the numbers $a^n - a$, where a is an integer.

15. Let n be a natural number greater than 1. Prove that the greatest common divisor of the numbers $a^n - a$, where a runs through all the integers, coincides with the greatest common divisor of the numbers $a^n - a$, where $a = 1, 2, 3, \ldots, 2^n$. Notice that this fact implies that the greatest common divisor of the numbers $a^n - a$, where a is an integer, coincides with the greatest common divisor of the numbers of the type listed, where a is a natural number.

The general case

Let us write the numbers 1, 2, 3, ..., $p - 1$, multiply each of them by a, where a is not divisible by p, and consider the remainders upon division of these products by p. For example, for $p = 19$ and $a = 4$, we obtain table 1.

The lower row of the table contains the same numbers as its upper row. However, they are arranged in a different order! It turns out that this is a general rule. Not only for $p = 19$ and $a = 4$, but also *for any prime p and any integer a which is not a multiple of p, we obtain the same remainders* 1, 2, 3, ..., $p - 1$ *in some order.*

Why is it so? First of all, the lower row cannot contain 0, because the product of two numbers a and b, neither of which are multiples of a prime p, cannot be divisible by p. Second, all numbers in the lower row are different. This fact can be proved indirectly: if the numbers xa and ya have the same remainders upon division by p (where x and y are both less than p), then their difference $xa - ya = (x - y)a$ is divisible by p. However, this cannot be true because $x - y$ is not divisible by p (it's too small). These two simple observations are enough: there exist exactly $p - 1$ different remainders obtained upon division by p, and all of them must appear once in the lower row of the table.

Exercises.

16. Does a natural number n exist such that the number $1999n$ ends in the digits 987654321?

17. Prove that if an integer k is relatively prime to a natural number n, then there exists a natural number x such that $kx - 1$ is divisible by n.

18. Prove that if the integers a and b are relatively prime, then any integer c can be represented as $c = ax + by$, where x and y are integers.

We recall that Fermat's little theorem asserts that, for any integer a and prime p, the number $a^p - a = a(a^{p-1} - 1)$ is divisible by p. Thus, for the numbers a that are not multiples of p, this theorem can be formulated as follows.

Theorem 1. *If an integer a is not divisible by the prime p, then the remainder obtained upon division of a^{p-1} by p is 1.*

Proof. The remainders obtained upon division by p of the numbers a, $2a$, $3a$, ..., $(p - 1)a$ are 1, 2, 3, ..., $p - 1$ (to within a permutation). Therefore,

$$a \cdot 2a \cdot 3a \cdot \ldots \cdot (p-1)a \equiv 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (p-1) \pmod{p}.$$

We thus have

$$a^{p-1}(p-1)! \equiv (p-1)! \pmod{p}.$$

Dividing both sides by $(p - 1)!$, we obtain the desired congruence

$$a^{p-1} \equiv 1 \pmod{p}.$$

Without using exercise 4(ii) or congruences, we may also reason as follows. Since the product $(a^{p-1} - 1)(p - 1)!$ is divisible by p and since $(p - 1)!$ is not divisible by p, the number $a^{p-1} - 1$ is divisible by the prime p.

Exercises.

19. Find the remainder when 3^{2000} is divided by 43.

20. Prove that if an integer a is not divisible by 17, then either $a^8 - 1$ or $a^8 + 1$ is divisible by 17.

21. Prove that $m^{61}n - mn^{61}$ is divisible by 56,786,730 for any integers m and n.

22. Find all primes p such that $5^{p^2} + 1$ is divisible by p.

23. Let p be a prime number other than 2. Prove that $7^p - 5^p - 2$ is divisible by $6p$.

24. Prove that for any prime p the sum $1^{p-1} + 2^{p-1} + \ldots + (p-1)^{p-1}$ leaves a remainder of $p - 1$ upon division by p.

25. Choose any six-digit number divisible by 7. Remove its leftmost digit, and append it to the right. Prove that the number thus obtained is also divisible by 7. For example, from the number 632,387, which is divisible by 7, we obtain 323,876, which is also divisible by 7, and from the number 200,004 we obtain 42, both of which are divisible by 7.

26. Let p be a prime number distinct from 2, 3, and 5. Prove that the number consisting of $p - 1$ digits 1 is divisible by p (for example, 111,111 is divisible by 7).

27*. Prove that, for any prime number p, the number 11...1122...22...99...99 consisting of $9p$ digits (first p ones, then p twos, then p threes ..., and, finally, p nines) gives the same remainder upon division by p as the number 123,456,789.

×	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	9	10
2	2	4	6	8	10	1	3	5	7	9
3	3	6	9	1	4	7	10	2	5	8
4	4	8	1	5	9	2	6	10	3	7
5	5	10	4	9	3	8	2	7	1	6
6	6	1	7	2	8	3	9	4	10	5
7	7	3	10	6	2	9	5	1	8	4
8	8	5	2	10	7	4	1	9	6	3
9	9	7	5	3	1	10	8	6	4	2
10	10	9	8	7	6	5	4	3	2	1

Table 3

Multiplication tables

Let us consider all $n - 1$ different nonzero remainders upon division by n. Let us construct a multiplication table by putting the remainder obtained as a result of division of ab by n at the intersection of column a and row b. For example, for $n = 5$ we have

×	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

Table 2

For $n = 11$ we have table 3.

In both examples, n is a prime number. Therefore, every row and every column contain a permutation of the numbers 1, 2, ..., $n - 1$. If, however, we consider a composite number, the table will definitely include zero. For example, for $n = 4$ we have

×	1	2	3
1	1	2	3
2	2	0	2
3	3	2	1

Table 4

The situation for $n = 12$ is similar (table 5). Again, some rows contain zeros! In general, for any composite $n = ab$, where $1 < a < b < n$, the intersection of row a and column b contains the remainder obtained upon division of ab by n, which is 0.

×	1	2	3	4	5	6	7	8	9	10	11
1	1	2	3	4	5	6	7	8	9	10	11
2	2	4	6	8	10	0	2	4	6	8	10
3	3	6	9	0	3	6	9	0	3	6	9
4	4	8	0	4	8	0	4	8	0	4	8
5	5	10	3	8	1	6	11	4	9	2	7
6	6	0	6	0	6	0	6	0	6	0	6
7	7	2	9	4	11	6	1	8	3	10	5
8	8	4	0	8	4	0	8	4	0	8	4
9	9	6	3	0	9	6	3	0	9	6	3
10	10	8	6	4	2	0	10	8	6	4	2
11	11	10	9	8	7	6	5	4	3	2	1

Table 5

Thus, if n is composite, there exist *divisors of zero*, i.e., nonzero remainders a and b such that the product ab is divisible by n. However, even for composite n, some rows of the multiplication table do not contain zeros. In table 4, they are the first and the third rows; in table 5 they are the first, fifth, seventh, and eleventh rows. It is easy to see that the rows with numbers that have a common divisor with n (different from 1), and only such rows, include zeros.

Exercise 28. Prove this statement.

Let us remove from the table all rows that contain zeros. (If n is prime, then there is nothing to remove.) For $n = 4$, we obtain a table consisting of only two rows and two columns:

×	1	3
1	1	3
3	3	1

Table 6

For $n = 12$, we have a 4 × 4 table:

×	1	5	7	11
1	1	5	7	11
5	5	1	11	7
5	5	11	1	5
11	11	7	5	1

Table 7

Exercise 29. Note that each of the tables 2–7 is symmetric with respect to its two diagonals. Prove that this is the case for any n.

Euler's theorem

To generalize Fermat's little theorem for the case of composite numbers n, we retain in the multiplication table only the rows and columns that contain no zeros. In other words, we consider the remainders that are relatively prime to n and that are obtained upon division by n. In the new table, the rows (or the columns) differ from each other only in the order of the numbers.

Exercise 30. If we denote by a_1, a_2, ... , a_r the remainders obtained when we divide by the natural number a and which are relatively prime to n, and then multiply each remainder by some number k that is relatively prime to n, we obtain the numbers ka_1, ka_2, ... , ka_r, which are also relatively prime to n and which also give different remainders when divided by n. Prove this statement.

Because of the result of exercise 30, the sequence of the remainders obtained upon division of the numbers $ka_1, ka_2, \ldots, ka_r$ by n can therefore differ from the sequence $a_1, a_2, \ldots, a_r$ only in the order of the elements. By analogy with the case for prime numbers p, we obtain the following relationship for the composite number n:

$$ka_1\, ka_2 \ldots ka_r \equiv a_1\, a_2 \ldots a_r \pmod{n}.$$

We thus have

$$(k^r - 1)\, a_1\, a_2 \ldots a_r \equiv 0 \pmod{n}.$$

Therefore, the product $(k^r - 1)\, a_1\, a_2 \ldots a_r$ is divisible by n. Since $a_1, a_2, \ldots, a_r$ are coprime to n, $k^r - 1$ is divisible by n. If n is a prime number, then $r = n - 1$, and we have the assertion of Fermat's little theorem. The general statement of Euler's theorem is the following.

Theorem 2. *If k is an integer relatively prime to a natural n, then $k^r - 1$ is divisible by n, where r is the number of natural numbers that are less than and relatively prime to n.*

Exercises

31. Prove that if k is not divisible by 3, then

(i) k^3 gives a remainder of 1 or 8 upon division by 9;

(ii) k^{81} gives a remainder of 1 or 242 upon division by 243.

32. (i) Prove that if $a^3 + b^3 + c^3$ is divisible by 9, then at least one of the numbers a, b, or c is divisible by 3.

(ii) Prove that the sum of squares of three integers is divisible by 7 if and only if the sum of the fourth powers of these numbers is divisible by 7.

33. Prove that $7^{7^{7^{7^{7}}}} - 7^{7^{7}}$ is divisible by 10.

34. What are the three last digits in the number 7^{9999}?

35. Prove that, for odd positive integers n, $2^{n!} - 1$ is divisible by n.

36*. Find all natural $n > 1$, for which the sum $1^n + 2^n + \ldots + (n-1)^n$ is divisible by n.

37*. For all natural numbers s, there exists a natural number n that is a multiple of s and such that the sum of the digits of n is equal to s.

Euler's function

In 1763, Leonard Euler (1707–1783) introduced the notation $\varphi(n)$ for the number r of the remainders (upon division by n) that are relatively prime to n. For example, $\varphi(1) = 1$, $\varphi(4) = 2$, and $\varphi(12) = 4$.

If p is prime, then $\varphi(p) = p - 1$. It is easy to calculate $\varphi(p^m)$, where m is a natural number. Let us write all p^m possible remainders: 0, 1, 2, ... , $p^m - 1$. Among these numbers, only the remainders $0, p, 2p, \ldots, p^m - p$ are divisible by p. Therefore,

$$\varphi(p^m) = p^m - p^{m-1} = p^m\left(1 - \frac{1}{p}\right).$$

We now calculate $\varphi(1000)$, i.e., the number of numbers from the first thousand that are divisible neither by 2 nor by 5. For this purpose, we subtract from 1000 the number of even numbers in the first thousand; we also subtract 200, which is the number of multiples of 5 in the first thousand. We must also take into account that some numbers (those ending in the digit 0) are divisible by 2 and by 5. There are 100 such numbers, each of which was counted twice. The correct result, therefore, is given by the formula

$$\varphi(1000) = 1000 - 500 - 200 + 100 = 400.$$

Exercises

38. Find $\varphi(2^a 5^b)$, where a and b are natural numbers.

39. Let p and q be different prime numbers. Find (i) $\varphi(pq)$ and (ii) $\varphi(p^a q^b)$, where a and b are natural numbers.

40. Solve the equations (i) $\varphi(7^x) = 294$; (ii) $\varphi(3^x 5^y) = 360$.

In principle, the method we have used allows us to calculate $\varphi(n)$ for any natural number n. For example, to calculate $\varphi(300)$ we write the numbers 1 through 300 and remove from this list 150 even numbers, 100 numbers divisible by 3, and 60 numbers divisible by 5. We recall that some numbers were removed twice (and some even three times). We must also restore this injustice by subtracting a count of 50 numbers divisible by $2 \cdot 3 = 6$, 30 numbers divisible by $2 \cdot 5 = 10$, and 15 numbers divisible by $3 \cdot 5 = 15$. However, each of the ten numbers that are divisible by $2 \cdot 3 \cdot 5 = 30$ was first removed three times (because each is divisible by 2, 3, and 5) and then restored three times (a multiple of 6, 10, and 15). We still must remove these ten numbers. We thus have

$$\varphi(300) = 300 - 150 - 100 - 60 + 50 + 30 + 20 - 10 = 80.$$

As you see, this method is rather simple. However, as the number of prime divisors of n increases, the number of terms in the resulting formula also increases.

Theorem 3. *The Euler's function is* multiplicative. *That is, if m and n are relatively prime natural numbers, then*

$$\varphi(mn) = \varphi(m)\varphi(n).$$

Corollary. *If $n = p_1^{a_1} p_2^{a_2} \cdot \ldots \cdot p_s^{a_s}$, where $p_1, p_2, \ldots, p_s$ are different prime numbers and $a_1, a_2, \ldots, a_s$ are natural numbers, then*

$$\varphi(n) = \varphi\left(p_1^{a_1}\right)\varphi\left(p_2^{a_2}\right)\cdot \ldots \cdot \varphi\left(p_s^{a_s}\right)$$
$$= \left(p_1^{a_1} - p_1^{a_1-1}\right)\left(p_2^{a_2} - p_2^{a_2-1}\right)\cdot \ldots$$
$$\ldots \cdot \left(p_s^{a_s} - p_s^{a_s-1}\right).$$

Proof of Theorem 3. Consider the numbers $mx + ny$, where $0 \le x < n$ and $0 \le y < m$. They can be represented as a table of dimension $n \times m$. For example, for $n = 5$ and $m = 8$

we have

x/y	0	1	2	3	4	5	6	7
0	0	5	10	15	20	25	30	35
1	8	13	18	23	28	33	38	43
2	16	21	26	31	36	41	46	51
3	24	29	34	39	44	49	54	59
4	32	37	42	47	52	57	62	67

Table 8

The remainders obtained upon division of the numbers from this table by mn are all different. If two remainders were equal, then the following congruence would be true:

$$mx_1 + ny_1 \equiv mx_2 + ny_2 \pmod{mn},$$

where $0 \le x_1, x_2 < n$, and $0 \le y_1, y_2 < m$. This congruence implies two other congruences:

$$mx_1 + ny_1 \equiv mx_2 + ny_2 \pmod{m}$$

and

$$mx_1 + ny_1 \equiv mx_2 + ny_2 \pmod{n}$$

The first of them implies the congruence

$$ny_1 \equiv ny_2 \pmod{m}.$$

Since m and n are relatively prime, we find from this congruence that

$$y_1 \equiv y_2 \pmod{m}.$$

Recalling that $0 \le y_1$ and $y_2 < m$, we obtain $y_1 = y_2$. By analogy, the congruence modulo n yields the equality $x_1 = x_2$.

Thus all mn numbers in the table give different remainders upon division by mn. But there are as many numbers in the table as there are different remainders upon division by mn. Therefore, the numbers in the table exhaust all possible remainders. In other words, *for any number* $d = 0, 1, \ldots, mn - 1$, *there exists a unique pair of integers* x, y *such that* $0 \le x < n$, $0 \le y < m$, *and* $d \equiv mx + ny \pmod{mn}$.

In table 8, the even numbers occupy four columns, while multiples of 5 occupy only one row. This is part of a more general rule:

$$\text{GCD}(mx + ny, m) = \text{GCD}(ny, m) = \text{GCD}(y, m).$$

By analogy, $\text{GCD}(mx + ny, n) = \text{GCD}(x, n)$. For this reason, the table that we are considering has $\varphi(m)$ columns containing numbers that are relatively prime to m (these are columns, where y is relatively prime to m) and $\varphi(n)$ rows containing numbers that are relatively prime to n.

The proof of Theorem 3 now becomes easy: in order for d to be relatively prime to mn, it is necessary and sufficient that d be relatively prime to m and n. Such numbers d occupy the intersections of $\varphi(m)$ columns (which contain numbers that are relatively prime to m) and $\varphi(n)$ rows (which contain numbers that are relatively prime to n). We thus have a "grid" that contains a total of $\varphi(m)\varphi(n)$ numbers. Q.E.D.

Exercises

41. Consider the table of m rows and n columns (table 9) that contains the numbers 0 through $mn - 1$.

(i) Construct such a table for $m = 3$ and $n = 4$. In this table, strike out all even numbers and then those of the remaining numbers that are divisible by 3. Notice that the remaining numbers are exactly those that are relatively prime to 12; Notice that the remaining numbers do not form a grid.

0	1	2	...	$n-1$
n	$n+1$	$n+2$	...	$2n-1$
$2n$	$2n+1$	$2n+2$	...	$3n-1$
...	...	...	...	...
$(m-1)n$	$(m-1)n+1$	$(m-1)n+2$	...	$mn-1$

Table 9

42. Complete the following outline of a second proof of Euler's theorem:

(1) the numbers that are relatively prime to n occupy $\varphi(n)$ columns in table 9;

(2) the remainders upon division of all m numbers of any row in table 9 by m are all different;

(3) each column contains exactly $\varphi(m)$ numbers that are relatively prime to m;

(4) a number is relatively prime to mn if and only if it is relatively prime to n [such numbers are in the $\varphi(n)$ columns] and relatively prime to m [each column contains $\varphi(m)$ such numbers].

43. A circle is divided into n equal parts by n points. How many different closed polygonal paths exist that are made up of n equal segments whose vertices are at these points? (Two polygonal lines that can be obtained from each other by a rotation are considered identical. Figure 1 shows all such lines for $n = 20$.)

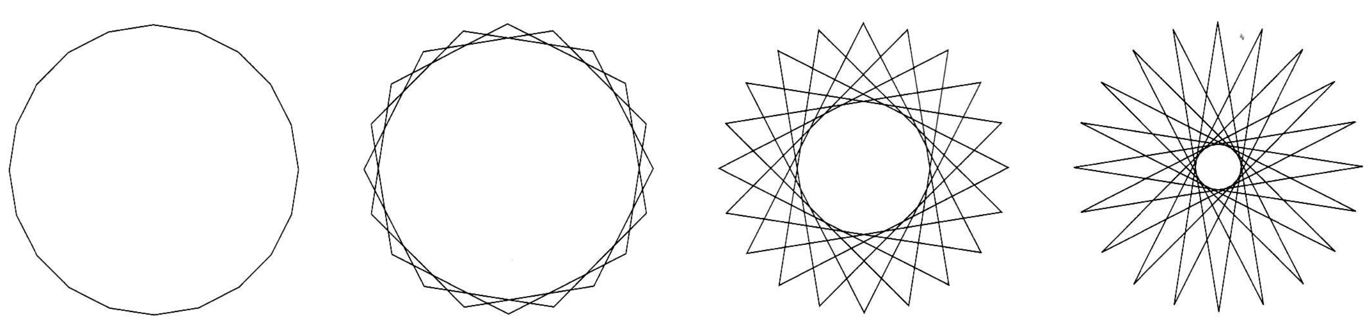


Figure 1

44. For any natural numbers m and n, prove the following equalities:

(i)

$$\varphi(m)\varphi(n) = \varphi(\mathrm{LCM}[m, n])\varphi(\mathrm{GCD}(m, n));$$

(ii)

$$\varphi(mn) = \varphi(\mathrm{LCM}\,[m, n])\,\mathrm{GCD}(m, n);$$

(iii)

$$\varphi(m)\varphi(n)\mathrm{GCD}(m, n) = \varphi(mn)\varphi(\mathrm{GCD}(m, n)).$$

(iv) Let m and n be natural numbers, and $\mathrm{GCD}(m, n) > 1$. Prove that $\varphi(mn) > \varphi(m)\varphi(n)$.

45. Solve the equations

(i) $\varphi(x) = 18$;
(ii) $\varphi(x) = 12$;
(iii) $x - \varphi(x) = 12$;
(iv)* $\varphi(x^2) = x^2 - x$;
(v) $\varphi(x) = x/2$;
(vi) $\varphi(x) = x/3$;
(vii)* $\varphi(x) = x/n$, where n is a natural number and $n > 3$;
(viii) $\varphi(nx) = \varphi(x)$, where n is a natural number and $n > 1$.

Open-key encryption

Let us assume that we need to receive an encoded message from a friend, but we did not come to an agreement with him in advance about the cipher to be used. What shall we do? Does a method exist such that it could be made public, and anyone (both your friends and your enemies) could encrypt messages using this method, but only your friends could decipher those messages?

This would be a remarkable method. Other encryption methods are based on the use of a secret key for encoding and decoding messages. This new method would have an "open key." That is, anybody could encode a message, but only the author of the encryption system can decode it.

RSA encryption. An encryption method with an open key was invented in 1978. That year, three mathematicians, R. A. Rivest, A. Shamir, and L. Adleman, encrypted an English phrase and promised a reward of 100 dollars to the first person who could decode it. Here is the encrypted phrase:

y = 96869613754622061477140922 25435588290575999112457431987 46951209308162982251457083569 31476622883989628013391990551 829945157815154.

The encryption method was explained in detail. First, the original phrase was represented as a sequence of digits by straightforward substitution (the letter a was encoded by 01, the letter b by 02, the letter c by 03, ..., the letter z by 26, and a space by 00.) Thus, the original phrase was written as a 78-digit number x. A prime 64-digit number p and a prime 65-digit number q were then multiplied (obviously by computer) to obtain

pq = 11438162575788886766932577 99761466120102182967212423625 62561842935706935245733897830 59712356395870505898907514759 9290026879543541.

Now, here is the key point:

$$y \equiv x^{9007} \pmod{pq}.$$

Do you see it? They published the product pq, the number 9007, the encryption method, and, of course, the number y. They even revealed that the number p consists of 64 digits, and q of 65 digits. Only the numbers p and q remained unknown. The task was to find x.

The solution was given in 1994 by Atkins, Craft, Lenstra, and Leiland. The numbers p and q were

p = 34905295108476509491478496 19903898133417764638493387843 990820577,
q = 32769132993266709549961988 19083446141317764296799294253 9798288533.

The book, entitled *Introduction to Cryptography*, says: "This remarkable result (the decomposition of a 129-digit number into prime factors) was obtained by using an algorithm called the quadratic-sieve method (this method was designed for prime factorization). The calculations required enormous resources. About 600 people participated in the project, headed by four persons who thought of it. About 1600 computers, joined by the internet, were used."

A description of the quadratic-sieve method, however, would divert us considerably from the topic of this article. We shall therefore leave its description for another time and briefly discuss here the basic idea of the RSA encryption method (the acronym RSA stands for the authors' names — Rivest, Shamir, and Adleman).

This idea is extremely elegant. First, if p and q are known, we can find $\varphi(pq) = (p - 1)(q - 1)$. Secondly, (and this is an important idea) if

$$ef = 1 + k\varphi(pq),$$

where e, f, and k are natural numbers, then for any number x that is relatively prime to pq, we find, by Euler's theorem, that

$$x^{ef} = x \cdot \left(x^k\right)^{\varphi(pq)} \equiv x \cdot 1 = x \pmod{pq}.$$

Do you see the importance of the numbers e and f? In our example, e = 9007 [the only requirement for e is that it must be relatively prime to $(p - 1)(q - 1)$; however, it is not wise to use $e = 1$ or $e = (p - 1)(q - 1)$ if you want to keep your secrets]. As has already been mentioned above, f is the solution of the congruence

$$ef \equiv 1 \pmod{\varphi(pq)}.$$

An algorithm for solving such congruences based on Euclid's algorithm is described in the Appendix.

The congruences

$$y^f \equiv x^{ef} \equiv x \pmod{pq}$$

show that it is sufficient to find the remainder upon division of y^f by pq to calculate x. The numbers are chosen in such a way that $x < pq$, where x is divisible by neither p nor q. This circumstance, however, does not impose any severe restrictions on us: if p and q are large numbers, the probability that x is divisible by p or q is negligible. In addition, we can arrange that the encryption algorithm will, if necessary, slightly change the message being encrypted (without changing its meaning) such that x and pq become relatively prime.

Why do many people think that the RSA encryption algorithm has an open key? Because the numbers pq and e can be made public. Anyone having a computer (and a program that can operate with multidigit numbers) can then encrypt a message. The message can be easily decoded if the number f is known. However, the only available method for finding f requires that the numbers p and q be known. That is, pq must be factorized into primes. Currently, no efficient methods are available for solving the latter problem. The successes achieved in 1994 do not count: if the numbers p and q would consist of, say, 300 digits or more, then no internet resources would be sufficient. On the other hand, we have no proof that it is impossible to find an efficient algorithm (that is, an algorithm whose execution time depends on the number of digits in a polynominal fashion) for factoring integer numbers into primes.

Appendix

A method for raising to a higher power. In order to raise a number x to the power 9007, it is sufficient, by definition, to perform 9006 multiplications. It is possible, however, to reduce the number of operations: we can calculate x^2, $(x^2)^2 = x^4$, $(x^4)^2 = x^8$, ... , $(x^{2048})^2 = x^{4096}$, and, finally, $(x^{4096})^2 = x^{8192}$, and then use the formula

$$x^{9007} = x \cdot x^2 \cdot x^4 \cdot x^8 \cdot x^{32} \cdot x^{256} \cdot x^{512} \cdot x^{8192},$$

which is based on the binary representation of 9007:

$$9007_{10} = 10\ 0011\ 0010\ 111_2.$$

We represented 9007 as the sum 1 + 2 + 4 + 8 + 32 + 256 + 512 + 8192 and got by with only 20 multiplications (13 squarings and 7 multiplications) instead of 9006. The saving of computational effort is enormous. For an alert (fault-finding) reader, we note that in the above considerations, we should have discussed multiplication modulo pq, rather than conventional multiplication. To keep the number of digits manageable, we must not only calculate the product at each step, but also the remainder upon division by pq.

The advantages of this method for raising to a power increase as the power increases. For example, if the exponent of the power consists of several dozen or hundreds of digits, the straightforward method of raising to this power is impractical even if the most powerful computers are used. But the method based on the binary representation works even for such big numbers.

Exercise 46.* Suppose that we may perform two operations: *multiply a number by* 2 and *increase a number by* 1. If the binary representation of the number n is $\overline{a_m a_{m-1} \ldots a_1 a_0}$, what is the minimum number of operations required for obtaining from 0 the number (i) 100; (ii) 9907; (iii) n ?

Euclidean algorithm. Euclidean algorithm is a method for finding the largest common divisor. It is based on the formula

$$\text{GCD}(a, b) = \text{GCD}(a - bq, b),$$

which holds for any integer numbers a, b, and q.

Exercise 47. Prove this fact.

Here we need a method for solving linear equations based on Euclidean algorithm rather than the algorithm itself.

Suppose we are given relatively prime e and m [in the case given above, $m = \varphi(pq)$]. We must find numbers f and k such that

$$ef = 1 + km.$$

If m is not very large, an exhaustive search of all m remainders is possible. For large m, an exhaustive search is impractical. It turns out that Euclidean algorithm provides a fast method for solving this problem.

Let us demonstrate how this method works by considering the example with e = 9007 and m = 19876. (We had originally wanted to use a value of m with more than one hundred digits, but lost our nerve at the last moment.) The equation

$$9007f = 1 + 19876k$$

can be written as

$$9007f = 1 + 9007 \cdot 2k + 1862k,$$

or as

$$9007(f - 2k) = 1 + 1862k.$$

Let $a = f - 2k$. Then

$$9007a = 1 + 1862k.$$

Notice that this is an equation of the same type as the original equation, but with smaller coefficients. The next step is

$$1862 \cdot 4a + 1559a = 1 + 1862k,$$

i.e.,

$$1559a = 1 + 1862(k - 4a).$$

Let $k - 4a = b$. Then

$$1559a = 1 + 1862b.$$

We rewrite this equation as

$$1559(a - b) = 1 + 303b.$$

Letting $a - b = c$, we obtain the equation

$$1559c = 1 + 303b.$$

Similar transformations yield

$$44c = 1 + 303(b - 5c),$$
$$44(c - 6d) = 1 + 39d,$$
$$5x = 1 + 39(d - x),$$

$$d = b - 5c,$$
$$x = c - 6d,$$
$$y = d - x,$$

$$44c = 1 + 303d,$$
$$44x = 1 + 39d,$$
$$5x = 1 + 39y.$$

The computer would continue the calculations until the coefficient of one of the unknowns becomes 1. We can stop here since it is clear that x = 8, y = 1 is a solution to the last equation. If x and y are known, we find

$$d = x + y = 9,$$
$$c = x + 6d = 62,$$
$$b = d + 5c = 319,$$
$$a = b + c = 381,$$
$$k = b + 4a = 1843,$$
$$f = a + 2k = 4067.$$

We have triumphed! Here are the values of f and k, and here is the verification:

$$9007 \cdot 4067 = 36{,}631{,}469 = 1 + 19876 \cdot 1843.$$

HOW DO YOU FIGURE?

Challenges

Math

M291

Corrupt copies. A student solved all the problems in a mathematics olympiad. Before sending his solutions by mail, he gave them to his two friends to copy. The next day, these two students copied the solutions. However, each of them made several (different) errors when copying. Before sending these copied solutions, the two students, in turn, gave the solutions to four other students (each of them to two friends). The next day, these four students repeated the procedure, and so on. Each new copy retained all the previous errors and may have added new errors. It is known that some day each new copy will contain at least ten errors. Prove that there is a day when a total of at least eleven new errors are made in the copies.

M292

What's your angle? I. In triangle ABC angle A is equal to α. The circle that passes through A and B and is tangent to line BC intersects the median drawn to side BC (or its extension) at a point M, which is different from A. Express the measure of angle BMC in terms of α.

M293

Roots of the problem. Let x_1 and x_2 be the roots of the equation $x^2 + px + q = 0$. Prove that if t satisfies the inequalities

$$x_1 \le \frac{t^2 - q}{2t + p} \le x_2,$$

then t must in fact be equal to either x_1 or x_2.

M294

What's your angle? II. In triangle ABC, $\angle BAC = \alpha$ and $\angle ABC = 2\alpha$. The circle with center C and radius CA intersects the line containing the bisector of the exterior angle at vertex B at points M and N. Express the measures of the angles of triangle AMN in terms of α.

M295

Proof positive. Let positive numbers $a_1, a_2, \ldots, a_{100}$ be such that

$$\frac{1}{1+a_1} + \frac{1}{1+a_2} + \ldots + \frac{1}{1+a_{100}} \le 1.$$

Prove that $a_1 \cdot a_2 \cdot \ldots \cdot a_{100} \ge 99^{100}$.

Physics

P291

Tugged boat. A motor is located on a cliff at the shore of a lake. The motor winds a rope uniformly on its drum and pulls a boat directly to the shore. At a certain instant the rope makes an angle α with respect to the horizontal, and the boat's speed is v. At this instant, a small knot on the rope is half as far from the boat as it is from the motor. Find the velocity and acceleration of the knot at this instant of time. (S. Varlamov)

P292

Water world. The surface of a planet, which has the same size, mass, and atmospheric composition as Earth, is covered entirely with an ocean whose temperature is +10°C and whose constant depth is 230 m. An internal process heats the ocean to a temperature +100 °C, but the depth of the ocean and the size of the planet's hard core remain the same. Assuming that the size of the solid part of the planet remains the same during the heating, find the mean thermal expansion of water in the indicated temperature range. (S. Varlamov)

P293

Spear-ited debate. What speed is required for a long, thin spear of mass M, which is uniformly charged by a positive charge Q along its length L, for the spear to pass completely through two adjacent layers of thickness h, in which the electric field is directed both against the velocity of the spear (in the first layer) and along it (in the second layer). In each case the intensity of the electric field is E. The total thickness of the two layers is less than the length of the spear. (O. Savchenko)

P294

Birth of the Earth. At present, natural uranium contains η_1 = 99.28% uranium-238 and η_2 = 0.72% uranium-235. The nuclear half-life of ^{238}U is $\tau_1 = 4.56 \cdot 10^9$ years and that of ^{235}U is $\tau_2 = 0.71 \cdot 10^9$ years. Assuming that the numbers of each uranium isotope were identical at the Earth's birth, find the age of our planet. (V. Mozhaev)

CONTINUED ON PAGE 26

High-speed hazards

Crafting a solution

by I. Vorobyov

EVEN THE NEAREST STARS are very distant objects. As for the galaxies, the distances to them are mind-boggling. For example, it takes light about a million years to travel from the Andromeda Galaxy to our Solar System.

Can we ever hope to reach the Andromeda Galaxy, say, in a week? According to clocks on Earth, this is utterly impossible, as no spacecraft can travel faster than the speed of light. However, this is possible for an interstellar traveler aboard the spacecraft if the spacecraft moves at a sufficiently high speed. At relativistic speeds clocks (and all other processes) slow down. The higher the speed, the more the clocks slow down.

How should one travel to Andromeda if one must get there in a week? The distance of one million light years must be covered in 1/52 of a year. Calculations based on the theory of relativity give an incredible result: the flight must occur with a constant acceleration $a = 700\ g$! Here g is acceleration due to gravity at Earth's surface.

However, how can one deal with such accelerations? At an acceleration of 700 g our weight increases 700 fold! This would be like having a steamroller roll over the star traveler. Our blood, which must circulate throughout the body, would also increase in weight 700 times. Hearts cannot handle such a load. Steep turns sometimes cause pilots to pass out temporarily because blood does not reach the brain, although in this case the acceleration a is no greater than 5–10 g. The load on the heart could be reduced by assuming a supine position, thereby reducing the height to which the blood must be raised. But even an acceleration of 10 g in this position can be endured for only three to five minutes.

Can some alternative method of traveling at such large accelerations be devised? Can a spacecraft be designed in such a way as to eliminate such large increases in weight?

Before attacking the enemy, one should get to know him. Where does the weight come from?

Standing on a spring scale causes the arrow to be deflected from its zero position. The Earth attracts the body, but since the body is stationary relative to the scale, this force must be counterbalanced by the spring within the scale. The spring therefore shows the weight $W = mg$ (where m is the mass of the body).

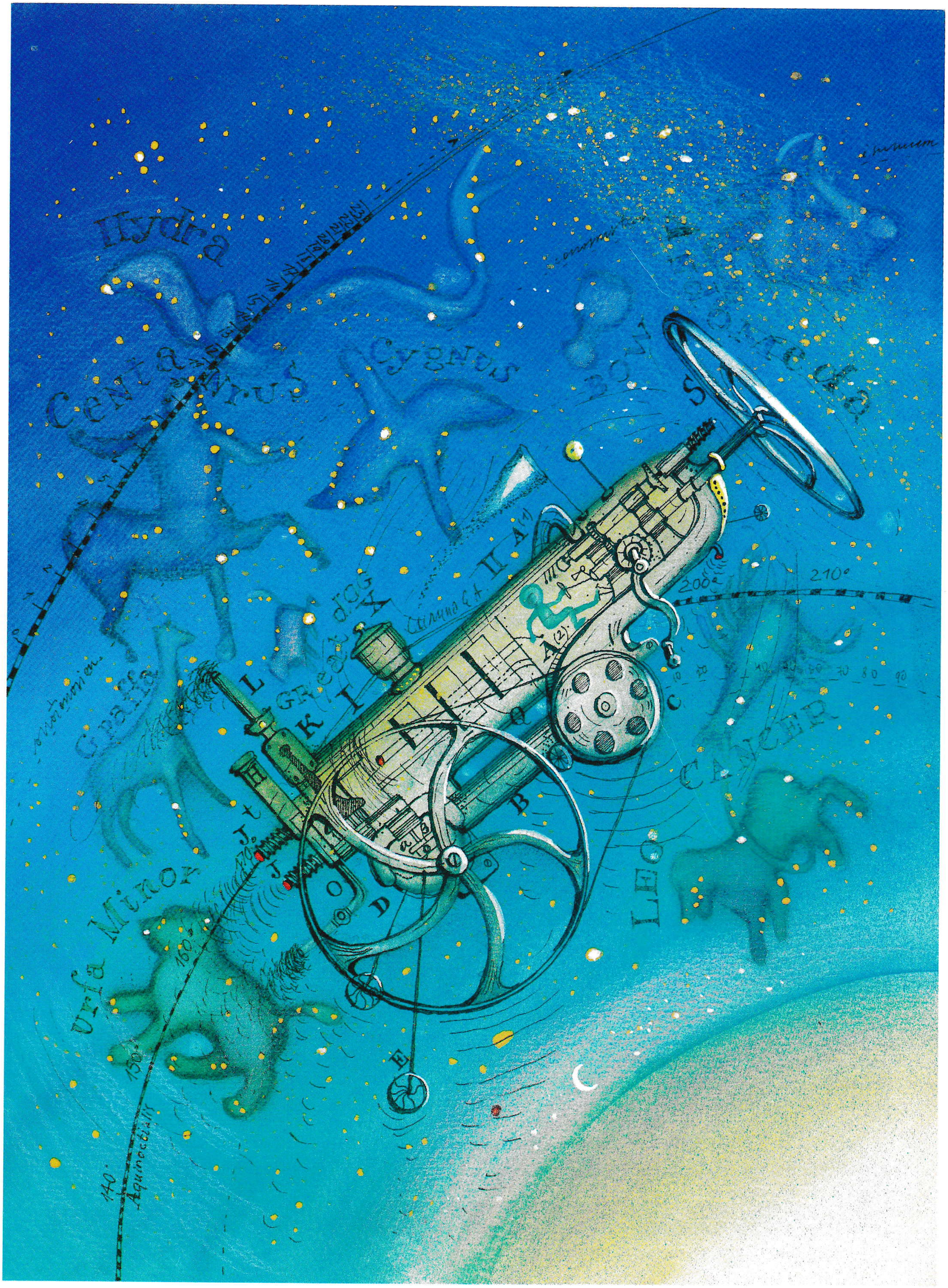
Let's assume that we are measuring the weight (the reading on the scale) in an elevator that is rising with acceleration a. The scale now reads $m(a + g)$, which indicates an increase in weight. In this case the force of the spring not only counterbalances the gravitational attraction of the Earth but also accelerates the body.

If, on the other hand, the elevator is in free fall, the scale will read zero, because all bodies fall with the same acceleration. The body's displacement therefore is exactly the same as that of the scale; the spring remains relaxed. Holding an apple in the hand in an elevator in free fall, one would not feel its weight because the apple falls freely with the hand without exerting any pressure on it.

This simple experiment shows that weight is $W = m(a + g)$ if the acceleration a is directed upward. If the acceleration a is in the same direction as the acceleration due to gravity g, and if $|a| = |g|$ (free fall), we have a state of weightlessness. This is the state experienced by astronauts when they are orbiting the Earth with the engines turned off.

It is important that weight is determined both by acceleration and

Art by Ekaterina Silina


Hydra
Cygnus
Great dog
CANCER
LEO
210°
200°
Aquinoctialis

by gravitational attraction. One component can now be canceled by the other. The state of weightlessness can be attained because the acceleration due to gravity does not depend on mass or the composition of the falling bodies. This fundamental property was discovered by Galileo. Even today, however, some researchers occasionally cannot resist the temptation to check this property with greater accuracy. In 1964, John Roll, Robert Krotkov, and Robert Dickey measured the relative gravitational acceleration of gold and aluminum to an accuracy of 0.00000000001!

Let's now consider a massive body (the spacecraft) that produces a gravitational acceleration a equal to the acceleration desired for traveling. Let's then place engines on this massive body that are capable of attaining this acceleration. The traveler is comfortably situated in a private cabin, which falls freely toward this body. The cabin falls with acceleration a, but since the engines are running, the body is accelerating at the same rate. If the initial velocity is zero, the distance between the passenger's cabin and the spacecraft remains the same. The passenger experiences weightlessness in the free-falling cabin, which is moving with the appropriate acceleration at the same time.

At the halfway point of the journey, however, deceleration of the spacecraft must be initiated in order to slow the spacecraft down. However, we must maintain the match between the acceleration of the spacecraft and the free-fall acceleration of the cabin, thus preventing the cabin from ramming the spacecraft. In addition, there is the danger of lagging behind or colliding with the spacecraft as a result of a change in its acceleration. The spacecraft, therefore, needs to be better designed.

Let's drill a shaft through the center of the massive body, which has the shape of a sphere, in the direction of acceleration. The gravitational acceleration changes as a function of the distance from the center of the sphere. Denote the acceleration due to gravity at the sphere's surface by a_0 and the radius of the sphere by r_0. Figure 1 shows how the acceleration a varies with distance r from the sphere's center.

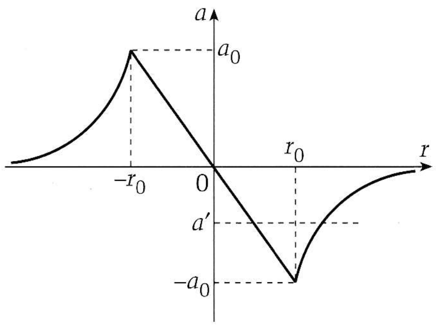


Figure 1

At the distances between $-r_0$ to r_0 (that is, inside the sphere) acceleration due to gravity varies according to the law

$$a = a_0 \frac{r}{r_0},$$

while at $r > |r_0|$ (outside the ball) the gravitational acceleration is described by the equation

$$a = -a_0 \left(\frac{r}{r_0}\right)^2.$$

The negative sign means that the acceleration is always directed toward the center of the sphere.

When the sphere is moving uniformly (without acceleration), its center is the point of stable equilibrium. Small deviations of the cabin from the sphere's (spacecraft's) center are corrected by the restoring force. If the acceleration a' of the spacecraft is slower than a_0, the cabin will always find a position at which the gravitational acceleration is a'. This position can be found from the plot by drawing a straight line parallel to the r axis at a distance a' below it. The ordinates of the two points of intersection give the positions. However, the point in the shaft corresponds to the stable equilibrium point. If the change in acceleration of the spacecraft is small, the cabin will oscillate near a new equilibrium position, monitoring the agreement between the two accelerations. The cabin remains in the state of weightlessness all this time.

If the sphere decelerates at the same rate, the equilibrium position will be on the other side of its center.

We have proposed a radical method to combat the effects of very large accelerations. What hurdles must be surmounted to realize this fantastic project? It does not violate the laws of nature. The main problem is the energy source. But this is the Achilles' heel common to all interstellar space projects. How can such a massive body be constructed? One can use either a body of very large volume or a material of very high density.

Quantum on gravitation, space voyages, and weightlessness:

A. Stasenko, "From the Edge of the Universe to Tartarus," March/April, 1996, pp. 4–8.

A. Byalko, "A Flight to the Sun," November/December 1996, pp. 16–20.

V. Surdin, "Swinging from Star to Star," March/April 1997, pp. 4–8.

V. Mozhaev, "In the Planetary Net," January/February 1998, pp. 4–8.

S. Pikin, "Weightlessness in a Car?", July/August 1998, p. 31.

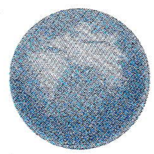
CONTINUED FROM PAGE 23

P295

The plate thickens. The luminous flux from a point source is measured with the help of a small photosensitive detector located at a distance $L = 0.1$ m from the source. A plane-parallel glass plate is placed between the source and the detector in such a way that its plane is perpendicular to the line connecting the light source and the detector. The refractive index of glass is $n = 1.5$. For what thickness of the plate will the reading of the detector not change? Glass is transparent. The coefficient of reflection k at the glass-air boundary for a normal angle of incidence is given by $k = (n-1)^2/(n+1)^2$. (S. Varlamov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 49

Do you know the bi



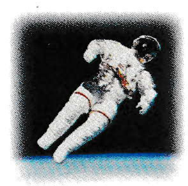
IF THERE WAS A COMPETITION between physical concepts for the top prize, then one of the strong candidates would surely be the binding energy. We usually encounter it in the final basic physics course involving the study of forces that bind nuclear particles. Let's, however, take a broader view of the binding energy and consider it as the work needed to pull apart two bodies attracting each other to a distance at which they no longer attract each other. We shall then see that in many cases the binding energy is indispensable in understanding the principles of the construction of the universe.

Is it not the binding energy that is "responsible" for the stability of planetary systems, molecules, atoms, and their nuclei? Carefully examining such seemingly dissimilar phenomena and processes as melting and evaporation, ionization and the photoelectric effect, the flight of a spacecraft and radioactive decay, we note that the concept of binding energy helps us to discover many common features in a heterogeneous world. In other words, binding energy is one of the wonderfully universal notions that unifies various types of physical interactions.

We hope that by reading this *Kaleidoscope* you will see the physical world in a less pessimistic light than suggested by the verses of the famous XVII century English poet John Donne. Had he been acquainted with the binding energy, he would certainly have been more optimistic.

"Into elements the universe is split. The bonds are torn, and the world hath gone to pieces."
John Donne

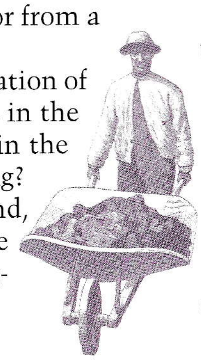
Problems and questions



1. An astronaut is on board a spaceship orbiting the Earth. Does the state of weightlessness that he experiences attest to the loss of connection with the Earth?

2. The kinetic energy of a satellite in a circular orbit is positive. What is the sign of its total mechanical energy?

3. Does a rocket require more energy to escape the planet's gravitation when it is launched from the surface of the planet or from a circular orbit?



4. Why does evaporation of a liquid in a jar result in the cooling of the liquid in the absence of any heating?

5. Why can wet sand, but not dry sand, be used to sculpture a figure?

6. The dissociation of molecules that takes place during the dissolving of salt crystals in water leads to a growth of the potential energy of interaction between the ions. To what is this process attributed?

7. What can cause a drastic increase in the number of electron-hole pairs in semiconductors?

8. Two uncharged plates of dissimilar metals have identical concentrations of free electrons. Which plate will become negative when the two plates are brought in contact with each other?

9. Are thermionic emission and evaporation of a liquid similar processes?

10. How can the saturation current in a vacuum diode be changed?

11. Why is a relatively small voltage sufficient to maintain the glow of an electric arc?

12. Why do electrons rather than heavy ions (although they are also accelerated by the electric field) play the major role in collision ionization that leads to self-maintained discharge in gases?

13. Can a hydrogen atom absorb a photon whose energy is larger than its binding energy?

14. Is more energy required to free the first or the second electron from a helium atom?

15. Can a free proton capture an electron (and produce a hydrogen atom) without any radiation?

16. In what part of the atom—its nucleus or its electron shell—do the processes that lead to the emission of β-rays occur?

binding energy?

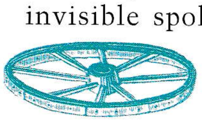
17. The light emitted by the surface of a star has a higher frequency than that of the light viewed by an observer. What is the reason for this difference in frequencies?

Microscopic experiment

Drop some vegetable oil into water in a large pan. What shape will the oil drops take? What are the forces that bind them and prevent them from spreading uniformly on the surface of the water?

It is interesting that...

...the opponents of the theory of Copernicus believed that the Earth was too heavy, too inert, and too sluggish to be able to rotate about its axis. Otherwise, they believed, the planet would break up into pieces, like a flywheel spun out of control. Later Kepler had to contrive special invisible spokes that bound the planets to the Sun and moved them in their orbits.

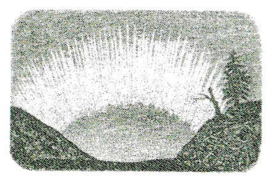


...the energy needed to launch an object with mass of one kilogram outside the range of the Earth's gravitation can be obtained by burning one and a half liters of gasoline (without allowance for the losses).

...the stability of most of the objects that surround us is determined by the fact that the energy of thermal motion of their molecules is not sufficient to break the chemical bonds that hold the molecules together.

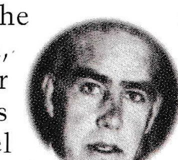
...in the 1920s quantum mechanics was used to explain the nature of chemical bonds. Laborious calculations of many years, with the help of quantum mechanics, have finally yielded complete agreement between experimental data and theory. This was the birth of quantum chemistry, which today uses powerful computers to perform the calculations.

...analysis of the radiation of the northern lights led to the conclusion that in the high atmospheric layers the oxygen molecules are dissociated by the solar ultraviolet radiation into atoms, which then scintillate individually.



...when the temperature is higher than five or six thousand degrees, thermal ionization takes place in gases. The electrons are ejected from the atoms and the substance undergoes a transition to the plasma state. Plasma radiation can be used to determine the nature of stars, of the ionosphere, and of a gaseous discharge; it may also provide a key to solving the nature of ball lightning.

...Niels Bohr, the author of the famous model of the structure of the atom, which was named after him, published one of his papers on this model with the title "Binding of an electron by the positive nucleus."



...the extreme chemical inertness of rare (noble) gases was explained in the study of outer electron shells of their atoms. When these shells are completely filled, the binding of the electron with the nucleus is strongest. In helium the energy of this bond is the strongest among all atoms.

... A century ago, a young physicist named Ernest Rutherford was able to explain the phenomenon of the ionization of gases by recently discovered radioactive substances. In the experiments he used an electroscope that rapidly discharged in the ionized air. This important device was made from a silk brush. It was electrified by stroking its base with a "warm, dry tobacco pouch." Evaluate the level of experimental techniques used a hundred years ago!

...the inevitable failures of alchemists to convert one chemical element into another, that is, to transform the nuclei, is explained by the fact that the binding energy in the nuclei per particle is about one million (!) times greater than the energy of chemical bonds between the atoms.

...The atomic nuclei that contain a so-called magic number of protons and neutrons have larger binding energies and therefore a greater resistance to decay. The search for such nuclei, which are islands of stability outside the range of Mendeleev's table, has recently been rewarded: the 114th element was synthesized in a laboratory at Dubna.

...Quarks, which are the smallest structures in the particles inside the nucleus, do not exist in a free state, although experiments have firmly convinced researchers of their existence. The forces "gluing" the quarks together are so unusual that inability of a quark to escape was given the name "confinement." ◘

ANSWERS, HINTS & SOLUTIONS ON PAGE 53

PHYSICS CONTEST

Rolling wheels

by Arthur Eisenkraft and Larry D. Kirkpatrick

The fly sat upon the axel tree of the chariot-wheel and said, What a dust do I raise!
—Aesop (6th century B.C.)

INVENTING THE BETTER downhill vehicle confounds engineers. If the measure of success is the speed of the car or skateboard at the bottom of an incline, how does one choose the proper wheels? Solid wheels, spoked wheels, cylinders, rotating casters, spheres, hoops, and a combination of these shapes inhabit our palette of possibilities.

The decision making certainly requires an understanding of rolling bodies. Although most people appear quite familiar with wheels, a few simple puzzles reveal the subtlety of the motion. Ask someone to trace the motion of a point on the outside of a bicycle wheel as it moves across your line of sight. Compare this with the actual trace as you roll a disk and construct the path. The path is a cycloid—not the first guess of most people!

A simpler puzzle requires you to place one quarter next to another with the faces of Washington upright. As the quarter on the right rolls without slipping on the other, what will be the orientation of Washington when the coin reaches the left-hand side of the fixed quarter? Try it and then explain it.

Finally, take a ruler and rest it on the wheels of a cart. As the cart moves forward, will the ruler move with the cart? If not, why not?

When viewing a rolling (without slipping) wheel, we notice that the bottom of the wheel is at rest with respect to the ground. There is no skidding here. The speed of the center of the wheel is equal to the speed of the wheel. When the wheel moves one meter to the right, the center of the wheel moves one meter to the right. The top of the wheel is also moving to the right, but at what speed?

An analysis of the kinetic energy of the wheel will help reveal the motion of the wheel. If the wheel is a solid cylinder, the kinetic energy of this fixed rotating cylinder is equal to the kinetic energy of each of the small elements

$$K = \sum \frac{1}{2} m v^2 .$$

Since each mass has the same angular velocity ω, the velocity of each small mass is equal to $r\omega$, where r is the distance from the center of rotation.

$$K = \sum \frac{1}{2} m r^2 \omega^2$$
$$= \frac{1}{2}\left(\sum m r^2\right)\omega^2 = \frac{1}{2} I \omega^2,$$

where I is defined as the rotational inertia of the rotating object. The rotational inertia is a measure of the difficulty of starting the rotation of an object in the same way mass is indicative of the difficulty of accelerating an object.

For a rolling wheel, the instantaneous rotation is about the point in contact with the ground. The definition of kinetic energy is unchanged but the moment of inertia about this new point must be determined. The parallel axis theorem can be used to determine this new rotational inertia

$$I = I_{cm} + md^2,$$

where d is the distance of the rotating axis from the center of mass. In the case of a rolling body, the instantaneous point of rotation is the ground, which is one radius from the center of mass of the wheel. The kinetic energy of this rotating wheel is

$$K = \frac{1}{2} I \omega^2 = \frac{1}{2}\left(I_{cm} + mR^2\right)\omega^2$$
$$= \frac{1}{2} I_{cm}\omega^2 + \frac{1}{2} mR^2\omega^2$$
$$= \frac{1}{2} I_{cm}\omega^2 + \frac{1}{2} m v_{cm}^2 .$$

Our analysis of the rolling wheel has become much simpler. The kinetic energy of the rolling wheel is equal to the rotational kinetic energy of the wheel about its center of mass plus the kinetic energy of the entire mass moving with the velocity of the center of mass.

Art by Tomas Bunk


WALL STREET
BLA BLA BLA
UNITED STATES OF AMERICA
QUARTER DOLLAR
TOM BUNK

That being the case, we can see that the center of the wheel moves with v_{cm}, and the wheel spins with an instantaneous angular velocity ω about the point in contact with the ground. All points have a tangential speed equal to ωr. For the point on the ground, the velocity is zero, for the central point the velocity is ωR, and for the top point the velocity is $2\omega R$. The top of the wheel moves at twice the speed of the center of mass.

We can use conservation of energy to find the speed of a wheel as it rolls down an incline. The loss in potential energy is equal to the gain in both translational and rotational kinetic energy.

$$mgh = \frac{1}{2}mv_{cm}^2 + \frac{1}{2}I\omega^2.$$

Different wheels will have different speeds owing to their different rotational inertias.

To fully understand the rolling wheel, we must delve into the dynamics of the motion. Why would the wheel rotate—why would anything rotate? If I were to place a pen on the table, how might you apply a force to rotate the pen? A force at the center of mass will accelerate the pen, but a force at any other point is required to rotate the pen. This "off-center" force is called a torque and is defined as the cross product of the distance and the applied force.

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}.$$

The force of gravity acts on the center of mass and cannot be responsible for the rotation. Similarly, the normal force of the surface on the wheel has no r (off-center component). It is the frictional force at the bottom of the wheel that creates the rolling.

Designing a wheel requires other considerations than the analysis above. The engineer must take into account the strength of materials needed, the ease of manufacturing, the cost of materials and production, the need for maintenance and repair, the durability, and the aesthetics. Maximizing the speed of a wheel and ensuring safety, low cost, and durability often requires compromise and creativity.

Part of this month's Quantum problem was originally given in Bucharest, Romania, in 1972.

A. Show that all solid spheres will arrive at the bottom of the incline with the same speed, independent of their radii. The rotational inertia of a solid sphere is $2/5\ mR^2$.

B. Determine the relative speeds of a cylinder, a hoop, and a solid sphere (all of the same mass) at the bottom of an incline. The rotational inertias are $1/2\ mR^2$, mR^2, and $2/5\ mR^2$.

C. Consider three cylinders of the same length, outer radius, and mass. The first cylinder is solid. The second is a hollow tube with walls of finite thickness. The third is a hollow tube with walls of the same finite thickness, filled with a liquid of the same density (both ends are closed by thin plates of negligible mass). Find and compare the linear and angular accelerations for the cylinders when they are placed on an inclined plane of angle α. The coefficient of friction between the cylinders and the inclined plane is μ. Friction between the liquid and the wall of the cylinder is negligible.

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.

A question of complexity

Our problem in the November/December 1999 issue of *Quantum* was taken from last summer's International Physics Olympiad that was held in Padua, Italy. A vertical cylinder filled with gas and capped by a moveable glass plate is illuminated for a finite time by a laser. As the gas absorbs the light, the glass plate is observed to move upward.

A. In order to find the final temperature and pressure of the gas, let's begin by noting that the initial temperature is the same as room temperature, which is given to be 20.0°C. The initial and final pressures are the same. The difference in the pressures inside and outside the cylinder must be large enough to support the weight of the glass plate. Therefore,

$$P_f = P_i = P_0 + \frac{mg}{\pi r^2},$$

where P_0 = 101.3 kPa is the atmospheric pressure. With a radius r = 50 mm and a mass m = 800 g, we find that P_f = 102.3 kPa.

Now let's find the initial volume of the gas. According to the ideal gas law, we have

$$V_i = \frac{nRT_0}{P_i}.$$

Since we are only given the displacement Δs of the glass plate, we will need to find the initial height of the glass plate in order to find the final volume.

$$h_i = \frac{V_i}{\pi r^2} = \frac{nRT_0}{P_0\pi r^2 + mg}.$$

Therefore,

$$V_f = V_i\left(\frac{h_i + \Delta s}{h_i}\right).$$

Using the ideal gas law, we have

$$T_f = T_0\left(\frac{V_f}{V_i}\right) = T_0\left(1 + \frac{\Delta s}{h_i}\right)$$
$$= T_0 + \frac{\Delta s\left(P_0\pi r^2 + mg\right)}{nR}.$$

Plugging in Δs = 30.0 mm, we get T_f = 322 K = 49°C.

B. The mechanical work performed is given by the force exerted by the gas on the glass plate multiplied by the displacement

$$W = (P_0\pi r^2 + mg)\Delta s,$$

which gives a value of 24.3 J.

C. The internal energy of the gas increases by

$$\Delta U = nc_V\Delta T,$$

and using the first law of thermodynamics, the heat absorbed must be

$$\begin{aligned} Q &= \Delta U + W \\ &= nc_V \frac{T_0 \Delta s}{h_i} + \left(P_0 \pi r^2 + mg\right)\Delta s \\ &= \Delta s\left(P_0 \pi r^2 + mg\right)\left(\frac{c_V}{R} + 1\right). \end{aligned}$$

This gives a numerical value of 85.3 J.

D. Given that the laser is on for only 10.0 s, the power of the laser is $P = Q/\Delta t = 8.53$ W. The wavelength λ of the laser light is 514 nm, so the energy of each photon $E = hc/\lambda$, where $h = 6.63 \cdot 10^{-34}$ J·s is Planck's constant. Therefore, the number of photons emitted per unit time is

$$\frac{P\lambda}{hc} = 2.2 \times 10^{19}\ \text{s}^{-1}.$$

E. We can now calculate the efficiency of this "heat" engine for converting light energy into gravitational potential energy to be

$$\eta = \frac{mg\Delta s}{Q} = 0.28\%.$$

F. When we rotate the cylinder so that its axis is horizontal, we have an adiabatic change in pressure from P_f to P_0. We know that PV^γ is constant for an adiabatic expansion, where

$$\gamma = \frac{c_P}{c_V} = 1 + \frac{R}{c_V}.$$

This gives us

$$\frac{V_{rot}}{V_f} = \left(\frac{P_f}{P_{rot}}\right)^{1/\gamma}.$$

Because both the volume and the pressure change, we find the temperature after rotation by

$$\begin{aligned} T_{rot} &= T_f\left(\frac{P_{rot}}{P_f}\right)\left(\frac{V_{rot}}{V_f}\right) \\ &= T_f\left(\frac{P_{rot}}{P_f}\right)\left(\frac{P_f}{P_{rot}}\right)^{1/\gamma} = T_f\left(\frac{P_{rot}}{P_f}\right)^{(\gamma-1)/\gamma}. \end{aligned}$$

This gives a temperature of 321 K, a drop of only one kelvin.

The Death of a Star

(Part 2)

by David Arns

Now, I hope that you remember what I talked about before,
When I spoke at length of dying stars and such.
Well, here I am again; I'm back to give you an encore,
'Bout what happens to a star
When its behavior gets bizarre
As it dies, because it weighs so very much.

I'm talking here about those supermassive types of stars:
A dozen solar masses at the least.
What happens to them as they say their final au revoirs?
Well, they often do go nova,
But the story's not yet "ova,"
There is much to do before they are deceased.

Now these novas are exploding stars; this happens at the end,
When a massive star exhausts it nuclear fuel.
It collapses, making temperatures too high to comprehend,
Then the shell becomes so dense
It stops neutrinos cold, and hence,
It just explodes, and then its "ashes" spread and cool.

But what happens to the *core* that such a nova'd leave behind,
Having thrown most of its mass out into space?
Well, assuming that its mass is large enough, then you will find
That there is truth more strange than fiction
Hid in physics' jurisdiction:
The core would vanish, leaving not a trace!

Okay, that part I said about its "leaving not a trace"
Is not *entirely* true; I'll tell you why.
Although by merely looking, you'd see nothing in its place,
If you measured *gravitation*,
You would see a demonstration
Where you'd swear that something's badly gone awry.

There'd be a gravitational well that simply would not quit,
Consuming everything that happened by;
And nothing could come out that went into this bottomless pit—
Even photons can't get out,
But in steep'ning downward route,
They would vanish with a tiny, tortured cry.

See, a "black hole," as they call these things, is just exactly that:
Its escape velocity is more than *c*.
And, of course, this means that *nothing* can go past the line whereat
Even light is bound securely;
And if *light* is bound, then surely,
Nothing else could hope to e'er again be free.

CONTINUED ON PAGE 35

Sink or swim

by N. Rodina

THE SIGHT OF A GOOSE waddling awkwardly on land gives the impression that its weight is a burden to carry around. But once in water the goose moves quickly and freely: even a light puff of wind can change its speed. How does such a dramatic change occur?

To see this effect more clearly, set a cork upright on a table and blow on it lightly from one side—it will not move. Once placed in water, however, a breath of air moves the cork easily. Evidently, the force of friction between a solid body and water is much less than that between two solid bodies.

The biggest mammal on Earth, the blue whale, is perfectly designed for an aquatic life. It may weigh up to 130 metric tons, but in water it can reach speeds up to 20 knots, or 37 km/h. In comparison, motor boats cruise at up to 30 km/h, or about 16 knots. Sixty-ton sperm whales have been seen leaping several meters out of the water. How do such behemoths move with such grace and ease?

"The whale... doesn't merely inspire superlatives—it is a living superlative," writes Jacques Cousteau in *Whales, the Lord of Seas.* The length of a blue whale may be as much as 33 m, almost 10 m longer than a railway car. The largest caught whale weighed 150 metric tons, while the largest terrestrial animal, the elephant, weighs in at only 3 to 6 metric tons (merely the tongue of some whales!).

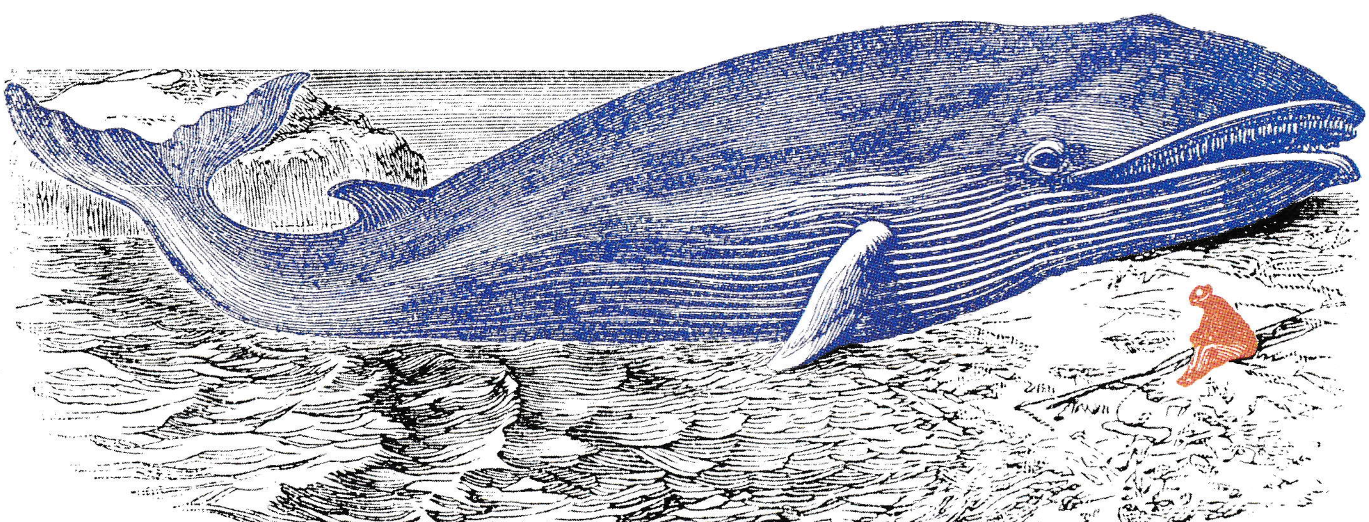
If an elephant's mass were doubled, it would need legs two times as thick to support itself. The cross-sectional area of each leg is 4 dm^2. Can you explain why terrestrial giants need such thick legs?

A body is neutrally buoyant (neither floats nor sinks) if the forces of buoyancy and gravity acting upon it are equal. Let's evaluate and compare these two forces. The force of buoyancy equals the weight of liquid displaced by a body, or

$$F_B = g\rho_l V,$$

where g is about 10 N/kg, ρ_l is the density of the liquid, and V is the volume of the body. How can we determine the volume of a whale? If we use a cylinder to approximate the shape of a whale, its volume is $V = \pi d^2 h/4$, where d is the diameter of the cylinder and h its height, or in our case, the length of the whale. Let's assume the diameter of our whale-cylinder is the average diameter of the whale's body, which is about one tenth of its length.

Make the calculations yourself and you'll discover that the force of buoyancy that keeps the whale neutrally buoyant amounts to millions of newtons. (Of course these are rough figures, but the force is somewhere between one and ten million newtons.) Such a huge force easily



supports a body weighing a hundred metric tons or so. So we see that in water a whale is actually weightless, because the force of gravity acting upon it is counterbalanced by buoyancy. No wonder whales seem to swim effortlessly.

On land, however, these giants face unsurmountable problems. Whale strandings are well known, if not entirely understood, phenomena. Out of water, the skeleton of a whale cannot bear the weight of the muscles and blubber that serve the whale perfectly in the dense medium of water. Just breathing requires enormous effort on land.

Once during an expedition Cousteau and his friends tried to save a stranded calf that weighed "only" two metric tons. To lift it onboard their ship they had to use a special hammock, because even a newborn whale can "break" under gravity if the underlying support is uneven.

If you were to watch a whale sleeping, you'd notice that it sticks partly out of the water. This means that its buoyancy is less than if it were completely submerged (because it equals the weight of the liquid displaced by the whale). However, the force of gravity (weight) remains the same. Has the equilibrium of neutral buoyancy been broken? Not at all: the whale sleeps serenely and doesn't sink. Therefore, the force of buoyancy equals that of gravity, just as before. How can this apparent contradiction be explained?

This is a good time to explain how a whale dives and resurfaces. The horizontal blades of a whale's tail help generate about 500 horsepower (1 horsepower is a unit of power equal to 736 watts). So to an underwater swimmer, a brush with a whale more closely resembles an encounter with an oncoming truck than a friendly puppy.

By a mighty movement of its tail a whale dives to the depths of the sea. Whales regularly dive to depths of dozens of meters, and sperm whales can reach depths of 1,000–1,200 meters. At such depths the water pressure is very high (calculate it yourself using the fact that the density of seawater is 1,030 kg/m^3). Under this pressure the lungs of a whale shrink to a residual volume. As its lungs shrink the whale's total volume decreases, and therefore its buoyancy drops.

As a whale rises, its buoyancy gradually increases (Why?). At the surface the whale takes a deep breath, further increasing its volume. Whether a whale swims at the surface or down below, it's supported by the same force of buoyancy needed to counterbalance its constant weight. However, to generate this force at the surface the whale need not be completely submerged to attain neutral buoyancy.

Here are a few more questions for you to ponder:

If we divide a whale's mass by its volume, we get its average density. Can you confirm that wherever a whale swims—in the depths of the ocean, at middle depths, or on the surface—its density is always equal to that of water? What mechanism is used to change the average density of a whale?

Sometimes whales visit brackish, coastal lagoons. How would a whale have to adjust its buoyancy to the different water composition in these areas?

Hot-air balloons are sometimes used to observe and photograph whales in shallow lagoons. Explain how such balloons ascend.

The flame of one hot-air balloon was adjusted so that the balloon was perfectly balanced: In windless weather it could hang at a set height above the water as long as necessary. What can be said about the relationship between the mass of air displaced by the balloon and the mass of the balloon?

Consider a phenomenon observed by Cousteau, in which he saw seawater bubble as if it were champagne. "It was a school of small fry going down, then up to the surface, forcing bursts of air from their swim bladders." Why did the fish force air from their swim bladders, and when did they do it—when diving or when coming up to the surface? ◘

Quantum on buoyancy:

V. Nevgod, "The Adventures of Hans Pfaal and Fatty Pyecraft," January 1990, pp. 14–15.

A. Buzdin and S. Krotov, "Boy-Oh-Buoyancy!," September/October 1990, pp. 27–31.

A. Eisenkraft and L. D. Kirkpatrick, "The Tip of the Iceberg," September/October 1992, pp. 24–26; March/April 1993, p. 41.

A. Eisenkraft and L. D. Kirkpatrick, "Up, Up and Away," September/October 1998, pp. 34–36; March/April 1999, p. 32.

CONTINUED FROM PAGE 33

This line beneath which light is trapped is the "event horizon"
Because whatever happens underneath
Can never be perceived, because (and this is not surprisin')
Since no signals can get out,
You can see, without a doubt,
The event horizon's just a one-way sheath.

So what happens to the star-stuff that's inside this sphere of black?
It shrinks and shrinks, 'til it's completely gone.
But how can a singularity—this spatial cul-de-sac-
Exhibit gravity like that?
Such a cosmic Cheshire cat
Goes away, but leaves a "smile" of gravitons.

Well, we're not exactly sure what astrophysics are required
To give answer to these questions; this we know.
And indeed, the more we learn, we find the more we get inspired
To learn more, for every answer, while an intellect enhancer,
Helps us see how terribly far we've got to go.

AT THE BLACKBOARD I

The quadratic trinomial

by A. Bolibruch, V. Uroev, and M. Shabunin

THE EXPRESSION

$$ax^2 + bx + c,$$

where a, b, and c are given numbers and $a \neq 0$, is called a quadratic trinomial in x. The values of x for which the quadratic trinomial vanishes (becomes zero) are called its *roots.*

Problems that require a knowledge of the properties of quadratic trinomials are often encountered in examinations. Many students can easily write various formulas, plot the function $y = ax^2 + bx + c$, and are familiar with its basic properties. This knowledge, however, is often superficial, and students don't know how to use it to solve problems.

In this article we show through examples the importance of combining algebraic and geometric reasoning to solve problems that involve quadratic trinomials.

1. *Find the maximum value of the quadratic trinomial* $y = -2x^2 + 4x - 5$.

A person familiar with the calculus can use the derivative to solve this problem. However, we can easily do without the calculus. Let us complete the square:

$$\begin{aligned} y &= -2x^2 + 4x - 5 \\ &= -2(x^2 - 2x + 1) + 2 - 5 \\ &= -2(x - 1)^2 - 3. \end{aligned}$$

We see that the maximum value of this quadratic trinomial is –3, and it is attained for $x = 1$.

The method of completing the square is used to derive the formula for the roots of a quadratic equation. It can also be used to plot the graph of the general quadratic function $y = ax^2 + bx + c$. Indeed, if we complete the square, we have

$$y = \left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{2a}.$$

Thus, we see that the graph of the quadratic function is obtained by translating the parabola $y = ax^2$ by the vector

$$\left(-\frac{b}{2a}, \frac{4ac - b^2}{2a}\right).$$

2. *Figure 1 shows four parabolas. Each can be described by an equation of the form* $y = ax^2 + bx + c$. *In each case, determine the sign of the numbers* a, b, *and* c.

We consider case (i) in detail. The coefficient a is less than 0, since the branches of the parabola are directed downwards. The abscissa of parabola's vertex is $-b/2a$. Since this is negative, we know that $b < 0$. The ordinate of the point where the parabola intersects the y-axis is equal to the value of $f(x) = ax^2 + bx + c$ for $x = 0$. Therefore, $c = f(0)$ is positive. Thus, we have $a < 0$, $b < 0$, and $c > 0$.

The same result can be obtained by considering the sum and the product of the roots of the equation $ax^2 + bx + c = 0$. However, this

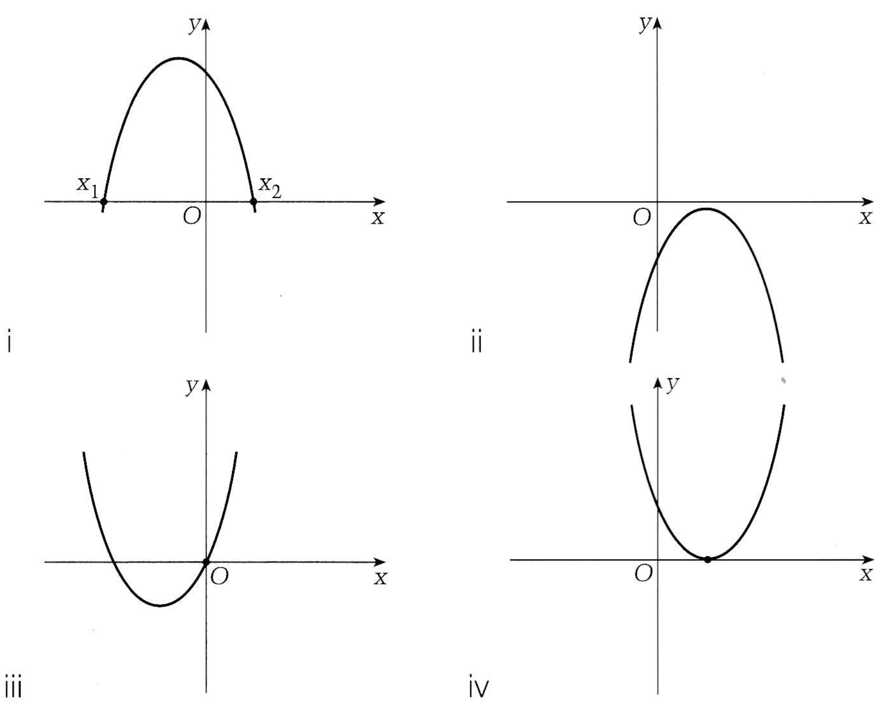


Figure 1

method cannot be used in case (ii), when the roots are complex.

We invite the reader to analyze cases (ii), (iii), and (iv).

3. *Suppose that the roots x_1 and x_2 of the quadratic equation $x^2 - 2rx - 7r^2 = 0$ satisfy the condition $x_1^2 + x_2^2 = 18$. Find the value of r.*

First, we represent $x_1^2 + x_2^2$ in terms of the sum and the product of the roots. We have

$$x_1^2 + x_2^2 = (x_1 + x_2)^2 - 2x_1x_2 = (2r)^2 - 2 \cdot (-7r^2) = 18r^2.$$

Thus, the problem reduces to solving the equation $r^2 = 1$, from which we find $r_1 = 1$ and $r_2 = -1$. We now need to check only that the roots exist for both values of r.

4. *Find necessary and sufficient conditions for the roots x_1 and x_2 of the equation $f(x) = x^2 + px + q = 0$ to have different signs and be greater than 1 in absolute value.*

A solution based on the discriminant and the quadratic formula is rather tedious. The problem, however, can be easily solved by using geometrical considerations.

First, we find a necessary condition. Let $x_1 < x_2$ (here the roots are different). We are given that $x_1 < -1$ and $x_2 > 1$; i.e., the interval $[-1, 1]$ belongs to the interval $[x_1, x_2]$ (see figure 2). This condition is equivalent to the following system of inequalities:

$$\begin{cases} f(-1) < 0, \\ f(1) < 0. \end{cases} \quad (1)$$

Substituting 1 and –1 into the expression for $f(x)$, we obtain the necessary conditions

$$\begin{cases} -p + q < -1, \\ p + q < -1. \end{cases} \quad (2)$$

The relationship between p and q can be illustrated graphically by a set of points (p, q), whose coordinates satisfy inequalities (2) (see figure 3).

Let us prove that the necessary conditions (2) are also sufficient. That is, if inequalities (2) are satisfied, the roots of the quadratic trinomial $f(x) = ax^2 + px + q$ satisfy the inequalities $x_1 < -1$ and $x_2 > 1$. Since conditions (2) are equivalent to conditions (1), the function $y = f(x)$ takes negative values at two different points ($x = 1$ and $x = -1$). Since the coefficient of x^2 is positive, the branches of the parabola $y = f(x)$ are directed upwards. Thus, the parabola intersects the x axis at two different points x_1 and x_2 ($x_1 < x_2$), and the points –1 and 1 belong to the interval $[x_1, x_2]$. Therefore, $x_1 < -1$ and $x_2 > 1$.

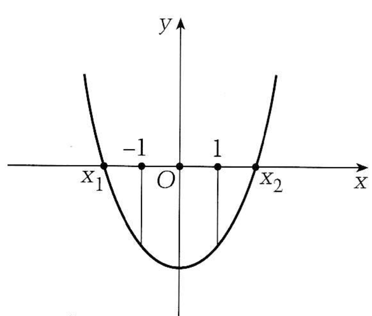


Figure 2

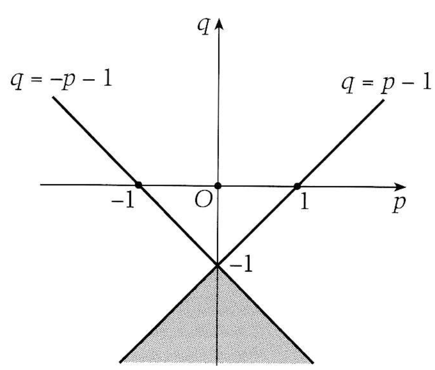


Figure 3

5. *Find all values of r for which the roots of the equation $(r - 4)x^2 - 2(r - 3)x + r = 0$ are greater than –1.*

Consider the case $r = 4$ separately. Then the equation becomes $-2x + 4 = 0$, so $x = 2$. Since $2 > -1$, the value $r = 4$ satisfies the condition of the problem. If $r \neq 4$, we have a quadratic equation.

We will solve a more general problem: *we will find necessary and sufficient conditions for the roots of the quadratic trinomial $f(x) = ax^2 + bx + c$ to be real and greater than a given real number d.*

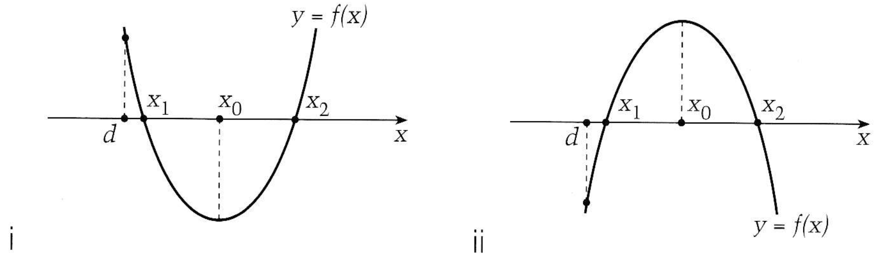


Figure 4

Here is a geometric solution. The roots x_1 and x_2 must exist. Therefore,

$$D = b^2 - 4ac \;^3\; 0. \quad (3)$$

Let us sketch the graph of $y = f(x)$. Figure 4 shows the two possible situations. Since both roots are greater than d, the abscissa of the parabola's vertex is greater than d. That is, $x_0 = (x_1 + x_2)/2 > d$. Using the formulas for the sum and product of the roots, we find:

$$-\frac{b}{2a} > d. \quad (4)$$

The point $x = d$ does not belong to the interval $[x_1, x_2]$. This means that parabola's branches are directed upwards in the case $a > 0$ and $f(d) > 0$ (figure 4i) or downwards in the case $a < 0$ and $f(d) < 0$ (figure 4ii). The numbers a and $f(d)$ are therefore identical in sign. That is,

$$a f(d) > 0. \quad (5)$$

We invite the reader to check that conditions (3)–(5) are not only necessary but also sufficient.

Here is an algebraic solution to the same problem. Two real numbers $x_1 - d$ and $x_2 - d$ are both positive if and only if their sum and product are both positive. Therefore, the condition given in the problem is equivalent to the following three conditions:

$$D = b^2 - 4ac > 0,$$

$$(x_1 - d) + (x_2 - d) > 0,$$

$$(x_1 - d)(x_2 - d) > 0.$$

Using the sum and product of the roots, we can rewrite the second condition as

$$x_1 + x_2 - 2d > 0,$$

$$\frac{x_1 + x_2}{2} > d,$$

$$-\frac{b}{2a} > d.$$

The third condition can be written as

$$x_1x_2 - (x_1 + x_2)d + d^2 > 0,$$

$$a(ad^2 + bd + c) > 0,$$

$$af(d) > 0.$$

Thus, we have again proved that the combination of conditions (1)–(3) is equivalent to the given conditions of the more general problem.

Returning to problem 5, we write conditions (1)–(3) in this case. We have the following system of inequalities:

$$\begin{cases} (r-3)^2 - r(r-4) = 9 - 2r \geq 0, \\ \dfrac{r-3}{r-4} > -1, \\ (r-4)(4r-10) > 0. \end{cases}$$

Solving these inequalities, we obtain $r < 5/2$, $4 < r \leq 9/2$. We also must add $r = 4$ to these solutions.

Answer:

$$\left]-\infty; \frac{5}{2}\right[\cup \left[4, \frac{9}{2}\right].$$

Problems.

6. Let x_1 and x_2 be the roots of the equation $x^2 + px + q = 0$. Find p and q if it is given that $x_1 + 1$ and $x_2 + 1$ are the roots of the equation $x^2 - p^2x + pq = 0$.

7. The graph of the quadratic function $y = ax^2 + bx + c$ cuts off segments AB and CD along two parallel lines. Prove that the line passing through the midpoints of these segments is parallel to the y-axis.

8. If the quadratic trinomial $f(x) = ax^2 + bx + c$ has no real roots, and if its coefficients satisfy the inequality $a - b + c < 0$, find the sign of c.

9. Let the roots x_1 and x_2 of the quadratic trinomial $ax^2 + bx + c$ be different. Prove that the number x_0 lies between x_1 and x_2 if and only if $a(ax_0^2 + bx_0 + c) < 0$.

10. Let the equation $ax^2 + bx + c = 0$ have no nonnegative roots, and suppose $a < 0$. Find the sign of c.

11. Let the coefficients of the equations $x^2 + p_1x + q_1 = 0$ and $x^2 + p_2x + q_2 = 0$ satisfy the equation $p_1p_2 = 2(q_1 + q_2)$. Prove that at least one of these equations has real roots.

12. Is it possible that the equation $x^2 + px + q = 0$, where p and q are rational numbers, has the following roots:

(a) $x_1 = \sqrt{3}$, $x_2 = \dfrac{1}{\sqrt{3}}$?

(b) $x_1 = \sqrt{3} + 2$, $x_2 = 2 - \sqrt{3}$?

13. Prove that any rational root of the equation $x^2 + px + q = 0$ with integer coefficients p and q is an integer.

14. Let the equations $x^2 + p_1x + q_1 = 0$ and $x^2 + p_2x + q_2 = 0$ with integer coefficients p_i and q_i $(i = 1, 2)$ have a common noninteger root. Prove that $p_1 = p_2$ and $q_1 = q_2$.

15. Let x_1 and x_2 be the roots of the quadratic equation $ax^2 + bx + c = 0$ and $S_m = x_1^m + x_2^m$ (where m is a positive integer). Prove the formula $aS_m + bS_{m-1} + cS_{m-2} = 0$. ◘

Circle No. 1 on Reader Service Card

Who needs a lofty tower?

by A. Stasenko

GALILEO, WHO CARRIED out an intriguing experiment, needed one. He took a cannon ball and a musket bullet, and dropped them from a height of 60 meters. The two objects fell to the ground simultaneously, thus shattering Aristotle's theory. According to a legend, Galileo carried out the celebrated experiment from the top of the leaning tower of Pisa. The tower was clearly the best site for this type of experiment.

Towers of course were not built solely for physics experiments. Standing on top of a tower, one could see far into the distance—a very important feature when radios, telephones, and televisions did not exist.

Figure 1 shows that the visible range can be calculated from

$$AB = \sqrt{(R+h)^2 - R^2},$$

where R = 6400 km is the radius of Earth. The Pythagorean theorem is applied to the rectangular triangle OBA. The angle B is a right angle be-

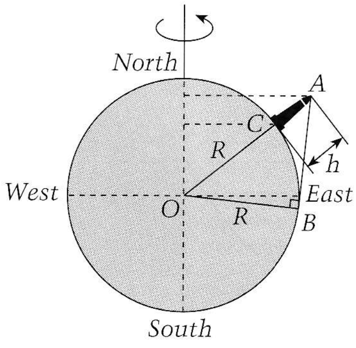


Figure 1

Art by Vasily Vlasov

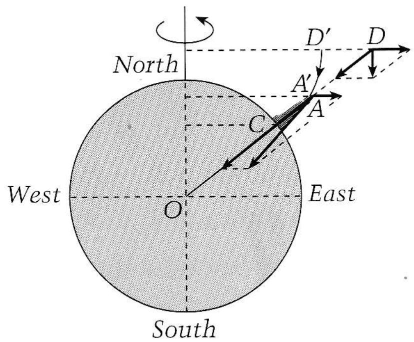


Figure 2

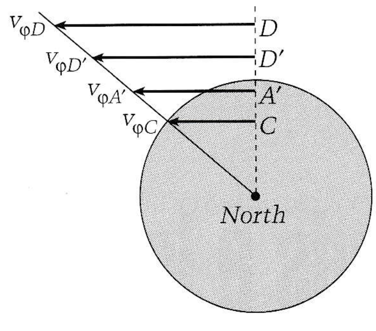


Figure 3

cause the line of sight is tangent to the surface of Earth, which is assumed to be a perfect sphere. Since the height of a tower is much less than the Earth's radius, we can simplify this equation by dropping the small quantity h^2:

$$AB \approx \sqrt{2Rh}.$$

As an example, the visual range of the leaning tower of Pisa (h = 60 m) is approximately 28 km, not bad for that time. (As an exercise, estimate the visible range of the Ostankino television tower in Moscow, whose height is $h \sim 300$ m.)

History and legend say that the first jumps with gliders and parachutes were made from high bell towers and watch towers, sometimes unsuccessfully. Today television antennas are built as high as possible (even on satellites), because unobstructed visibility is important in this range of wavelengths.

Let's consider a tower that is built progressively higher such that its top rises from point A to point D (figure 2). We recall that the Earth rotates around its north-south axis. This means that the builder of this tower (and hence his professionally inseparable plumb bob) are situated on a gigantic merry-go-round.

We know that a person standing on a rotating platform feels a force acting radially away from the axis of rotation. However, an observer on the ground observes the person traveling in a circle and therefore, must experience a centripetal (that is, toward the axis of rotation) force. The force experienced by the person on the rotating platform is an attribute of any rotating system. Rotating systems belong to a large group of non-inertial reference systems. Life in the non-inertial systems is much more enigmatic than in the ordinary (Galilean) inertial reference systems. The appearance of "strange" forces generated "from nothing" is a common feature of such systems. Here we consider some of these forces.

The force generated on a rotating platform is called the centrifugal inertial force. This force is always directed away from the axis of rotation. The centrifugal force increases with distance from the axis. As we climb the tower, this force increases as the force of gravity decreases. While the force of gravity is directed to the Earth's center, the centripetal force is directed to the axis of rotation. Therefore, as the height of the tower increases, the vector sum of these forces deviates progressively more from the direction toward the Earth's center. Since the builder carefully follows the reading of his plumb bob, he constructs not a straight (vertical) tower CAD, which is directed strictly along the radius, but a curved tower $CA'D'$, which is the straightest possible tower from the viewpoint of the art of building towers. Strictly speaking, the walls of very tall buildings are not flat for the same reason.

Towers have a straight vertical axis only at the Earth's poles and at the equator. There is simply no centrifugal force in the first case, while in the second case it is directed strictly along the vertical axis. However, what do we mean by the word "strictly"? If, say, Mt. Kilimanjaro is on one side of the equatorial tower and there is no mountain on the other side, the tower's axis will bend slightly toward the mountain. A highly precise science called gravimetry studies such deviations of the gravitational field. It detects not only the effect of mountains (they can be seen) but also that of dense materials (such as ore-bearing strata, which is of greater practical importance) inside the Earth's crust.

Where, in fact, does the body fall if it is dropped from the top of the tower? Let's look at the falling body from above along the axis of rotation (figure 3). We plot at several points the linear velocities [with the subscript ϕ, which represents the angle of rotation around the axis of rotation (the north-south axis)]. Clearly, the linear velocity increases with increasing distance from the axis of rotation. According to Newton's first law, a falling body conserves the velocity $V_{\phi D'}$ at the point at which it is dropped (if the air resistance is disregarded). Since the velocity at the foot of the tower $V_{\phi C}$ is smaller than $V_{\phi D'}$, during its fall the body travels farther east than the foot of the tower and therefore misses it, falling in the direction of sunrise.

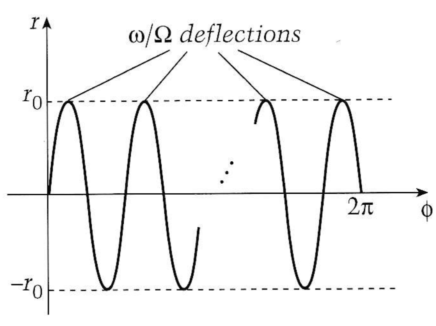


Figure 4

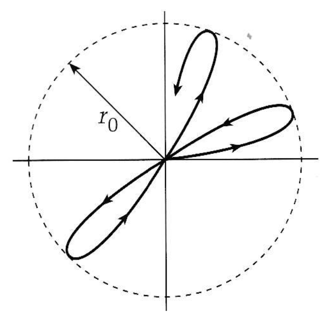


Figure 5

This simple reasoning can help to explain many interesting phenomena. Why, for example, do rivers in the northern hemisphere, which flow in the north-south direction, have steep right banks, while the rivers in the southern hemisphere have steep left banks? With the help of physics, this phenomenon can be explained without visiting these rivers.

Can the falling body perhaps be also deflected to the south by the centrifugal inertial force? This would be the case if the builder constructed a strictly vertical (radial) tower. The builder's plumb bob, however, would take this force into account as a component of the resultant force.

Imagine that the straightest possible tower of height h is built on the Earth's pole so that a pendulum of length h could be attached to its top. The period of the pendulum's oscillation, as we know, is $2\pi\sqrt{h/g}$, and its angular frequency is $\omega = \sqrt{g/h}$. If the Earth did not rotate, the deflection of the pendulum from the pole would always be in a particular meridian plane and would change with time consistent with the law

$$r(t) = r_0 \sin \omega t,$$

where r_0 is the deflection amplitude. However, since the Earth rotates with an angular velocity Ω under the swinging pendulum, the motion of the pendulum with respect to the Earth is complicated. The angle ϕ of the meridian plane in which the oscillations begin increases linearly with time in the Earth's reference frame: $\phi = \Omega t$.

Therefore,

$$r(\phi) = r_0 \sin\left(\frac{\omega}{\Omega}\phi\right).$$

The plot of this function is shown in the Cartesian coordinate system in figure 4. This is an ordinary sine curve, but the interval $0 \le \phi \le 2\pi$ contains not just a single oscillation, as in a sine function, but ω/Ω oscillations. This number is not necessarily an integer. The dependence of r on ϕ is shown in polar coordinates in figure 5 (this is only a qualitative plot—a more exact plot can be drawn by using various values of h).

Build high towers for physics' sake, but please don't drop heavy objects from them! ◘

Quantum on rotation and non-inertial reference frames:

V. Surdin, "A Venusian Mystery," July/August 1996, pp. 4–8.

A. Leonovich, "Are You Relatively Sure?", September/October 1996, pp. 32–33.

A. Stasenko, "Merry-Go-Round Kinematics," September/October 1996, pp. 48–49.

M. Emelyanov, A. Zharkov, V. Zagainov, and V. Matochkin, "In Foucault's Footsteps," November/December 1996, pp. 26–27.

L. Kirkpatrick and A. Eisenkraft, "Around and Around She Goes," March/April 1998, pp. 30–33.

A. Stasenko, "Rivers, Typhoons, and Molecules," July/August 1998, pp. 38–40.

Modeling a tornado

by V. Mayer

TORNADO, CYCLONE, WATER-spout, and sand-storm belong to the most spectacular and enigmatic natural phenomena. Their energy is so high that nothing can withstand them.

How can a tornado carry heavy objects over large distances? How is a tornado generated? Modern science cannot provide comprehensive answers to these and many other questions.

Can a tornado be produced in a laboratory? We will describe two experimental setups in which aqueous models of a tornado can be easily produced even with a crude setup.

1. Solder a disk made of brass or tin plate with a diameter of 40 mm and a thickness of 0.5–1 mm to the shaft of a small electric motor (such as a motor used in toys). The disk must be fixed exactly perpendicular to the shaft in order to prevent wobbling of the rotor. To seal the motor, oil the bearings with a lubricant or vaseline and cover the electric contacts where the wires are attached with a layer of modeling clay. Take a glass or a glass jar with a diameter of about 9 cm and height of 18 cm and place a piece of modeling clay on its bottom. Attach the motor to the clay without letting the lower end of the shaft touch the clay. Attach the electric wires to the sides of the glass with tape. Figure 1 shows such a setup ready for experiments.

Pour water into the glass and a 1 to 2 cm layer of sunflower seed oil on top. Connect the motor's leads to a flashlight battery. The disk will start to rotate and spin the water above it. After a while, the boundary between the water and the oil will begin to bend downward, creating an oil-filled crater, which will grow until it comes into contact with the disk. At this time, the disk will break up the oil into droplets and the water in the glass will become turbid. After the electric motor is shut off, the oil droplets will float to the surface and form a layer of oil on the surface of the water. The experiment can then be repeated.

Figures 2 and 3 are photographs demonstrating the process of formation of an air crater in a slightly different experiment, in which the glass is filled with oil-free water.

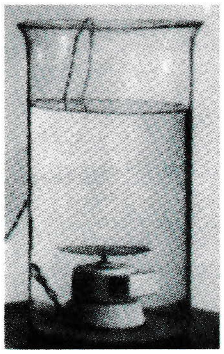


Figure 1

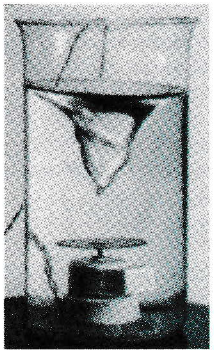


Figure 2

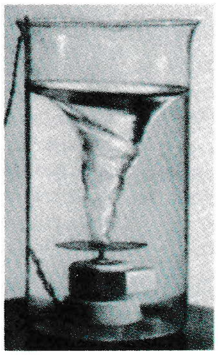


Figure 3

2. A closer approximation of a real tornado can be observed in the following experiments.

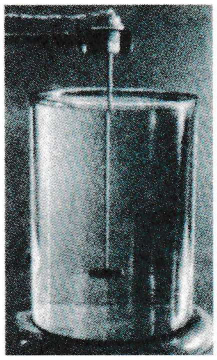


Figure 4

Solder a copper wire with a length of 25 cm and a diameter of 2 mm (a knitting needle will work) to the shaft of an electric motor. Solder a brass or tin square plate (10 × 25 × 0.5 mm) to the other end of the wire at right angles to it (figure 4). Turn on the motor to check the rotation of the propeller. If necessary, straighten the wire to minimize wobbling.

Now immerse the propeller vertically into a water-filled jar (diameter 15–20 cm, height 25–30 cm) and turn on the motor. You will see a gradual formation of a crater at the surface of the water and growth of a tornado in the direction of the rotating propeller (figures 5–7). When the bottom end of the vortex comes into contact with the propeller, many air bubbles, which trace the vortex surrounding the propeller, are produced.

Holding the motor in your hand, it is fascinating to watch the "predatory" movements of the tornado's cone.

Let's continue the experiment. Place a wooden block on the surface of the water—it will be swallowed by the vortex! By changing speed of the motor, make the block whirl in the crater at a constant depth under the surface of the wa-

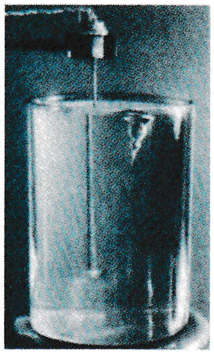


Figure 5

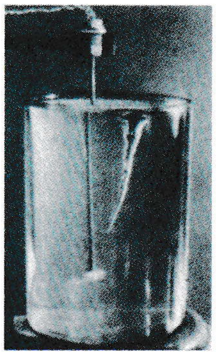


Figure 6

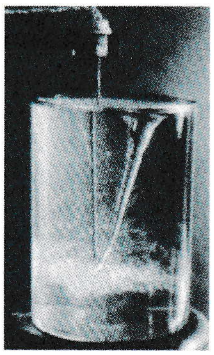


Figure 7

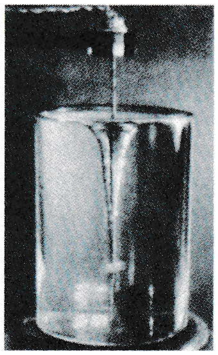


Figure 8

ter. In a similar way, a tornado will pull in objects that are denser than water (in contrast to the wooden block) lying on the bottom of the jar before the appearance of the vortex.

Now hold the electric motor so that its shaft is aligned with the jar's axis. You will see that the crater descends along the shaft and that under the propeller its path is traced by the air bubbles (figure 8). Put some well-washed river sand at the bottom of the jar and watch the structure of the vortex under the propeller.

These experiments show that the key process producing a tornado in a fluid is always a vortex. ◘

Bulletin Board

The following were the first ten correct entries in this month's CyberTeaser contest, *The Hose Knows*:

Jerold Lewandowski
(Troy, New York)
Theo Koupelis
(Wausau, Wisconsin)
Nick Fonarev
(Staten Island, New York)
Christopher Franck
(Redondo Beach, California)
Bruno Konder
(Rio de Janeiro, Brazil)
Maxim Bachmutsky
(Kfar-Saba, Israel)
John E. Beam
(Bellaire, Texas)
Chris Resmondo
(Pensacola, Florida)
Lue Chen
(New York, New York)
Anastasia Nikitina
(Pasadena, California)

Our congratulations to this month's winner, who will receive a copy of this issue of *Quantum* and the coveted *Quantum* button. Everyone who submitted a correct answer (up to the time the answer is posted on the Web) is entered into a drawing for a copy of *Quantum Quandaries*, a collection of 100 *Quantum* brainteasers. Our thanks to everyone who submitted an answer—right or wrong. The new CyberTeaser is waiting for you at http://www.nsta.org/quantum.

AT THE BLACKBOARD III

Equation of the Gaseous State

by V. Belonuchkin

THE EQUATION OF STATE of an ideal gas (or simply the ideal gas law) describes the relationship between the pressure, temperature, and volume of one of the simplest physical systems. Because the system is so simple, its equation of state is very simple, too:

$$PV = \frac{m}{\mu} RT,$$

where P is the pressure of the gas, V is its volume, m is its mass, μ is its molar mass, T is the temperature, and R is the universal gas constant.

According to this law, the pressure of an ideal gas is proportional to its temperature. What does this mean? Will the pressure always follow the ups and downs of the temperature? Of course not. The proportionality is valid only when the other parameters, namely, the volume, mass, and molar mass, are fixed. The pressure can fall instead of rising as the temperature increases even for a constant mass of the gas if the gas expands rapidly enough. There are other possibilities, as well. We shall discuss them in detail while solving some problems taken from the entrance exams for a physics and engineering institute.

Problem 1. *The temperature and pressure of a fixed mass of an ideal gas vary as shown in figure 1. Will its volume vary, and if yes, how will it change?*

Solution. Figure 1 shows that the pressure grows linearly with the temperature. Will the volume remain constant? If no, how can we recognize the diagram of a process that takes place at constant volume?

At constant volume the pressure rises in direct proportion to the temperature. The corresponding plot differs from other linear diagrams in that it passes through the origin. Thus, in P-T coordinates the isochor (line of constant volume) passes through the origin.

Let's plot the isochores which pass through points 1 and 2 (figure 2). These isochores are different, so the volumes corresponding to states 1 and 2 are also different. Which volume is greater? A simple way to answer this question is to use the plot again. Let's connect two isochores by, say, an isobar 1-3 (or by an isotherm 1-4, or by any other isobar or isotherm). We see that temperature rises along path 1-3 at constant pressure. Clearly the gas expands. Therefore, point 2 belongs to the isochor with greater volume, so the volume of the gas increases in going from state 1 to state 2.

Can the plot in figure 1 still represent a process taking place at a constant volume? Such a situation is considered in the next problem.

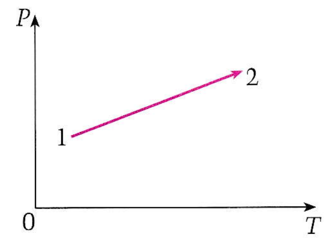


Figure 1

Problem 2. *Helium is put into a vessel of constant volume. The vessel is connected to a manometer and a thermometer, the readings of which vary according to figure 1. What can be said about the state of the gas? Specifically, is the pressure inside the vessel higher or lower than atmospheric pressure?*

Solution. In this case the volume of the gas is constant, but its pressure is not proportional to the temperature. How can this be possible?

Let's look at the ideal gas equation once again. In addition to the parameters P, V, and T, it also contains m and μ. The molar mass μ of helium cannot change, because helium is a monatomic gas and so cannot dissociate. Condensation can also be ruled out because the temperature is rising. Thus only one possibility is left: the mass m of helium gas in the vessel must be changing.

Since the pressure grows more slowly than in the isochoric process for a constant mass, the mass of gas must be decreasing. In other words, the vessel is leaky, and helium is flowing into the surroundings. This means that the pressure in the vessel is higher than the atmospheric

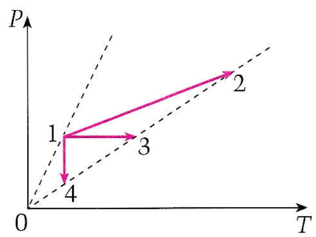


Figure 2

pressure—otherwise there would be a reverse flow of air into the vessel.

The ideal gas equation of state is a tool that can be used to compare not only the states of the same gas but also the states of different gases. Here is another example.

Problem 3. *Ball lightning is a dimly glowing gaseous ball that floats freely in the air. According to Stakhanov's model of ball lightning (one of the models constructed to explain the nature and behavior of this phenomenon), the gas inside this ball is made up of complex molecular aggregates: each particle consists of a nitrogen atom bound to a number of water molecules. The electrons lost by the nitrogen atoms are accepted by the water molecules, so every complex molecule is electrically neutral. Determine how many molecules of water are bound to each nitrogen atom if the temperature inside the ball is* $T = 600°C$ *and the temperature of the surrounding air is* $T_0 = 20°C$.

Solution. Since the ball lightning floats in the air, its density is equal to that of the air. Probably the pressure inside the ball is equal to atmospheric pressure, as well. These two conditions yield

$$\frac{P}{\rho} = \frac{P_0}{\rho_0},$$

or

$$\frac{PV}{m} = \frac{P_0 V_0}{m_0}.$$

Here all the variables with subscripts pertain to the air and those without subscripts pertain to the gas of molecular aggregates. The ideal gas equation yields

$$\frac{T}{\mu} = \frac{T_0}{\mu_0},$$

from which the molar mass of the molecular aggregate can be determined using the molar mass of air ($\mu_0 = 29 \cdot 10^{-3}$ kg/mol):

$$\mu = \mu_0 \frac{T}{T_0} \approx 86 \cdot 10^{-3} \text{ kg/mol}.$$

The molar mass of atomic nitrogen is $14 \cdot 10^{-3}$ kg/mol, and that of water is $18 \cdot 10^{-3}$ kg/mol; therefore, each nitrogen atom binds with four molecules of water.

The external conditions around a gas can be changed in many ways, but in every case the state of the gas can be described by the ideal gas equation of state.

Problem 4. *A gas-filled cylinder with a cross-sectional area* $S = 10$ cm^2 *is closed by a massive piston. The cylinder begins to ascend with an acceleration of 2g. When the temperature of the gas becomes equal to the initial temperature, the volume under the piston has decreased by 1.5 times. Find the mass m of the piston. The external pressure is* 10^5 Pa.

Solution. In the motionless state (at rest) the piston's weight was counterbalanced by a pressure difference between the inside and outside of the cylinder:

$$mg = (P - P_0)S.$$

In the accelerating cylinder the total force applied to the piston imparts an upward acceleration of 2g. The volume of the gas decreases 1.5-fold at a constant temperature, so the pressure of the gas must have risen by the same factor. Therefore,

$$2mg = (1.5P - P_0)S - mg,$$

or

$$3mg = (1.5P - P_0)S.$$

Solving both equations simultaneously, we get

$$m = \frac{P_0 S}{3g} \approx 3.4 \text{ kg}.$$

We have considered some examples of problems for which the ideal gas equation applies (of course these do not cover the full range of problems that can be solved with the help of this equation). However, it is also important to know the conditions for which this famous equation cannot be used (in other words, to find its domain of applicability).

The model of an ideal gas is based on the assumption that the energy of molecular interaction is negligible in comparison with the average kinetic energy of the molecules. This approach envisages molecules as small elastic balls, which interact only during the time of a collision. The diameter of the balls is much smaller than the mean free path.

Certainly, such a model is oversimplified. Indeed, what is the "size" of a molecule? The effective molecular diameter is reasonably assumed to be the distance at which the motion of a molecule is altered by another. Clearly, within the framework of this definition, the molecular size should depend on many physical conditions; in particular, the size should be temperature-dependent. This is confirmed by experimental evidence: the effective molecular diameter does indeed decrease as the temperature rises. This is not unexpected: the kinetic energy grows with temperature, so the molecules must approach each other more closely in order for the potential energy of their interaction to become comparable to the kinetic energy of their motion (otherwise the molecular trajectories will not change significantly). However, this dependence is very weak, so it makes sense to treat the molecular diameter as having a definite constant value.

The properties of real gases begin to deviate noticeably from the ideal gas model under conditions such that the molecules collide with each other frequently (so that the assumption that most of the time they are not interacting is wrong). Under such conditions the mean free path (the average distance between two successive collisions) becomes comparable with the size of a molecule. Since the majority of the time the molecules of such a gas are found close to one another, their interaction cannot be neglected.

Under what conditions does this occur? Before answering this question, let's solve the following problem.

Problem 5. *Evaluate the mean free path of air molecules under standard conditions. Assume the molecular diameter to be* $d = 3.7 \cdot 10^{-10}$ m.

Solution. Two molecules collide when the distance between their centers becomes less than the molecular diameter d. Suppose that a molecule travels a distance l during some period of time. Along the way it collides with molecules whose centers are located inside a broken (zig-zag) cylinder, where the break points (zigs and zags) correspond to collisions.

The total length (height) of the cylinder is l and the area of its base is πd^2. The number of molecules in this volume is $nl\pi d^2$ (n is the number of molecules per unit volume) and is equal to the total number of collisions. Let's calculate the distance between successive collisions, that is, the mean free path in the gas:

$$\lambda = \frac{l}{nl\pi d^2} = \frac{1}{n\pi d^2} = \frac{RT}{N_A P\pi d^2} \approx 8.75 \cdot 10^{-8}\ \text{m}.$$

Here the pressure $P = 10^5$ Pa, the temperature $T = 273$ K, and Avogadro's number $N_A = 6.02 \cdot 10^{23}\ \text{mol}^{-1}$.

The experimental value for these conditions is $6.20 \cdot 10^{-8}$ m. The difference between the experimental and theoretical values is explained mainly by the fact that we considered all the molecules (except the chosen one) to be motionless. A detailed analysis (which is beyond the elementary physics course) shows that the relative molecular motion changes the theoretical value of the mean free path by a factor of $1/\sqrt{2}$. Multiplying our theoretical value by this factor, we get $6.19 \cdot 10^{-8}$ m.

Right now we are more interested in a different matter. Under normal conditions the mean free path is about 200 times greater than the molecular diameter, so the ideal gas approximation works rather well. In contrast, when the pressure rises by a factor of 100–200 at constant temperature, the mean free path becomes comparable to the molecular size. This means that under such high pressure (or, strictly speaking, at high density) the molecules are hardly ever far apart, so one cannot neglect their interaction. In this case the ideal gas model does not work. Incidentally, the densities of condensed phases (liquids or solids) are about 1000 times greater than the densities of gases. A gas with a density which is only a factor of 5–10 times smaller than that of a liquid is no longer ideal. However, if the temperature is high enough, a gas can in many cases be considered ideal even at high pressures because the key role is played by density and not by pressure.

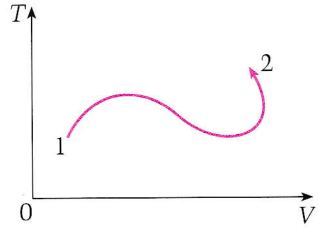


Figure 3

Problems

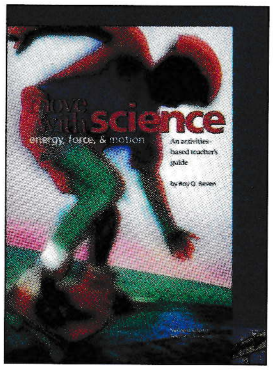
1. Figure 3 shows a plot which represents changes of state of some mass of an ideal gas. Find the fragments of the plot which correspond to increasing and decreasing pressure.

2. An electrical discharge is produced in a vessel containing oxygen. As a result, all the oxygen is transformed into ozone, and the temperature is doubled. How has the pressure in the vessel changed? Assume the volume of the vessel to be constant.

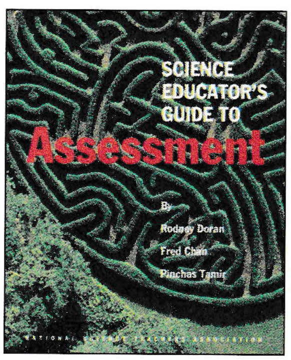
3. A piston of mass $m = 5$ kg and cross-sectional area $S = 10\ \text{cm}^2$ can move inside a gas-filled cylinder. When the cylinder is moved downwards with an acceleration of $4g$, the volume of gas under the piston doubles. Find the external pressure if temperature of the gas does not vary.

4. A satellite with cross-sectional area $S = 1\ \text{m}^2$ moves near the Earth with orbital velocity $v = 7.8$ km/s. Atmospheric pressure at the orbital altitude of 200 km is $P = 1.37 \cdot 10^{-4}$ Pa, and the temperature $T = 1226$ K. Find the number of collisions of the satellite with air molecules during one second.

ANSWERS, HINTS & SOLUTIONS ON PAGE 53



Move with Science:
Energy, Force & Motion
Roy Q. Beven
This resource uses methods of transportation that are most familiar to high school students to connect basic concepts of physics and human biology to the concrete sights, sounds, and physical sensations that students experience nearly everyday. *Move with Science* brings real transportation situations—representing such concepts as inertia, stability, and the relationships between mass, energy, and motion—into the classroom through hands-on activities. (Grades 9–12, 1998, 160 pp.)
#PB144X $21.95



Science Educator's Guide to Assessment
R. Doran, F. Chan, and P. Tamir
Give students regular and accurate feedback, reinforce productive learning habits, and help students reflect on their own learning. Incorporate ready-to-use assessment activities keyed directly to the *National Science Education Standards*. (Grades 7–12, 1998, 220 pp.)
#PB145X $27.95

P500Q2

The NSTA Science Store

❖ ❖ ❖ ❖

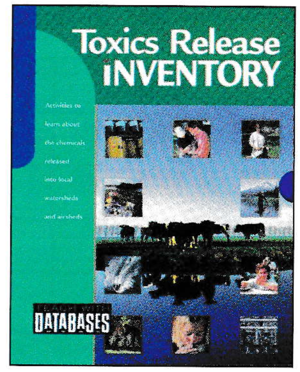
Discover innovative ideas to enhance your science education program

NSTA

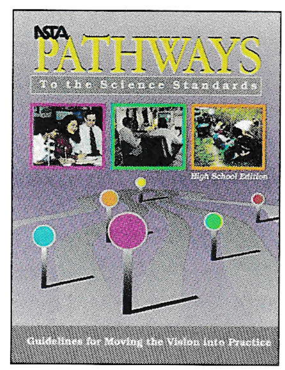
Call (800) 722–NSTA to order your copies today!

❖ ❖ ❖ ❖

National Science Teachers Association
www.nsta.org



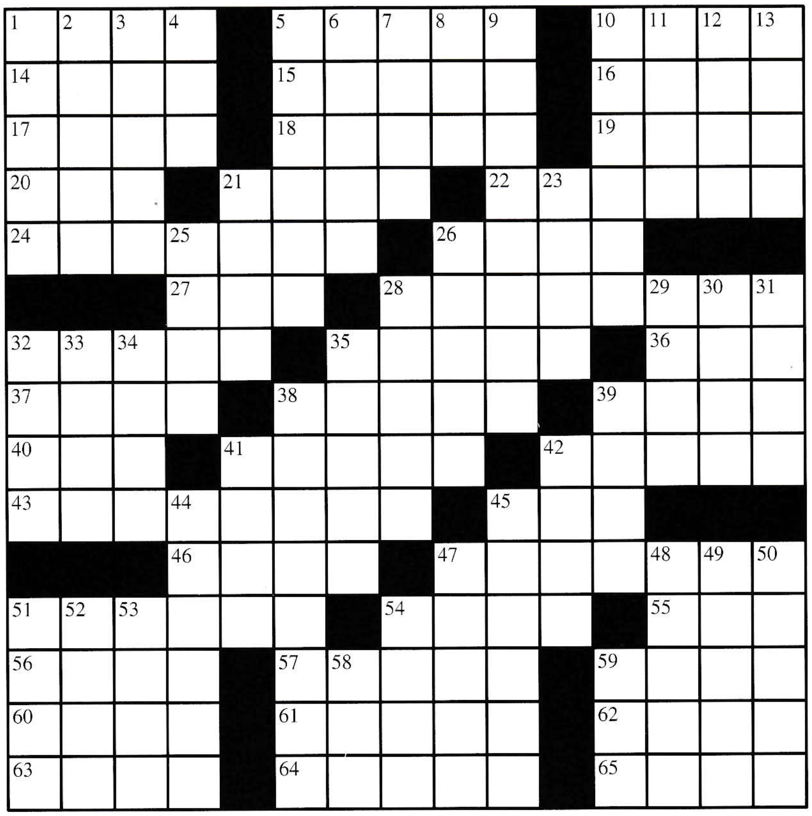
Teach with Databases:
Toxics Release Inventory
Jay Barracato
This package offers *Toxics Release Inventory* as a tool for student investigation, with a range of labs covering sampling and analysis methods, computer-based explorations, and guidance for using this real data to investigate the chemical history of your local watershed. Includes: Getting Started, Teacher's Guide, EPA Guide, CD-ROM, and *Database Basics*. (Grades 9–12, 1998)
#PB143X01 $35.00



NSTA Pathways to the Science Standards
This practical guidebook demonstrates how you can carry the vision of the Standards—for teaching, professional development, assessment, content, program, and system—into the real world of the classroom and school. Filled with specific suggestions and clear examples on how to implement each of the Standards, *Pathways* is a valuable resource for everyone involved in science education. (Grades 9–12, 1996, 196 pp.) **#PB126X $34.95**

Criss × cross science

by David R. Martin



ACROSS

1 Refrain syllables
5 Degrade
10 Archaeologist ___ Kathleen Kenyon
14 60 coulombs: abbr.
15 Dissident
16 Nucleotide sequence
17 Fruit's skin
18 Mountain ridge
19 Bend
20 Computer language
21 ___ group (chem. group)
22 Duplicates
24 Dewey ___ System
26 Ukrainian river
27 Choose
28 Element 24
32 Mesons
35 Move fast
36 Southern constellation
37 Mistral, in France
38 Lime-bearing silicate
39 Prayer ending
40 Basketball org.
41 Practitioner: suff.
42 Force a boat to shelter
43 Salt derived from SiO_2
45 ___ rata
46 City west of Charleroi
47 Angle units
51 Relax
54 Gram-force
55 Cesium iodide
56 Element 82
57 Like Argon
59 Metallurgy fuel
60 Compacted snow
61 Prior to
62 Smelting product
63 God of war
64 Chaotic
65 Waxed fabric

DOWN

1 Capacitance unit
2 Organic compound
3 Linear accelerator
4 ___ gate (logic circuit element)
5 Site of Noah's Ark
6 A gemstone
7 English chemist Frederick Augustus ___ (1827-1902)
8 Collection
9 Fermion
10 Distort
11 Nerve cell arm
12 Chemical measure
13 Odd's partner
21 Current units
23 French astronomer Bernard ___ (1897–1952)
25 Plasma particles
26 Shaven
28 838,318 (in base 16)
29 Verse segment
30 Carbamide
31 Various
32 Potassium thiocyanate
33 Phosphatase unit
34 Type of exam
35 Short comic sketches
38 Element 21
39 Certain asteroid
41 Computer screen image
42 Energy units
44 Chemical compounds
45 Boldly
47 Actress Day
48 Intestinal bacteria
49 Gravel ridge (alt. sp.)
50 Surround a castle
51 Arm bone
52 Poet's never
53 Periodic disturbance
54 Circuit elements: abbr.
58 Compass direction
59 Trig. function

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE MARCH/APRIL PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M291

Assume that no more than ten new errors were made at each stage. On the first day the total number of errors will then be no greater than ten, on the second day it will be no greater than $2 \cdot 10 + 10$, on the third day no greater than $2(2 \cdot 10 + 10) + 10 = 2^2 \cdot 10 + 2 \cdot 10 + 10$, and so on. On the nth day, the number of errors will be no greater than $(2^{n-1} + 2^{n-2} + \ldots + 2 + 1)10 = (2^n - 1)10$.

However, we know that there was a day when each paper contained at least 10 errors. If this was on the nth day, then the total number of errors on that day will be no less than $2^n \cdot 10$ (since the number of students who sent their solutions that day was 2^n), which contradicts the estimate obtained above.

M292

Denote by P the midpoint of BC (figure 1). The triangles BMP and BPA have a pair of congruent angles: $\angle MBP = \angle BAP$ (both of them are measured by the half of arc BM), and $\angle BPM = \angle BPA$. Therefore, these triangles are similar and

$$\frac{BP}{PM} = \frac{AP}{BP}.$$

However, $BP = PC$. Thus,

$$\frac{CP}{PM} = \frac{AP}{CP}.$$

Since triangles CMP and CPA contain these proportional line segments, and since the angles they include in each triangle are equal, these triangles are also similar, and we obtain the equality $\angle MCP = \angle CAP$. Thus, $\angle BMC = 180° - \angle MBC - \angle MCB = 180° - \angle MBP - \angle MCP = 180° - \angle BAP - \angle CAP = 180° - \angle BAC = 180° - \alpha$.

M293

Let us rewrite the left inequality as

$$0 \le \frac{t^2 - q}{2t + p} - x_1 = \frac{t^2 - q - 2tx_1 - px_1}{2t + p}$$
$$= \frac{t^2 - 2tx_1 - x_1x_2 + (x_1 + x_2)x_1}{2t + p}$$
$$= \frac{t^2 - 2tx_1 + x_1^2}{2t + p} = \frac{(t - x_1)^2}{2t + p}.$$

(We use here the formulas for the sum and product of the roots of quadratic equation.) Similar transformation of the right inequality yields the system of inequalities

$$\frac{(t - x_1)^2}{2t + p} \ge 0, \; \frac{(t - x_2)^2}{2t + p} \le 0.$$

If t is not equal to x_1 or x_2, then we obtain $2t + p > 0$ and $2t + p < 0$, which is impossible. Thus the assertion is proved.

M294

A trigonometric solution to this problem is not difficult to find. We give here a more geometric solution.

The segment CM (figure 2) must intersect either side CA or CB of triangle ABC. Suppose, without loss of generality, that it intersects side AB. Consider a point P on AB such that CM bisects angle BCP. Now in triangle BCP, point M is the intersection of the bisector of the interior angle at C and of the exterior angles at B. Hence M is equidistant from lines CB and CP, and also from lines CB and PB. Hence M is equidistant from the lines CP and PB, and must be on the bisector of angle CPA. (Point M is the center of an *escribed* circle for triangle BCP.)

Let $\angle BCP = 2x$ and $\angle BPC = 2y$, so that $2x + 2y + 2\alpha = 180°$. Then $\angle MCP = x$. What is the measure of $\angle MPC$? Well, the argument of the previous paragraph shows that $\angle MPB = (1/2)\angle APC = (1/2)(180° - 2y) = 90° - y$, and $\angle MPC = \angle MPB + \angle BPC = (90° - y) + 2y = 90° + y$. Finally, $\angle CMP = 180° - x - (90° + y) = 90° - x - y = \alpha$. Thus, $\angle CMP = \angle CAB$. This conclusion, together with the fact that $CM = CA$, shows, in particular, that point P lies inside segment AB rather than on its extension.

Now triangles CAP and CMP have a common side CP, $CM = CA$, and the angles at M and A are equal. Thus either these triangles are congruent, or the angles opposite sides CM and CA add up to 180°.

Consider the triangles CAP and

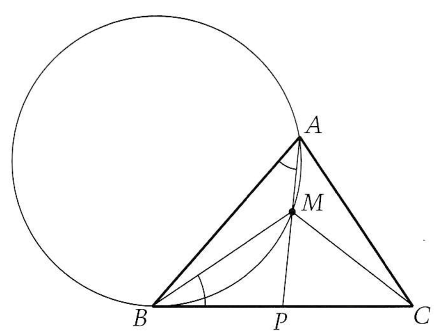


Figure 1

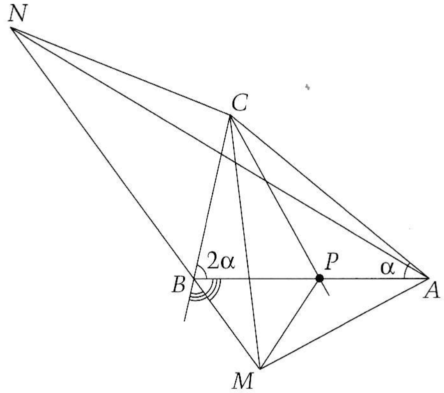


Figure 2

CMP. They have side *CP* in common, the pair of sides *CM* and *CA* are equal, and the angles at the vertices *M* and *A* are equal. Must they be congruent? Well, if two triangles agree in two sides and a non-included angle, they may or may not be congruent. But if we use law of sines to find sin $\angle CPM$ and sin $\angle CPA$, we will see that the expressions for these two sines are equal. Hence either $\angle CPM = \angle CPA$, and the two triangles are congruent, or $\angle CPM + \angle CPA = 180°$. If the latter were true, then point *M* would lie on line *AB*, which is not the case. Hence triangles *CAP*, *CMP* are in fact congruent.

It follows that *CP* bisects angle *ACM*, which means that *CM* and *CP* divide angle *ACB* into three equal parts. Since $\angle ACB = 180° - \angle BAC - \angle CBA = 180° - 3\alpha$, each of these parts measures $60° - \alpha$. We have $\angle CPM = \angle CPA = 180° - (60° - \alpha) - \alpha = 120°$, $\angle CMB = 180° - \angle BCM - \angle CBM = 180° - x - (90° + \alpha) = 90° - x - \alpha = y = (1/2)\angle CPB = 30°$, and $\angle MCN = 120°$. Point *C* is the center of the circle circumscribed about triangle *AMN*. Therefore, $\angle MNA = (1/2)\angle MCA = 60° - \alpha$ and $\angle MAN = (1/2)\angle MCN = 60°$. Finally, $\angle AMN = 60° + \alpha$.

The alert reader will note that this solution is purely geometric, except for the observation about triangles *CPM* and *CPA*. The reader is invited to rephrase this part of the proof to eliminate its dependence on results from trigonometry.

M295

We give a solution for the corresponding problem with arbitrary n (in our case, $n = 100$). It is clear how the condition of the problem is written for the general case: n positive numbers are considered, the left-hand side of the inequality in the assumption is the sum of n fractions, and we must prove that the product of the given numbers is not less than $(n-1)^n$. Notice that the inequality in the assumption becomes an equality when all given numbers are equal to $n - 1$.

We will use the following notation: $t = \sqrt[n]{a_1 a_2 \ldots a_n}$, S_k is the sum of all possible products of the given numbers by k numbers in each product (all numbers in each product are different), and G_k is the geometric mean of all possible products of the given numbers by k factors in each product (all numbers in each product are different). The well-known theorem on the geometric and arithmetic means gives

$$S_k \ge N_k\, G_k, \qquad (1)$$

where N_k is the number of all possible products of the given numbers by k factors in each product (in fact, this is the number of combinations of n items k at a time, but, for our solution, we need not know how to find this number). It is easy to see that

$$G_k = t^k. \qquad (2)$$

This follows from the fact that all terms in G_k have the power of k. Therefore, G_k also has the same power. Since all a_i are equivalent, G_k can be written as $G_k = t^\lambda$. However, the power of t is 1; therefore, $\lambda = k$.

Let us now transform the inequality in the assumption (for arbitrary n); i.e., we drop the fractions and collect similar terms, putting on the left side all the terms except the highest-power term $S_n = t^n$. We thus obtain the inequality

$$A_0 + A_1 S_1 + \ldots + A_{n-2} S_{n-2} + A_{n-1} S_{n-1} \le S_n. \qquad (3)$$

The coefficients A_i can be easily represented in terms of n, in particular, $A_{n-1} = 0$, although there is no need to do so. It is important, however, that all of them should be nonnegative. Using inequality (1) and equation (2), we obtain the following inequality from (3):

$$B_0 + B_1 t + \ldots + B_{n-1} t^{n-1} \le t^n. \qquad (4)$$

All the coefficients on the left-hand side of (4) are nonnegative. Therefore, for a positive value of t the solution of inequality (4) is $t \ge t_0$, where t_0 is the only number that turns inequality (4) into an equality. This number has already been obtained: $t_0 = n - 1$. With it, the inequality in the assumption of the problem becomes an equality, as does inequality (3) and all inequalities (1). Thus, we have proved that $t = \sqrt[n]{a_1 a_2 \ldots a_n} \ge n - 1$; i.e., $a_1 a_2 \ldots a_n \ge (n-1)^n$.

Physics

P291

While the velocity of the boat is always directed along the surface (figure 3), the projection of this velocity onto the rope is constant and equal to the speed at which the rope is reeled onto the drum:

$$v_0 = v \cos \alpha.$$

After a small time interval Δt, the rope turns through a small angle

$$\Delta\alpha = \frac{v \Delta t \sin\alpha}{L}.$$

Thus, the angular velocity of "rotation" of the rope is

$$\omega = \frac{\Delta\alpha}{\Delta t} = \frac{v \sin\alpha}{L}.$$

The velocity of the knot is the vector sum of the translational velocity along the rope (v_0) and the linear velocity of rotation, which is determined by the location of the knot:

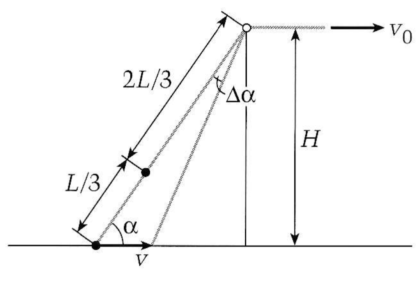
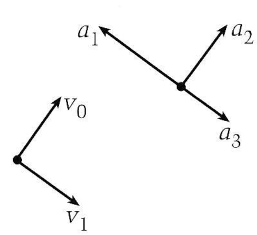


Figure 3



$$v_1 = \frac{2}{3} L\omega = \frac{2}{3} v \sin\alpha .$$

The total velocity of the knot is

$$v_2 = \sqrt{v_0^2 + v_1^2} = \sqrt{v^2 \cos^2\alpha + \frac{4}{9} v^2 \sin^2\alpha} = \frac{v}{3}\sqrt{9\cos^2\alpha + 4\sin^2\alpha}.$$

To find the acceleration, we must introduce another value, which was not specified in the conditions of the problem.

Let's designate the height of the motor above the surface of the water as H. The total acceleration of the knot is the sum of three components: the first and second components are determined by the rotation of vectors $\mathbf{v}_0$ and $\mathbf{v}_1$, respectively, while the third component results from the change in the magnitude of $\mathbf{v}_1$.

The first component, which is normal to $\mathbf{v}_0$, is

$$a_1 = v_0\omega = v_0 \frac{v\sin\alpha}{L} = \frac{v_0^2 \sin^2\alpha}{H\cos\alpha}.$$

The second component is normal to the linear velocity of rotation, that is, it is directed along the rope:

$$a_2 = v_1\omega = \frac{2}{3} v\sin\alpha \cdot \frac{v\sin\alpha}{L} = \frac{2}{3}\frac{v_0^2 \sin^3\alpha}{H\cos^2\alpha}.$$

The third component is directed along the velocity :

$$a_3 = \frac{2}{3}\frac{v_0\tan(\alpha+\Delta\alpha) - v_0\tan\alpha}{\Delta t} = \frac{2}{3}\frac{v_0\sin\Delta\alpha}{\cos^2\alpha\cdot\Delta t} = \frac{2}{3}\frac{v_0\omega}{\cos^2\alpha} = \frac{2}{3}\frac{v_0^2\sin^2\alpha}{H\cos^3\alpha}.$$

Finally, we obtain the vector sum (with allowance for the signs of the components) and calculate the magnitude of this sum:

$$a = \sqrt{(a_1 - a_3)^2 + a_2^2} = \frac{2}{3}\frac{v^2\sin^2\alpha}{H}\sqrt{\frac{(1.5\cos^2\alpha - 1)^2}{\cos^2\alpha} + \sin^2\alpha}.$$

P292

At a temperature of $+100\ ^\circ$C the pressure of saturated water vapor is 1 atm $\cong 10^5$ Pa. This pressure is the pressure developed by a 10-meter-high water column. In contrast, the pressure of saturated vapor at the initial temperature of 10°C is much lower than 1 atm. We may assume, therefore, that a pressure of 1 atm on the hot oceanic planet is produced by water molecules evaporated due to heating. Before heating, these molecules occupied the "upper" 10-meter layer of the ocean (since the thickness of the atmosphere is much less than the radius of the planet, the weight of the water layer does not change after its evaporation).

The remaining layer of water with the thickness of 220 m expands and compensates for the volume of water evaporated. The coefficient of thermal expansion is defined as the fractional change in the volume per degree change in temperature. The average thermal expansion coefficient of water in the specified temperature range is

$$\alpha = \frac{\Delta V}{V\Delta T} = \frac{10}{230\cdot 90}\ \text{K}^{-1} \approx 5\cdot 10^{-4}\ \text{K}^{-1}.$$

P293

Let's consider how the electric force affecting the spear depends on the position of the spearhead (figure 4). Since there is no field outside the layers, the force is zero when $x > 2h$ or $x < 0$. As the spear penetrates the layers, the braking force grows linearly with distance, attaining a maximum when the spearhead exits the first layer. As the spearhead penetrates the second layer, the braking force continually decreases. The braking force becomes zero when the spearhead exits the second layer. The force remains zero until the tail of the spear enters the first layer. At this time the electric force causes the spear to speed up. If the speed of the spear has not dropped to zero before this time, the spear will pass completely through both layers.

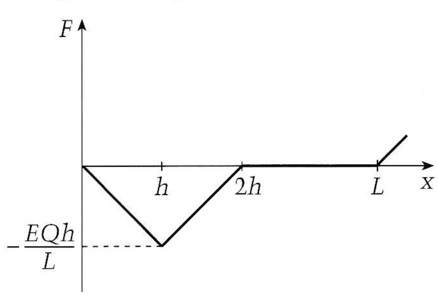


Figure 4

The minimum value for the initial speed required to pass through both layers can be determined from energy conservation by equating the initial kinetic energy to the work performed by the braking force. The latter can be easily calculated for such a simple dependence of force on distance:

$$\frac{Mv_0^2}{2} = F_{\text{mean}}\cdot 2h = \frac{1}{2}\frac{EQh}{L}2h = \frac{EQh^2}{L}.$$

Therefore

$$v_0 = h\sqrt{\frac{2EQ}{ML}}.$$

P294

The basic law of radioactive decay is

$$N(t) = N_0 \cdot 2^{-t/\tau},$$

where $N(t)$ is the number of parent nuclei at a time t after an arbitrary starting point, N_0 is the initial number of parent nuclei, and τ is the half-life of these nuclei. In our case we start counting time from the moment of the Earth's birth. If N_0 is the number of nuclei of each isotope in natural uranium at the moment of the Earth's birth, the number of these nuclei at the present time t is

$$N_1(t) = N_0 \cdot 2^{-t/\tau_1}$$

and

$$N_2(t) = N_0 \cdot 2^{-t/\tau_2}.$$

Dividing one equation by the other, we obtain

$$\frac{N_1(t)}{N_2(t)} = \frac{\eta_1}{\eta_2} = 2^{t\left(\frac{1}{\tau_2} - \frac{1}{\tau_1}\right)}.$$

Taking the logarithm of both terms of this equation gives the age of the Earth:

$$t = \frac{\ln(\eta_1/\eta_2)}{\ln 2}\frac{\tau_1\tau_2}{\tau_1 - \tau_2} \approx 6\cdot 10^9 \text{ years.}$$

P295

A part of the luminous flux will be lost due to multiple reflections, but this loss may be compensated by refraction, which "shifts" the source toward the photosensitive detector.

Let's consider the reflections. At the air-glass boundary, only $1-(n-1)^2/(n+1)^2 = 8/9$ of the incident light enters the plate. Since the same reflection occurs at the glass-air boundary, $(8/9)(8/9) = 64/81$ of the luminous flux leaves the plate. There are, however, other portions of light, which contribute to the final flux. Some of the light, which is reflected many times into the glass at each boundary, leaks from the plate. Initially 8/9 of the light enters the plate and $(1/9)(8/9)$ is reflected backward from the glass-air boundary, and after the next reflection at the opposite boundary only $(1/9)^2(8/9)$ stays within the glass, while $(1/9)^2(8/9)^2$ leaks out. As a result, we have the sum

$$\left(\frac{8}{9}\right)^2 + \left(\frac{8}{9}\right)^2\left(\frac{1}{9}\right)^2 + \left(\frac{8}{9}\right)^2\left(\frac{1}{9}\right)^4 + \dots$$
$$= \left(\frac{8}{9}\right)^2 \frac{1}{1-1/81} = \frac{4}{5}.$$

Thus, 4/5 of the incident light passes through the plate.

Let's now consider the refractive effects of the plate, which "shifts" the source toward the detector. Consider the beam that leaves point A at a small angle α to the horizontal (figure 5). After refraction, the angle will be α/n. The beam will leave the plate at the same angle, but it appears to originate from point B. The image of the source is therefore nearer to the detector by $\Delta L = AB$. Simple geometry and simplifications due to the small value of the angle α yield

$$\Delta L = d\left(1-\frac{1}{n}\right) = \frac{d}{3},$$

where d is the thickness of the plate.

Thus, we may write

$$\frac{I}{L^2} = \frac{4}{5}\frac{I}{(L-d/3)^2},$$

where I is the intensity of the luminous flux. This equation yields

$$d = 3L\left(1-\frac{2}{\sqrt{5}}\right) \approx 3.2 \text{ cm.}$$

This solution, in fact, is only an approximation, because we disregarded the fact that multiple reflections produce slightly different "shifts" of the light source to the detector. The correction, however, is very small.

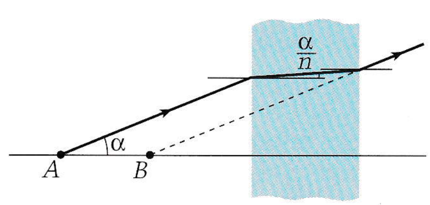


Figure 5

Brainteasers

B291

The first number can be represented as a sum of two numbers consisting of 14 digits 1 and 7 digits 1, respectively. We can then write the difference in the form

$$\frac{10^{14}-1}{9} + \frac{10^7-1}{9} - \frac{3(10^7-1)}{9}$$
$$= \frac{10^{14} - 2\cdot 10^7 + 1}{9} = \frac{(10^7-1)^2}{9}$$
$$= (3{,}333{,}333)^2.$$

B292

Let O be the center of the circle inscribed in triangle ABC (figure 6). We note that triangles POM, QOC are congruent, so that $OM = OC$. Similarly $ON = OC$. Hence O is also the center of the circle circumscribed about the triangle MCN. It is not hard to see that triangles POM, RON are congruent isosceles right triangles, and it follows that $\angle MON = 90°$. Therefore, $\angle MCN = 45°$, since an inscribed angle is half as large as a central angle with the same arc.

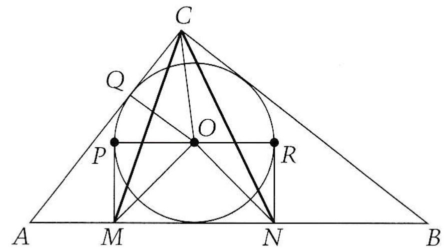


Figure 6

B293

There are two possible cases: (a) the smaller tank is filled first from the less powerful hose, or (b) it is filled first from the more powerful hose. Let us work with case (a) first. In this case, it will take 6.3/2.9 minutes to fill the first half of the small container, during which the larger container receives (6.3/2.9)(8.7) liters of water. Then it will take 6.3/8.7 minutes to fill the second half of the small container, and the same amount to top off the larger one (because the jobs are completed at the same time). So the larger container, during this time, receives (6.3/8.7)(2.9) liters of water. Since this fills the larger container, it must contain

$$(6.3/2.9)(8.7) + (6.3/8.7)(2.9) = 21 \text{ liters.}$$

Case (b) is left for the reader, who may notice that it requires exactly the same arithmetic as case (a).

B294

Let us number the caves from 1 to 16 consecutively. We will outline a strategy allowing the sheriff to catch Elusive Joe.

The sheriff must search the caves in order, starting with the first cave. Each day, he searches some cave, and Joe is hiding in some cave. Let us prove that the parity of the sum of the numbers of these two caves doesn't change (so long as the sheriff searches them in order).

Indeed, suppose Joe starts in an odd cave. The sheriff also starts in an odd cave (it's number 1), and the sum of these two numbers is even. Each day, both numbers change by 1, so the parity of their sum remains even. Similarly, if Joe starts in an even cave, the parity remains odd.

Suppose the parity is even. Can Joe slip through the sheriff's search? This can happen only if Joe finds himself in a cave adjacent to the one being searched, and moves that night to the one that has just been searched. But then the sheriff and Joe would have found themselves in adjacent caves, and the parity of the sum would not be even. Hence, if the parity is even, Joe must be caught.

This will happen when the sheriff searches the cave Joe is in, so that the sum of their cave numbers is simply double the number of one cave. How long will this take? If Joe lives up to his name, the sheriff will be forced to search up to cave 14. If he finds it empty, then Joe must be in cave 16 (he must be in an even cave, and cannot have slipped through). On the next day, the 15th, he will move to cave number 15 and will be caught. This is the largest number of days the search can take in this case. Note that this happens if Joe chooses an odd cave initially.

Things are more complicated if Joe has chosen an even cave initially. Then the parity, during the sheriff's sequential search, is odd, and Joe can slip past the Sheriff. By the above argument, the Sheriff (who reasons as well as we do) will know this has happened when he reaches the 15th cave and finds it empty. On the day this happens, Joe must be hiding in a different even-numbered cave. That night, he must move to an odd-numbered cave. The sheriff can come back and search cave 15 (into which Joe might have moved), and the sum of Joe's cave and the sheriff's cave will again become even.

Now the Sheriff can search down the cave numbers, from 15 to 1. As we noted earlier, during such a sequential search Joe can only slip through if the parity of the sum of the cave numbers is odd. Again, since it is even, he will be caught. The sheriff has essentially changed the parity of the sum by searching cave 15 twice.

How long will this take? The initial search, and the double search of cave 15, can take 16 days. If the sheriff reaches cave 3 without finding Joe, then Joe must be hiding in cave 1. That night he will move to cave 2, and be caught the next day. This will be the 29th day of searching, and May has 31 days. So the Sheriff could even catch Joe if he had started on February 1, 2000. (Solution by Jonathan Raasch.)

B295

The water in the small pan will not boil.

Kaleidoscope

Problems

1. No, because the astronaut is affected by the Earth's gravitation, which confines the spaceship to its orbit.

2. Negative.

3. The rocket orbiting the planet has received sufficient energy to lift it from the planet's surface. Therefore, in comparison with a rocket on the ground, it has more energy and requires less additional energy to reach the outskirts of the universe.

4. Only the molecules whose kinetic energy is larger than the work required to escape the surface of the liquid can enter the vapor phase. Therefore, the mean value of the kinetic energy of the remaining molecules decreases during evaporation, as does the temperature.

5. A wet film covers the grains of sand, and surface tension draws the grains together.

6. It occurs as a result of the decrease in the kinetic energy of thermal motion of the molecules, that is, a decrease in temperature.

7. Heating the semiconductor and/or shining light on it.

8. The plate whose work function is larger will be negatively charged.

9. Like the molecules in an evaporating liquid, only the fastest electrons, whose energy is greater than the work function, can escape the heated metal.

10. By changing the temperature of the cathode filament.

11. The electrons that are generated by intensive thermionic emission from the hot cathode produce an impact ionization of gas molecules, which decreases the electric resistance of the gaseous gap.

12. The lighter the incident particle, the smaller is the energy needed to ionize an atom.

13. Yes, it can. This process involves the ionization of the hydrogen atom.

14. More energy is needed to remove the second electron, because the binding energy of this electron is larger since it no longer feels the repulsion of the first electron.

15. No, because the energy that is equal to the binding energy of the hydrogen atom is released in this process.

16. The energy of the β-particles is so large that no transitions in the electron shell could generate it.

17. The photons are attracted to the star. Attempting to escape the potential well of the star's gravitational field, they lose energy.

Microscopic experiment

The oil molecules are bound into circles by the surface tension.

At the blackboard III

1. Transitions from rising to falling pressure occur at those points of the plot (figure 3, p. 46) where the isobar is tangent to the plot.

2. The pressure has increased by a factor of 4/3.

3. $P = 7mg/S = 3.5 \cdot 10^5$ Pa.

4. $z = PvSN_A/(RT) \approx 6.10^{19}\ \text{s}^{-1}$.

Hint: Since the mean velocity of thermal molecular motion is far less than the orbital velocity of the satellite, the molecules may be considered to be motionless.

Shortest path

by Don Piele

WHILE THE ALGORITHMS OF MATHEMATICS go back thousands of years, those of Informatics are relatively new. The basic Informatics algorithms were not even considered by the mathematical masters and consequently are not part of our school training. Perhaps we need to update what we believe is fundamental in light of the ever increasing contributions that Informatics algorithms are making to our understanding of the world. As surely as e-commerce is reshaping the way companies are doing business, so to are the new Informatics algorithms reshaping the way scientists are going about their work.

So it is fitting to focus the column this month on one of the pioneers in Informatics algorithms, Edsger Dijkstra. Dijkstra is still active, holding the Schlumberger Centennial Chair in Computer Sciences at the University of Texas, Austin. Dijkstra discovered his famous shortest path algorithm in 1956 at the age of 26. At the time, programming was not officially recognized as a profession. In fact, when he applied for a marriage license in 1957, he put down "theoretical physicist" as his profession, believing that "programmer" would not be understood.

After he was assigned the task of showing off the power of the *ARMAC* computer housed in the Mathematical Center in Amsterdam, he came up with a program for constructing the shortest distance between two vertices on a graph. He used similar ideas to find an algorithm for finding a way to convey electricity to all essential circuits while using as little expensive copper wire as possible. This he called his "shortest subspanning tree algorithm."

Simply put

For a problem to be considered fundamental, it should be able to be simply put. Here is the shortest path problem simply put. Suppose you have a large map of all the cities in the United States connected by a vast network of roads. Along each road, denoted by a blue line, is a number that represents the distance between the two adjacent cities. Your task is to start from New York and find a path of minimal distance to Los Angeles. How would you do it?

What you have is a huge weighted graph where the vertices are the cities and the edges are the roads between cities. The edges are weighted by the distance between adjacent cities. Two cities are adjacent if there is a road directly between them. It turns out that finding a minimal length path from *NY* to *LA* can be done only after you know the shortest distance from *NY* to *LA*. So our first task is to examine Dijkstra's shortest distance algorithm.

Shortest distance

The basic idea is simple to explain. Start with *NY* and add the number of cities to which you know the shortest distance, one at a time, until *LA* joins the list. But how do you do this? First, look at all cities that are adjacent to *NY*. The current shortest distance from each city to *NY* is assigned to each adjacent city. But only the closest city, say *A*, is added to the list of cities whose shortest distance is known. Clearly none can be closer. Now scan all the cities adjacent to *A*. Update the current shortest distance value for each of these cities by comparing the current value with the distance you would get by going through *A*. In other words, if going to *A* and then to *NY* is better, use the short cut. Now select the closest city, say *B*, from the list of cities whose current shortest distance to *NY* is known. Add it to the list {*NY*, *A*} to make {*NY*, *A*, *B*} and continue looking for the next shortest distance city to *NY*. Once *LA* joins the list, you know the shortest distance to *LA*.

Pseudo code

Here is the pseudo code for finding the shortest distance.

$G = \{V, E\}$ or {Cities, Roads};

S = Set of cities whose shortest distance to *NY* is known;

$V–S$ = Set of cities whose shortest distance is not yet known for sure;

d = an array of best estimates of shortest distance from each city on the map to *NY*;

adj$M$ = distance between adjacent city matrix: adj$M[u, v]$ = distance along road directly connecting u to v.

1. Initialize d: $d = \{0, \infty, \infty, ..., \infty\}$.
2. Set $S = \{\ \}$.
3. While *LA* is not in S.
   i. Find u, the closest city in $V–S$ to S and move it from $V–S$ to S.
   ii. Update all cities v to see if going through u is better (i.e. a shortcut).
   $d[v] = \text{Min}[d[v], d[u]+ \text{adj}M[u, v]]$.

Mathematica implementation

Now let's bring this algorithm to life by creating a *Mathematica* animation that will grow the set S from one vertice to all vertices. Our demonstration is built with seven cities starting in *NY* and going to *LA*. We begin by drawing a simple weighted graph with seven vertices on the circumference of a circle, drawing roads between certain cities, and assigning a distance to each road. The assignment of distance is stored in the array *BetweenCityDistances*. Here each triplet $\{i, j, k\}$ means there is a road from city i to city j with length k.

```
n = 7;

BetweenCityDistances = {{1, 2, 5}, {1, 3, 2}, {1, 5, 7}, {2, 3, 1}, {2, 4, 5}, {3, 4, 8}, {3, 5, 10}, {4, 5, 2}, {4, 6, 10}, {5, 6, 2}, {2, 6, 6}, {1, 4, 12}, {6, 7, 5}, {5, 7, 20}};
```

We will use letters of the alphabet to label the cities when the pictures are drawn. Our substitution rule will be as follows:

```
alpha = {1 →"NY", 2 →"A", 3 →"B", 4 →"C", 5 →"D", 6 →"E", 7 →"LA"};
```

The distances and the road were easily picked off from the *BetweenCityDistances* array.

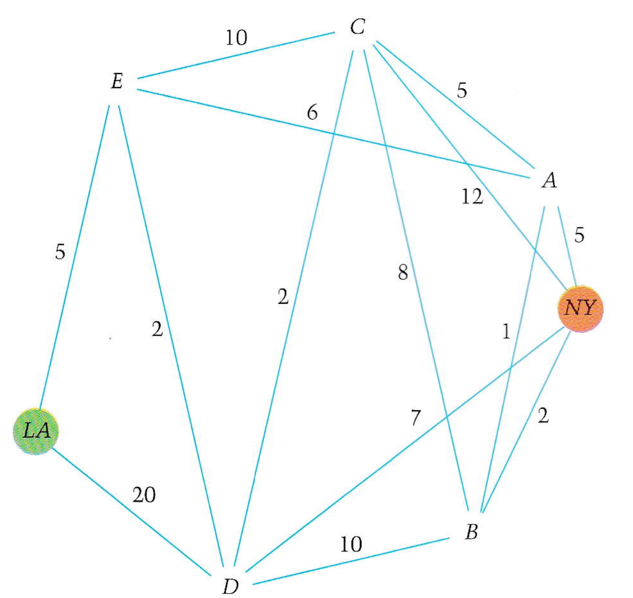
```
distances = BetweenCityDistances /. {i_, j_, k_} :> k

roads = BetweenCityDistances /. {i_, j_, k_} :> {i, j};

roads /. alpha

{5, 2, 7, 1, 5, 8, 10, 2, 10, 2, 6, 12, 5, 20}

{{NY, A}, {NY, B}, {NY, D}, {A, B}, {A, C}, {B, C}, {B, D}, {C, D}, {C, E}, {D, E}, {A, E}, {NY, C}, {E, LA}, {D, LA}}
```

Using the *BetweenCityDistances* array, an adjacency matrix adj$M$ is computed by placing ∞ between cities that are not connected and remembering that you can travel either way on a road. So the distance from vertex i to vertex j must be the same as the distance from vertex j to vertex i.

```
cities = {"NY", "A", "B", "C", "D", "E", "LA"};

adjM[i_, j_] := ∞;

adjM[i_, j_] := adjM[j, i] /; i > j;

adjM[i_, j_] := 0 /; i == j;

BetweenCityDistances /. {i_, j_, k_} :> (adjM[i, j] = k);

TableForm[Array[adjM, {n, n}], TableHeadings → {cities, cities}, TableSpacing → 1]
```

	NY	A	B	C	D	E	LA
NY	0	5	2	12	7	∞	∞
A	5	0	1	5	∞	6	∞
B	2	1	0	8	10	∞	∞
C	12	5	8	0	2	10	∞
D	7	∞	10	2	0	2	20
E	∞	6	∞	10	2	0	5
LA	∞	∞	∞	∞	20	5	0

A graphical representation for the cities and roads is

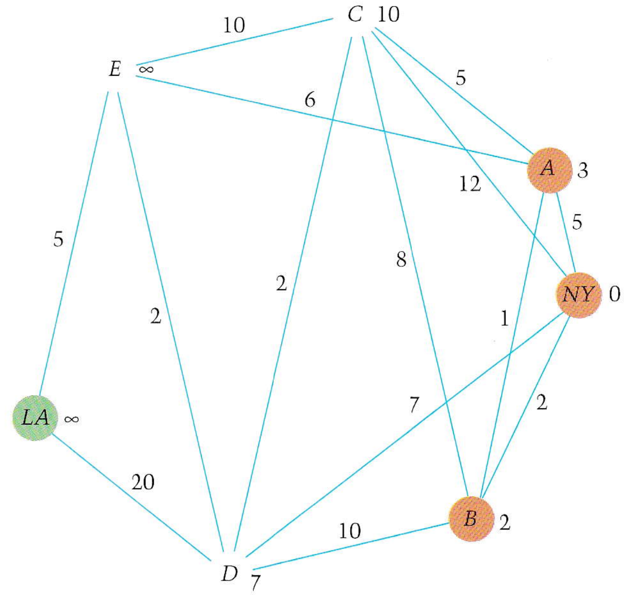


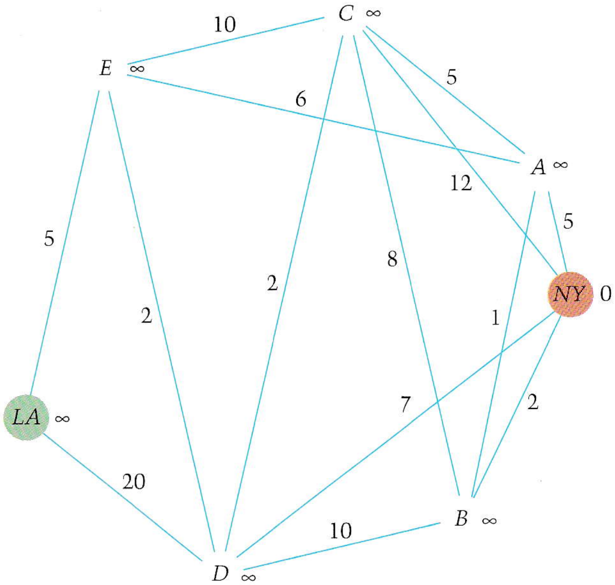
Dijkstra's algorithm in Mathematica

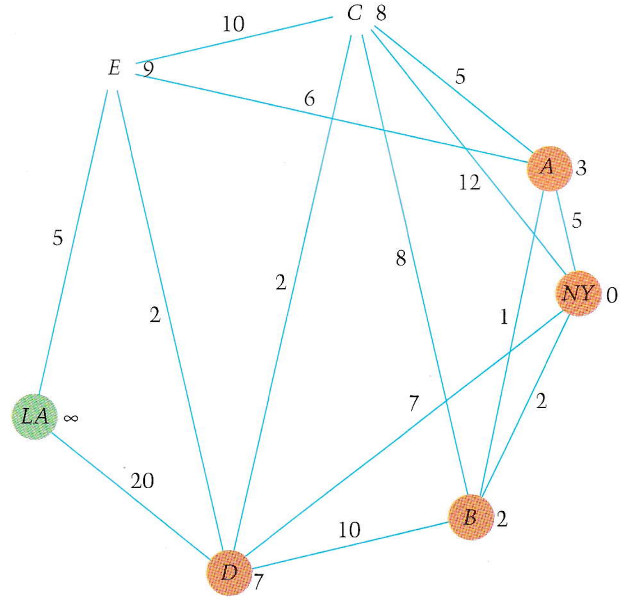
Using the pseudo code given above, the following program graphically illustrates the ever-expanding list of cities whose shortest distance to *NY* is known. Each city that is added to the list is colored red. Notice that cities are added to the list in order of increasing distance from *NY*. The current best estimate of the distance from a city to *NY* is indicated next to the city. This does not change once the city is colored red. The small numbers between cities is the distance between adjacent cities.

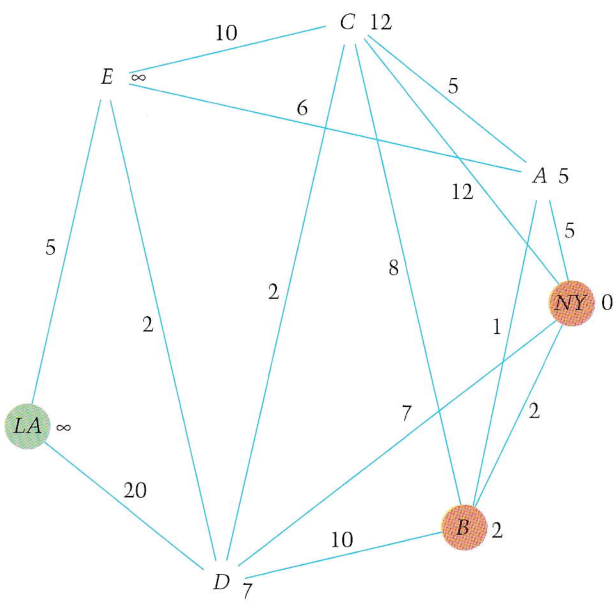
```
(* 1. Initialize d and Vertices V *)
d = Join[{0}, Table[∞, {n - 1}]];
```

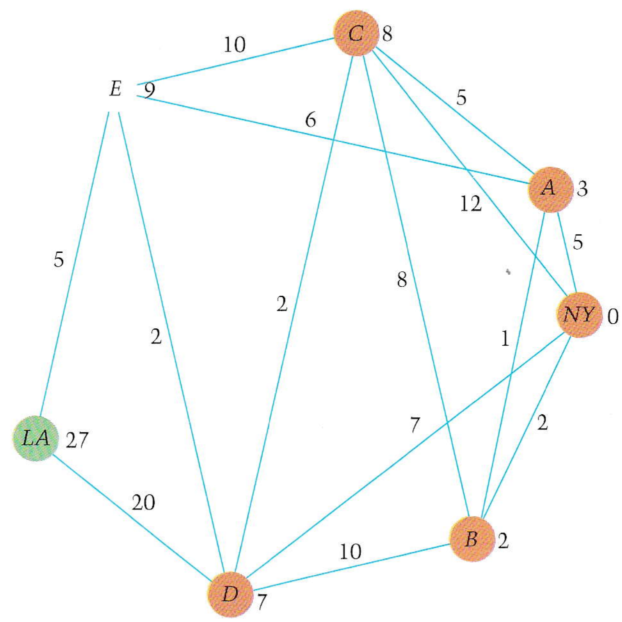
```
  V = Range[n];

  LA = n; (* LA is the last city *)
  (* define the update function *)

  update[v_] := d[[v]] = Min[d[[u]] +
adjM[u, v],d[[v]]];

  S = {};

  currentBest = d;

  While[! MemberQ[S, LA],
    u = V[[Position[currentBest,
Min[currentBest]][[1, 1]]]];
    S = Join[S, {u}];
    V = Complement[V, {u}];
      drawMap;
    currentBest = update /@ V
    ];
```

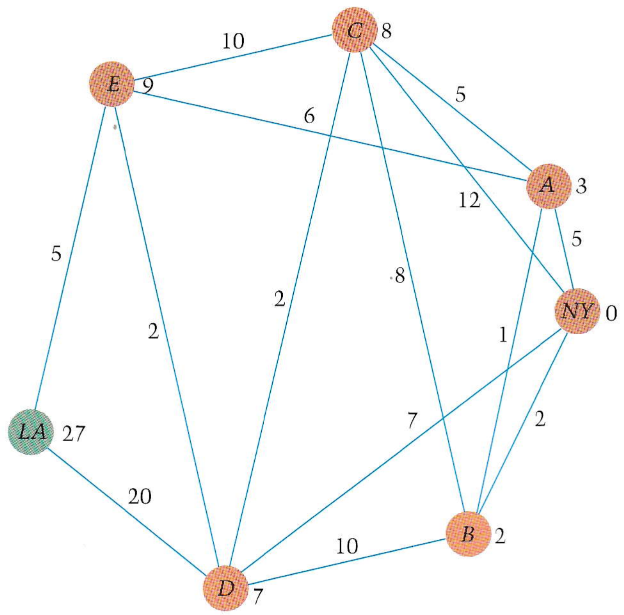


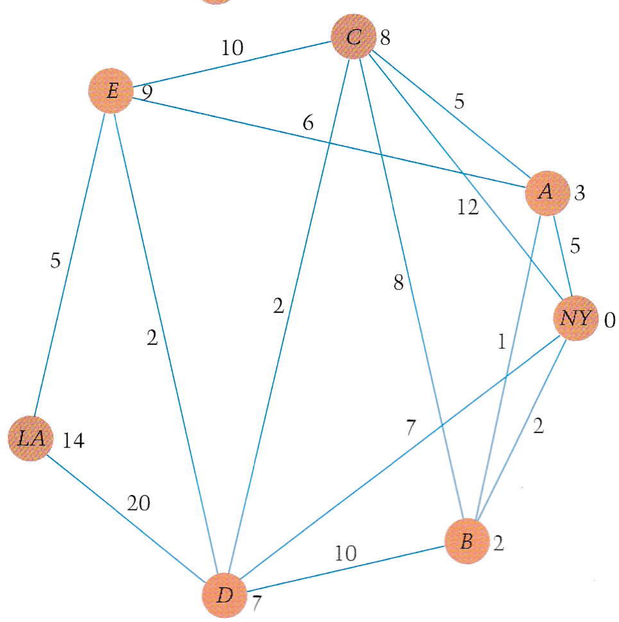












Your turn

We know the shortest distance from *NY* to all cities as far away as *LA*, but we do not have a path that actually has this distance. This problem has been left for you the reader. How would you modify the algorithm above so that a path—a list of cities from *NY* to *LA*—actually has a length equal to this shortest distance?

2000 USA Computing Olympiad

159 students from 34 countries took part in the USACO Winter Internet competition. In this programming competition, the problems are sent out via email to students who subscribe to the USACO mailing list at majordomo@delos.com. Participants were given one week to solve the problems and submit their solutions by email.

Four students from the United States were among the list of 13 students that had a perfect score on Senior Division problem. They were: David Arthur, Reid Barton, John Danaher, and Vladimir Novakovski.

Three Internet competitions are held each year in November, January, and March. The USACO National competition was held at high schools in the United States on April 12, 2000. To view the results, log on to our website at www.usaco.org.

Finally

Carla Laffra of Pace University has created a nice Java Applet that animates the solution to Dijkstra's shortest path algorithm. It is available on the Internet at http://www.mcs.csuhayward.edu/~morgan/notes_CS4590/Dijkstra_SPF_Applet/.

All solutions to the problems presented in this column are available at the Informatics website: http://www.uwp.edu/academic/mathematics/usaco/informatics/.

Notice that the *Mathematica* code for drawing the map (drawMap) is not included in this column. It can be seen by going to the Informatics website. If you would like, send me an email at piele@uwp.edu. ◘

Index of Advertisers