part of this solution and thereby obtain the solution (*) of the differential equation (30).[8]

21 Newton's Law of Gravitation and The Motion of the Planets

The inverse square law of attraction underlies so many natural phenomena—the orbits of the planets around the sun, the motion of the moon and artificial satellites about the earth, the paths described by charged particles in atomic physics, etc.—that every person educated in science ought to know something about its consequences. Our purpose in this section is to deduce Kepler's laws of planetary motion from Newton's law of universal gravitation, and to this end we discuss the motion of a small particle of mass m (a planet) under the attraction of a fixed large particle of mass M (the sun).

[8] The use of complex numbers in the mathematics of electric circuit problems was pioneered by the mathematician, inventor and electrical engineer Charles Proteus Steinmetz (1865–1923). As a young man in Germany, his student socialist activities got him into trouble with Bismarck's police, and he hastily emigrated to America in 1889. He was employed by the General Electric Company in its earliest period, and he quickly became the scientific brains of the Company and probably the greatest of all electrical engineers. When he came to GE there was no way to mass-produce electric motors or generators, and no economically viable way to transmit electric power more than 3 miles. Steinmetz solved these problems by using mathematics and the power of his own mind, and thereby improved human life forever in ways too numerous to count.

He was a dwarf who was crippled by a congenital deformity and lived with pain, but he was universally admired for his scientific genius and loved for his warm humanity and puckish sense of humor. The following little-known but unforgettable anecdote about him was published in the Letters section of *Life* magazine (May 14, 1965):

> Sirs: In your article on Steinmetz (April 23) you mentioned a consultation with Henry Ford. My father, Burt Scott, who was an employee of Henry Ford for many years, related to me the story behind that meeting. Technical troubles developed with a huge new generator at Ford's River Rouge plant. His electrical engineers were unable to locate the difficulty so Ford solicited the aid of Steinmetz. When "the little giant" arrived at the plant, he rejected all assistance, asking only for a notebook, pencil and cot. For two straight days and nights he listened to the generator and made countless computations. Then he asked for a ladder, a measuring tape and a piece of chalk. He laboriously ascended the ladder, made careful measurements, and put a chalk mark on the side of the generator. He descended and told his skeptical audience to remove a plate from the side of the generator and take out 16 windings from the field coil at that location. The corrections were made and the generator then functioned perfectly. Subsequently Ford received a bill for $10,000 signed by Steinmetz for G.E. Ford returned the bill acknowledging the good job done by Steinmetz but respectfully requesting an itemized statement. Steinmetz replied as follows: Making chalk mark on generator $1. Knowing where to make mark $9,999. Total due $10,000.

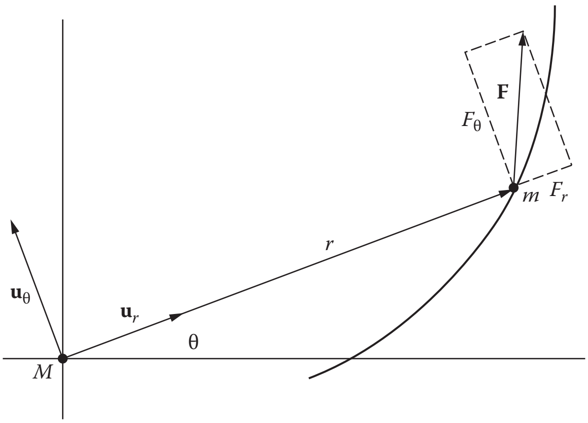

FIGURE 29

For problems involving a moving particle in which the force acting on it is always directed along the line from the particle to a fixed point, it is usually simplest to resolve the velocity, acceleration, and force into components along and perpendicular to this line. We therefore place the fixed particle M at the origin of a polar coordinate system (Figure 29) and express the radius vector from the origin to the moving particle m in the form

$$\mathbf{r} = r\mathbf{u}_r, \tag{1}$$

where $\mathbf{u}_r$ is the unit vector in the direction of $\mathbf{r}$.[9] It is clear that

$$\mathbf{u}_r = \mathbf{i}\cos\theta + \mathbf{j}\sin\theta, \tag{2}$$

and also that the corresponding unit vector $\mathbf{u}_\theta$, perpendicular to $\mathbf{u}_r$ in the direction of increasing θ, is given by

$$\mathbf{u}_\theta = -\mathbf{i}\sin\theta + \mathbf{j}\cos\theta. \tag{3}$$

The simple relations

$$\frac{d\mathbf{u}_r}{d\theta} = \mathbf{u}_\theta \quad \text{and} \quad \frac{d\mathbf{u}_\theta}{d\theta} = -\mathbf{u}_r,$$

obtained by differentiating (2) and (3), are essential for computing the velocity and acceleration vectors $\mathbf{v}$ and $\mathbf{a}$. Direct calculation from (1) now yields

[9] We here adopt the usual convention of signifying vectors by boldface type.

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = r\frac{d\mathbf{u}_r}{dt} + \mathbf{u}_r\frac{dr}{dt} = r\frac{d\mathbf{u}_r}{d\theta}\frac{d\theta}{dt} + \mathbf{u}_r\frac{dr}{dt} = r\frac{d\theta}{dt}\mathbf{u}_\theta + \frac{dr}{dt}\mathbf{u}_r \tag{4}$$

and

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \left(r\frac{d^2\theta}{dt^2} + 2\frac{dr}{dt}\frac{d\theta}{dt}\right)\mathbf{u}_\theta + \left[\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2\right]\mathbf{u}_r. \tag{5}$$

If the force $\mathbf{F}$ acting on m is written in the form

$$\mathbf{F} = F_\theta\mathbf{u}_\theta + F_r\mathbf{u}_r, \tag{6}$$

then from (5) and (6) and Newton's second law of motion $m\mathbf{a} = \mathbf{F}$, we get

$$m\left(r\frac{d^2\theta}{dt^2} + 2\frac{dr}{dt}\frac{d\theta}{dt}\right) = F_\theta \quad \text{and} \quad m\left[\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2\right] = F_r. \tag{7}$$

These differential equations govern the motion of the particle m, and are valid regardless of the nature of the force. Our next task is to extract information from them by making suitable assumptions about the direction and magnitude of $\mathbf{F}$.

Central forces and Kepler's Second Law. $\mathbf{F}$ is called a *central force* if it has no component perpendicular to $\mathbf{r}$, that is, if $F_\theta = 0$. Under this assumption the first of equations (7) becomes

$$r\frac{d^2\theta}{dt^2} + 2\frac{dr}{dt}\frac{d\theta}{dt} = 0.$$

On multiplying through by r, we obtain

$$r^2\frac{d^2\theta}{dt^2} + 2r\frac{dr}{dt}\frac{d\theta}{dt} = 0$$

or

$$\frac{d}{dt}\left(r^2\frac{d\theta}{dt}\right) = 0,$$

so

$$r^2\frac{d\theta}{dt} = h \tag{8}$$

for some constant h. We shall assume that h is positive, which evidently means that m is moving in a counterclockwise direction. If $A = A(t)$ is the area swept out by $\mathbf{r}$ from some fixed position of reference, so that $dA = r^2\, d\theta/2$, then (8) implies that

$$dA = \frac{1}{2}\left(r^2 \frac{d\theta}{dt}\right)dt = \frac{1}{2} h\, dt. \tag{9}$$

On integrating (9) from t_1 to t_2, we get

$$A(t_2) - A(t_1) = \frac{1}{2} h(t_2 - t_1). \tag{10}$$

This yields *Kepler's second law*: the radius vector $\mathbf{r}$ from the sun to a planet sweeps out equal areas in equal intervals of time.[10]

Central gravitational forces and Kepler's First Law. We now specialize even further, and assume that $\mathbf{F}$ is a central attractive force whose magnitude—according to Newton's law of gravitation—is directly proportional to the product of the two masses and inversely proportional to the square of the distance between them:

$$F_r = -G\frac{Mm}{r^2}. \tag{11}$$

The letter G represents the *gravitational constant*, which is one of the universal constants of nature. If we write (11) in the slightly simpler form

$$F_r = -\frac{km}{r^2},$$

where $k = GM$, then the second of equations (7) becomes

$$\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 = -\frac{k}{r^2}. \tag{12}$$

[10] When the Danish astronomer Tycho Brahe died in 1601, his assistant Johannes Kepler (1571–1630) inherited great masses of raw data on the positions of the planets at various times. Kepler worked incessantly on this material for 20 years, and at last succeeded in distilling from it his three beautifully simple laws of planetary motion—which were the climax of thousands of years of purely observational astronomy.

The next step in this line of thought is difficult to motivate, because it involves considerable technical ingenuity, but we will try. Our purpose is to use the differential equation (12) to obtain the equation of the orbit in the polar form $r=f(\theta)$, so we want to eliminate t from (12) and consider θ as the independent variable. Also, we want r to be the dependent variable, but if (8) is used to put (12) in the form

$$\frac{d^2r}{dt^2}-\frac{h^2}{r^3}=-\frac{k}{r^2}, \tag{13}$$

then the presence of powers of $1/r$ suggests that it might be temporarily convenient to introduce a new dependent variable $z=1/r$.

To accomplish these various aims, we must first express d^2r/dt^2 in terms of $d^2z/d\theta^2$, by calculating

$$\frac{dr}{dt}=\frac{d}{dt}\left(\frac{1}{z}\right)=-\frac{1}{z^2}\frac{dz}{dt}=-\frac{1}{z^2}\frac{dz}{d\theta}\frac{d\theta}{dt}=-\frac{1}{z^2}\frac{dz}{d\theta}\frac{h}{r^2}=-h\frac{dz}{d\theta}$$

and

$$\frac{d^2r}{dt^2}=-h\frac{d}{dt}\left(\frac{dz}{d\theta}\right)=-h\frac{d}{d\theta}\left(\frac{dz}{d\theta}\right)\frac{d\theta}{dt}=-h\frac{d^2z}{d\theta^2}\frac{h}{r^2}=-h^2z^2\frac{d^2z}{d\theta^2}.$$

When the latter expression is inserted in (13) and $1/r$ is replaced by z, we get

$$-h^2z^2\frac{d^2z}{d\theta^2}-h^2z^3=-kz^2$$

or

$$\frac{d^2z}{d\theta^2}+z=\frac{k}{h^2}.$$

The general solution of this equation can be written down at once:

$$z=A\sin\theta+B\cos\theta+\frac{k}{h^2}. \tag{14}$$

For the sake of simplicity, we shift the direction of the polar axis in such a way that r is minimal (that is, m is closest to the origin) when $\theta=0$. This means that z is to be maximal in this direction, so

$$\frac{dz}{d\theta} = 0 \quad \text{and} \quad \frac{d^2z}{d\theta^2} < 0$$

when $\theta = 0$. These conditions imply that $A = 0$ and $B > 0$. If we now replace z by $1/r$, then (14) can be written

$$r = \frac{1}{k/h^2 + B\cos\theta} = \frac{h^2/k}{1 + (Bh^2/k)\cos\theta};$$

and if we put $e = Bh^2/k$, then our equation for the orbit becomes

$$r = \frac{h^2/k}{1 + e\cos\theta}, \tag{15}$$

where e is a positive constant.

At this point we recall (Figure 30) that the locus defined by $PF/PD = e$ is the conic section with focus F, directrix d, and eccentricity e. When this condition is expressed in terms of r and 0, it is easy to see that

$$r = \frac{pe}{1 + e\cos\theta}$$

is the polar equation of our conic section, which is an ellipse, a parabola, or a hyperbola according as $e < 1$, $e = 1$, or $e > 1$. These remarks show that the orbit (15) is a conic section with eccentricity $e = Bh^2/k$; and since the planets remain in the solar system and do not move infinitely far away from the sun, the ellipse is the only possibility. This yields *Kepler's first law*: the orbit of each planet is an ellipse with the sun at one focus.

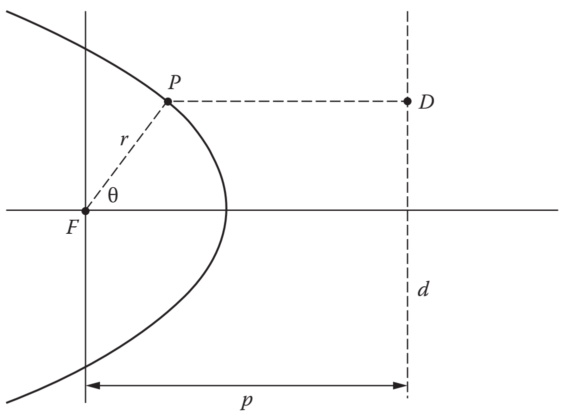

FIGURE 30

The physical meaning of the eccentricity. It follows from equation (4) that the kinetic energy of m is

$$\frac{1}{2}mv^2 = \frac{1}{2}m\left[r^2\left(\frac{d\theta}{dt}\right)^2 + \left(\frac{dr}{dt}\right)^2\right]. \tag{16}$$

The potential energy of the system is the negative of the work required to move m to infinity (where the potential energy is zero), and is therefore

$$-\int_r^\infty \frac{km}{r^2}\,dr = \frac{km}{r}\bigg|_r^\infty = -\frac{km}{r}. \tag{17}$$

If E is the total energy of the system, which is constant by the principle of conservation of energy, then (16) and (17) yield

$$\frac{1}{2}m\left[r^2\left(\frac{d\theta}{dt}\right)^2 + \left(\frac{dr}{dt}\right)^2\right] - \frac{km}{r} = E. \tag{18}$$

At the instant when $\theta = 0$, (15) and (18) give

$$r = \frac{h^2/k}{1+e} \qquad \text{and} \qquad \frac{mr^2}{2}\frac{h^2}{r^4} - \frac{km}{r} = E.$$

It is easy to eliminate r from these equations; and when the result is solved for e, we find that

$$e = \sqrt{1 + E\left(\frac{2h^2}{mk^2}\right)}.$$

This enables us to write equation (15) for the orbit in the form

$$r = \frac{h^2/k}{1 + \sqrt{1 + E(2h^2/mk^2)}\cos\theta}. \tag{19}$$

It is evident from (19) that the orbit is an ellipse, a parabola, or a hyperbola according as $E < 0$, $E = 0$, or $E > 0$. It is therefore clear that the nature of the orbit of m is completely determined by its total energy E. Thus the planets in the solar system have negative energies and move in ellipses, and bodies passing through the solar system at high speeds have positive energies and travel along hyperbolic paths. It is interesting to realize that if a planet like the earth could be given a push from behind, sufficiently strong to speed it up and lift its total energy above zero, it would enter into a hyperbolic orbit and leave the solar system permanently.

The periods of revolution of the planets and Kepler's Third Law. We now restrict our attention to the case in which m has an elliptic orbit (Figure 31) whose polar and rectangular equations are (15) and

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1.$$

It is well known from elementary analytic geometry that $e=c/a$ and $c^2=a^2-b^2$, so $e^2=(a^2-b^2)/a^2$ and

$$b^2=a^2(1-e^2). \tag{20}$$

In astronomy the semimajor axis of the orbit is called the *mean distance*, because it is one-half the sum of the least and greatest values of r, so (15) and (20) give

$$a=\frac{1}{2}\left(\frac{h^2/k}{1+e}+\frac{h^2/k}{1-e}\right)=\frac{h^2}{k(1-e^2)}=\frac{h^2a^2}{kb^2},$$

and we have

$$b^2=\frac{h^2a}{k}. \tag{21}$$

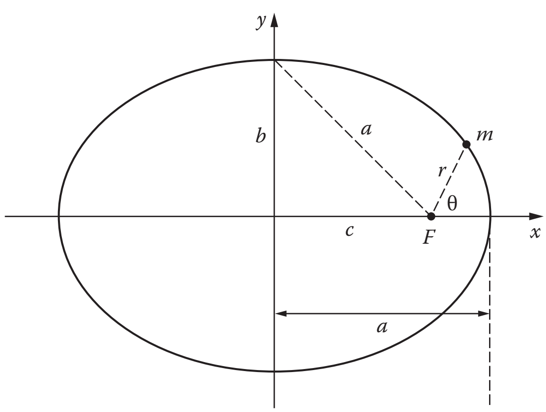

FIGURE 31

If T is the period of m (that is, the time required for one complete revolution in its orbit), then since the area of the ellipse is πab it follows from (10) that $\pi ab = hT/2$. In view of (21), this yields

$$T^2 = \frac{4\pi^2 a^2 b^2}{h^2} = \left(\frac{4\pi^2}{k}\right) a^3. \tag{22}$$

In the present idealized treatment, the constant $k = GM$ depends on the central mass M but not on m, so (22) holds for all the planets in our solar system and we have *Keplers' third law*: the squares of the periods of revolution of the planets are proportional to the cubes of their mean distances.

The ideas of this section are of course due primarily to Newton (Appendix B). However, the arguments given here are quite different from those that were used in print by Newton himself, for he made no explicit use of the methods of calculus in any of his published works on physics or astronomy. For him calculus was a private method of scientific investigation unknown to his contemporaries, and he had to rewrite his discoveries into the language of classical geometry whenever he wished to communicate them to others.

Problems

1. In practical work with Kepler's third law (22), it is customary to measure T in years and a in astronomical units (1 astronomical unit = the earth's mean distance $\cong$ 93,000,000 miles $\cong$ 150,000,000 kilometers). With these convenient units of measurement, (22) takes the simpler form $T^2 = a^3$. What is the period of revolution T of a planet whose mean distance from the sun is

(a) twice that of the earth?

(b) three times that of the earth?

(c) twenty-five times that of the earth?

2. (a) Mercury's "year" is 88 days. What is Mercury's mean distance from the sun?

(b) The mean distance of the planet Saturn is 9.54 astronomical units. What is Saturn's period of revolution about the sun?

3. Kepler's first two laws, in the form of equations (8) and (15), imply that m is attracted toward the origin with a force whose magnitude is inversely proportional to the square of r. This was Newton's fundamental discovery, for it caused him to propound his law of gravitation and investigate its consequences. Prove this by assuming (8) and (15) and verifying the following statements:

(a) $F_\theta = 0$;

(b) $\dfrac{dr}{dt} = \dfrac{ke}{h}\sin\theta$;

(c) $\dfrac{d^2r}{dt^2} = \dfrac{ke\cos\theta}{r^2}$;

(d) $F_r = -\dfrac{mk}{r^2} = -G\dfrac{Mm}{r^2}$.

4. Show that the speed v of a planet at any point of its orbit is given by

$$v^2 = k\left(\frac{2}{r} - \frac{1}{a}\right).$$

5. Suppose that the earth explodes into fragments which fly off at the same speed in different directions into orbits of their own. Use Kepler's third law and the result of Problem 4 to show that all fragments that do not fall into the sun or escape from the solar system will reunite later at the same point where they began to diverge.

22 Higher Order Linear Equations. Coupled Harmonic Oscillators

Even though the main topic of this chapter is second order linear equations, there are several aspects of higher order linear equations that make it worthwhile to discuss them briefly.

Most of the ideas and methods described in Sections 14 to 19 are easily extended to nth order linear equations with constant coefficients,

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1}\, y' + a_n y = f(x), \tag{1}$$

where $f(x)$ is assumed to be continuous on an interval $[a,b]$. The basic fact to keep in mind is that the general solution of (1) has the form we expect,

$$y(x) = y_g(x) + y_p(x),$$

where $y_p(x)$ is any particular solution of (1) and $y_g(x)$ is the general solution of the reduced homogeneous equation

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y' + a_n y = 0. \tag{2}$$

The proof is exactly the same as the proof for the case $n = 2$, and will not be repeated.

We begin by considering the problem of finding the general solution of the homogeneous equation (2). Our experience with the case $n = 2$ tells us that this equation probably has solutions of the form $y = e^{rx}$ for suitable values of the constant r. By substituting $y = e^{rx}$ and its derivatives into (2) and dividing out the nonzero factor e^{rx}, we obtain the *auxiliary equation*

$$r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n = 0. \tag{3}$$

The polynomial on the left side of (3) is called the *auxiliary polynomial*; in principle it can always be factored completely into a product of n linear factors, and equation (3) can then be written in the factored form

$$(r - r_1)(r - r_2) \cdots (r - r_n) = 0.$$

The constants $r_1, r_2, \ldots, r_n$ are the roots of the auxiliary equation (3). If these roots are distinct from one another, then we have n distinct solutions

$$e^{r_1 x}, e^{r_2 x}, \cdots, e^{r_n x} \tag{4}$$

of the homogeneous equation (2). Just as in the case $n = 2$, the linear combination

$$y(x) = c_1 e^{r_1 x} + c_2 e^{r_2 x} + \cdots + c_n e^{r_n x} \tag{5}$$

is also a solution for every choice of the coefficients $c_1, c_2, \ldots, c_n$.

Since (5) contains n arbitrary constants, we have reasonable grounds for hoping that it is the general solution of the nth order equation (2). To elevate this hope into a certainty, we must appeal to a small body of theory that we now sketch very briefly.

When the theorems of Sections 14 and 15 are extended in the natural way, it can be proved that (5) is the general solution of (2) if the solutions (4) are linearly independent.[11] There are several ways of establishing the fact that the solutions (4) are linearly independent whenever the roots $r_1, r_2, \ldots, r_n$ are

[11] This requires establishing the same connections as before among (1) satisfying n initial conditions, (2) the nonvanishing of the Wronskian, (3) Abel's formula, and (4) linear independence. A set of n functions $y_1(x), y_2(x), \ldots, y_n(x)$ is said to be *linearly dependent* if one of them can be expressed as a linear combination of the others, and *linearly independent* if this is not possible. In specific cases this is usually easy to decide by inspection. Equivalently, linear dependence means that there exists a relation of the form $c_1 y_1(x) + c_2 y_2(x) + \cdots + c_n y_n(x) = 0$ in which at least one of the c's is not zero, and linear independence means that any such relation implies that all the c's must be zero.

distinct, but we omit the details. It therefore follows that (5) actually is the general solution of (2) in this case.

Repeated real roots. If the real roots of (3) are not all distinct, then the solutions (4) are linearly dependent and (5) is not the general solution. For example, if $r_1 = r_2$ then the part of (5) consisting of $c_1 e^{r_1 x} + c_2 e^{r_2 x}$ becomes $(c_1 + c_2)e^{r_1 x}$, and the two constants c_1 and c_2 become one constant $c_1 + c_2$. To see what to do when this happens, we recall that in the special case of the second order equation, where we had only the two roots r_1 and r_2, we found that when $r_1 = r_2$ the solution $c_1 e^{r_1 x} + c_2 e^{r_2 x}$ had to be replaced by $c_1 e^{r_1 x} + c_2 x e^{r_1 x} = (c_1 + c_2 x)e^{r_1 x}$. It can be verified by substitution that if $r_1 = r_2$ for the nth order equation (2), then the first two terms of (5) must be replaced by this same expression.

More generally, if $r_1 = r_2 = \cdots = r_k$ is a real root of multiplicity k (that is, a k-fold repeated root) of the auxiliary equation (3), then the first k terms in the solution (5) must be replaced by

$$(c_1 + c_2 x + c_3 x^2 + \cdots + c_k x^{k-1})e^{r_1 x}.$$

A similar family of solutions is needed for each multiple real root, giving a correspondingly modified form of (5). In the next section we will show how to obtain these expressions by operator methods.

Complex roots. Some of the roots of the auxiliary equation (3) may be complex numbers. Since the coefficients of (3) are real, all complex roots occur in conjugate complex pairs $a + ib$ and $a - ib$. As in the case $n = 2$, the part of the solution (5) corresponding to two such roots can be written in the alternative real form

$$e^{ax}(A \cos bx + B \sin bx).$$

If $a + ib$ and $a - ib$ are roots of multiplicity k, then we must take

$$\begin{gathered} e^{ax}[(A_1 + A_2 x + \cdots + A_k x^{k-1})\cos bx \\ + (B_1 + B_2 x + \cdots + B_k x^{k-1})\sin bx] \end{gathered}$$

as part of the general solution.

Example 1. The differential equation

$$y^{(4)} - 5y'' + 4y = 0$$

has auxiliary equation

$$r^4 - 5r^2 + 4 = (r^2 - 1)(r^2 - 4) = (r - 1)(r + 1)(r - 2)(r + 2) = 0.$$

Its general solution is therefore

$$y = c_1 e^x + c_2 e^{-x} + c_3 e^{2x} + c_4 e^{-2x}.$$

Example 2. The equation

$$y^{(4)} - 8y'' + 16y = 0$$

has auxiliary equation

$$r^4 - 8r^2 + 16 = (r^2 - 4)^2 = (r - 2)^2(r + 2)^2 = 0,$$

so the general solution is

$$y = (c_1 + c_2 x)e^{2x} + (c_3 + c_4 x)e^{-2x}.$$

Example 3. The equation

$$y^{(4)} - 2y''' + 2y'' - 2y' + y = 0$$

has auxiliary equation

$$r^4 - 2r^3 + 2r^2 - 2r + 1 = 0,$$

or, after factoring,[12]

$$(r - 1)^2(r^2 + 1) = 0.$$

The general solution is therefore

$$y = (c_1 + c_2 x)e^x + c_3 \cos x + c_4 \sin x.$$

Example 4. **Coupled harmonic oscillators**. Linear equations of order $n > 2$ arise most often in physics by eliminating variables from simultaneous systems of second order equations. We can see an example of this by linking together two simple harmonic oscillators of the kind discussed at the beginning of Section 20. Accordingly, let two carts of masses m_1 and m_2 be attached to the left and right walls in Figure 32 by springs with spring constants k_1 and k_2. If there is no damping and these carts are left unconnected, then when disturbed each moves with its own simple harmonic motion, that is, we have two independent harmonic oscillators. We obtain *coupled harmonic oscillators* if we now connect the carts to each other by a spring with spring constant k_3, as indicated in the figure. By applying Newton's second law of motion, it can be shown (Problem 16)

[12] To factor the auxiliary equation, notice that $r = 1$ is a root that can be found by inspection, so $r - 1$ is a factor of the auxiliary polynomial and the other factor can be found by long division.

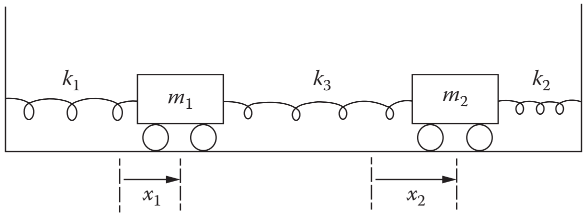

FIGURE 32

that the displacements x_1 and x_2 of the carts satisfy the following simultaneous system of second order linear equations:

$$\begin{aligned} m_1\frac{d^2x_1}{dt^2} &= -k_1x_1 + k_3(x_2 - x_1) \\ &= -(k_1 + k_3)x_1 + k_3x_2, \\ m_2\frac{d^2x_2}{dt^2} &= -k_3(x_2 - x_1) - k_2x_2 \\ &= k_3x_1 - (k_2 + k_3)x_2. \end{aligned} \tag{6}$$

We can now obtain a single fourth order equation for x_1 by solving the first equation for x_2 and substituting in the second equation (Problem 17).

We have not yet addressed the problem of finding a particular solution for the complete equation (1). In this context it suffices to remark that the method of undetermined coefficients discussed in Section 18 continues to apply, with obvious minor changes, for functions $f(x)$ of the types considered in that section. In the next section we shall examine a totally different approach to the problem of finding particular solutions.

Example 5. Find a particular solution of the differential equation $y''' + 2y'' - y' = 3x^2 - 2x + 1$.

Our experience in Section 18 suggests that we take a trial solution of the form

$$\begin{aligned} y &= x(a_0 + a_1x + a_2x^2) \\ &= a_0x + a_1x^2 + a_2x^3. \end{aligned}$$

Since $y' = a_0 + 2a_1x + 3a_2x^2$, $y'' = 2a_1 + 6a_2x$, and $y''' = 6a_2$, substitution in the given equation yields

$$6a_2 + 2(2a_1 + 6a_2x) - (a_0 + 2a_1x + 3a_2x^2) = 3x^2 - 2x + 1$$

or, after collecting coefficients of like powers of x,

$$-3a_2x^2 + (-2a_1 + 12a_2)x + (-a_0 + 4a_1 + 6a_2) = 3x^2 - 2x + 1.$$

Thus,

$$-3a_2 = 3,$$

$$-2a_1 + 12a_2 = -2,$$

$$-a_0 + 4a_1 + 6a_2 = 1,$$

so $a_2 = -1$, $a_1 = -5$, and $a_0 = -27$. We therefore have a particular solution $y = -27x - 5x^2 - x^3$.

Problems

Find the general solution of each of the following equations.

1. $y''' - 3y'' + 2y' = 0.$
2. $y''' - 3y'' + 4y' - 2y = 0.$
3. $y''' - y = 0.$
4. $y''' + y = 0.$
5. $y''' + 3y'' + 3y' + y = 0.$
6. $y^{(4)} + 4y''' + 6y'' + 4y' + y = 0.$
7. $y^{(4)} - y = 0.$
8. $y^{(4)} + 5y'' + 4y = 0.$
9. $y^{(4)} - 2a^2y'' + a^4y = 0.$
10. $y^{(4)} + 2a^2y'' + a^4y = 0.$
11. $y^{(4)} + 2y''' + 2y'' + 2y' + y = 0.$
12. $y^{(4)} + 2y''' - 2y'' - 6y' + 5y = 0.$
13. $y''' - 6y'' + 11y' - 6y = 0.$
14. $y^{(4)} + y''' - 3y'' - 5y' - 2y = 0.$
15. $y^{(5)} - 6y^{(4)} - 8y''' + 48y'' + 16y' - 96y = 0.$
16. Derive equations (6) for the coupled harmonic oscillators by using the configuration shown in Figure 32, where both carts are displaced to the right from their equilibrium positions and $x_2 > x_1$, so that the spring on the right is compressed and the other two are stretched.
17. In Example 4, find the fourth order differential equation for x_1 by eliminating x_2 as suggested.
18. In the preceding problem, solve the fourth order equation for x_1 if the masses are equal and the spring constants are equal, so that $m_1 = m_2 = m$ and $k_1 = k_2 = k_3 = k$. In this special case, show directly (that is, without using the symmetry of the situation) that x_2 satisfies the same

differential equation as x_1. The two frequencies associated with these coupled harmonic oscillators are called the *normal frequencies* of the system. What are they?

19. Find the general solution of $y^{(4)}=0$. Of $y^{(4)}=\sin x+24$.

20. Find the general solution of $y''' - 3y'' + 2y' = 10 + 42e^{3x}$.

21. Find the solution of $y''' - y' = 1$ that satisfies the initial conditions $y(0)=y'(0)=y''(0)=4$.

22. Show that the change of independent variable $x=e^z$ transforms the third order Euler equidimensional equation

$$x^3y''' + a_1x^2y'' + a_2xy' + a_3y = 0$$

into a third order linear equation with constant coefficients. (This transformation also works for the nth order Euler equation.) Solve the following equations by this method:

(a) $x^3y''' + 3x^2y'' = 0$;

(b) $x^3y''' + x^2y'' - 2xy' + 2y = 0$;

(c) $x^3y''' + 2x^2y'' + xy' - y = 0$.

23. In determining the drag on a small sphere moving at a constant speed through a viscous fluid, it is necessary to solve the differential equation

$$x^3y^{(4)} + 8x^2y''' + 8xy'' - 8y' = 0.$$

Notice that this is an Euler equation for y' and use the method of Problem 22 to show that the general solution is

$$y = c_1x^2 + c_2x^{-1} + c_3x^{-3} + c_4.$$

These ideas are part of the mathematical background used by Robert A. Millikan in his famous oil-drop experiment of 1909 for measuring the charge on an electron, for which he won the 1923 Nobel Prize.[13]

23 Operator Methods for Finding Particular Solutions

At the end of Section 22 we referred to the problem of finding particular solutions for nonhomogeneous equations of the form

$$\frac{d^ny}{dx^n} + a_1\frac{d^{n-1}y}{dx^{n-1}} + \cdots + a_{n-1}\frac{dy}{dx} + a_ny = f(x). \tag{1}$$

[13] For a clear explanation of this exceedingly ingenious experiment, with a good drawing of the apparatus, see pp. 50–51 of the book by Linus Pauling mentioned in Section 4 [Note 12].

In this section we give a very brief sketch of the use of differential operators for solving this problem in more efficient ways than any we have seen before. These "operational methods" are mainly due to the English applied mathematician Oliver Heaviside (1850–1925). Heaviside's methods seemed so strange to the scientists of his time that he was widely regarded as a crackpot, which unfortunately is a common fate for thinkers of unusual originality.

Let us represent derivatives by powers of D, so that

$$Dy = \frac{dy}{dx}, D^2y = \frac{d^2y}{dx^2}, \ldots, D^ny = \frac{d^ny}{dx^n}.$$

Then (1) can be written as

$$D^ny + a_1D^{n-1}y + \cdots + a_{n-1}Dy + a_ny = f(x), \tag{2}$$

or as

$$(D^n + a_1D^{n-1} + \cdots + a_{n-1}D + a_n)y = f(x),$$

or as

$$p(D)y = f(x), \tag{3}$$

where the differential operator $p(D)$ is simply the auxiliary polynomial $p(r)$ with r replaced by D. The successive application of two or more such operators can be made by first multiplying the operators together by the usual rules of algebra and then applying the product operator. For example, we know that $p(D)$ can be formally factored into

$$p(D) = (D - r_1)(D - r_2)\cdots(D - r_n), \tag{4}$$

where $r_1, r_2, \ldots, r_n$ are the roots of the auxiliary equation; and these factors can then be applied successively in any order to yield the same result as a single application of $p(D)$. As an illustration of this idea, we point out that if the auxiliary equation is of the second decree and therefore has only two roots r_1 and r_2, then formally we have

$$(D - r_1)(D - r_2) = D^2 - (r_1 + r_2)D + r_1r_2; \tag{5}$$

and since

$$(D - r_2)y = \left(\frac{d}{dx} - r_2\right)y = \frac{dy}{dx} - r_2y,$$

we can verify (5) by writing

$$\begin{aligned}(D-r_1)(D-r_2)y &= \left(\frac{d}{dx}-r_1\right)\left(\frac{dy}{dx}-r_2y\right)\\ &= \frac{d}{dx}\left(\frac{dy}{dx}-r_2y\right)-r_1\left(\frac{dy}{dx}-r_2y\right)\\ &= \frac{d^2y}{dx^2}-(r_1+r_2)\frac{dy}{dx}+r_1r_2y\\ &= D^2y-(r_1+r_2)Dy+r_1r_2y\\ &= [D^2-(r_1+r_2)D+r_1r_2]y,\end{aligned}$$

for this is the meaning of (5).

We have no difficulty with the meaning of the expression $p(D)y$ on the left of (3): it has the same meaning as the left side of (2) or (1). Our purpose now is to learn how to treat $p(D)$ as a separate entity, and in doing this to develop the methods for solving (3) that are the subject of this section. Without beating around the bush, we wish to "solve formally" for y in (3), obtaining

$$y = \frac{1}{p(D)}f(x). \tag{6}$$

Here $1/p(D)$ represents an operation to be performed on $f(x)$ to yield y. The question is, what is the nature of this operation, and how can we carry it out? In order to begin to understand these matters, we consider the simple equation $Dy = f(x)$, which gives

$$y = \frac{1}{D}f(x).$$

But $Dy = f(x)$, or equivalently $dy/dx = f(x)$, is easily solved by writing

$$y = \int f(x)dx,$$

so it is natural to make the definition

$$\frac{1}{D}f(x) = \int f(x)dx. \tag{7}$$

This tells us that the operator $1/D$ applied to a function means integrate the function. Similarly, the operator $1/D^2$ applied to a function means integrate the function twice in succession, and so on. Operators like $1/D$ and $1/D^2$

are called *inverse operators*. We continue this line of investigation and examine other inverse operators. Consider

$$(D - r)y = f(x), \tag{8}$$

where r is a constant. Formally, we have

$$y = \frac{1}{D-r} f(x).$$

But (8) is the simple first order linear equation

$$\frac{dy}{dx} - ry = f(x),$$

whose solution by Section 10 is

$$y = e^{rx} \int e^{-rx} f(x)\, dx.$$

(We suppress constants of integration because we are only seeking particular solutions.) It is therefore natural to make the definition

$$\frac{1}{D-r} f(x) = e^{rx} \int e^{-rx} f(x)\, dx. \tag{9}$$

Notice that this reduces to (7) when $r = 0$. We are now ready to begin carrying out the problem-solving procedures that arise from (6).

METHOD 1: SUCCESSIVE INTEGRATIONS. By using the factorization (4), we can write formula (6) as

$$\begin{aligned} y &= \frac{1}{p(D)} f(x) = \frac{1}{(D-r_1)(D-r_2)\cdots(D-r_n)} f(x) \\ &= \frac{1}{D-r_1} \frac{1}{D-r_2} \cdots \frac{1}{D-r_n} f(x). \end{aligned}$$

Here we may apply the n inverse operators in any convenient order, and by (9) we know that the complete process requires n successive integrations. That the resulting function $y = y(x)$ is a particular solution of (3) is easily seen; for by applying to y the factors of $p(D)$ in suitable order, we undo the successive integrations and arrive back at $f(x)$.

Example 1. Find a particular solution of $y'' - 3y' + 2y = xe^x$.

Solution. We have $(D^2 - 3D + 2)y = xe^x$, so

$$(D-1)(D-2)y = xe^x \quad \text{and} \quad y = \frac{1}{D-1}\frac{1}{D-2}xe^x.$$

By (9) and an integration by parts, we obtain

$$\frac{1}{D-2}xe^x = e^{2x}\int e^{-2x}xe^x\,dx = -(1+x)e^x,$$

so

$$y = \frac{1}{D-1}\left[-(1+x)e^x\right] = -e^x\int e^{-x}(1+x)e^x\,dx = -\frac{1}{2}(1+x)^2e^x.$$

Example 2. Find a particular solution of $y'' - y = e^{-x}$.

Solution. We have $(D^2 - 1)y = e^{-x}$, so

$$(D-1)(D+1)y = e^{-x}, \qquad y = \frac{1}{D-1}\frac{1}{D+1}e^{-x},$$

$$\frac{1}{D+1}e^{-x} = e^{-x}\int e^xe^{-x}\,dx = xe^{-x},$$

$$y = \frac{1}{D-1}xe^{-x} = e^x\int e^{-x}xe^{-x}\,dx = \left(-\frac{1}{2}x - \frac{1}{4}\right)e^{-x}.$$

METHOD 2: PARTIAL FRACTIONS DECOMPOSITIONS OF OPERATORS. The successive integrations of method 1 are likely to become complicated and time-consuming to carry out. The formula

$$y = \frac{1}{p(D)}f(x) = \frac{1}{(D-r_1)(D-r_2)\cdots(D-r_n)}f(x)$$

suggests a way to avoid this work, for it suggests the possibility of decomposing the operator on the right into partial fractions. If the factors of $p(D)$ are distinct, we can write

$$y = \frac{1}{p(D)}f(x) = \left[\frac{A_1}{D-r_1} + \frac{A_2}{D-r_2} + \cdots + \frac{A_n}{D-r_n}\right]f(x)$$

for suitable constants $A_1, A_2, \ldots, A_n$, and each term on the right can be found by using (9). The operator in brackets here is sometimes called the *Heaviside expansion* of the inverse operator $1/p(D)$.

Example 3. Solve the problem in Example 1 by this method.

Solution. We have

$$\begin{aligned} y &= \frac{1}{(D-1)(D-2)} xe^x = \left[\frac{1}{D-2} - \frac{1}{D-1}\right] xe^x \\ &= \frac{1}{D-2} xe^x - \frac{1}{D-1} xe^x \\ &= e^{2x}\int e^{-2x} xe^x dx - e^x \int e^{-x} xe^x dx \\ &= -1(1+x)e^x - \frac{1}{2}x^2 e^x = -(1 + x + \frac{1}{2}x^2)e^x. \end{aligned}$$

The student will notice that this solution is not quite the same as the solution found in Example 1. However, it is easy to see that they differ only by a solution of the reduced homogeneous equation, so all is well.

Example 4. Solve the problem in Example 2 by this method.

Solution. We have

$$\begin{aligned} y &= \frac{1}{(D-1)(D+1)} e^{-x} = \frac{1}{2}\left[\frac{1}{D-1} - \frac{1}{D+1}\right] e^{-x} \\ &= \frac{1}{2}e^x \int e^{-x}e^{-x} dx - \frac{1}{2}e^{-x}\int e^x e^{-x} dx \\ &= -\frac{1}{4}e^{-x} - \frac{1}{2}xe^{-x}. \end{aligned}$$

If some of the factors of $p(D)$ are repeated, then we know that the form of the partial fractions decomposition is different. For example if $D - r_1$ is a k-fold repeated factor, then the decomposition contains the terms

$$\frac{A_1}{D-r_1} + \frac{A_2}{(D-r_1)^2} + \cdots + \frac{A_k}{(D-r_1)^k}.$$

These operators can be applied to $f(x)$ in order from left to right, each requiring an integration based on the result of the preceding step, as in method 1.

METHOD 3: SERIES EXPANSIONS OF OPERATORS. For problems in which $f(x)$ is a polynomial, it is often useful to expand the inverse operator $1/p(D)$ in a power series in D, so that

$$y = \frac{1}{p(D)} f(x) = (1 + b_1 D + b_2 D^2 + \cdots) f(x).$$

The reason for this is that high derivatives of polynomials disappear, because $D^k x^n = 0$ if $k > n$.

Example 5. Find a particular solution of $y''' - 2y'' + y = x^4 + 2x + 5$.

Solution. We have $(D^3 - 2D^2 + 1)y = x^4 + 2x + 5$, so

$$y = \frac{1}{1 - 2D^2 + D^3}(x^4 + 2x + 5).$$

By ordinary long division we find that

$$\frac{1}{1 - 2D^2 + D^3} = 1 + 2D^2 - D^3 + 4D^4 - 4D^5 + \cdots,$$

so

$$\begin{aligned} y &= (1 + 2D^2 - D^3 + 4D^4 - 4D^5 + \cdots)(x^4 + 2x + 5) \\ &= (x^4 + 2x + 5) + 2(12x^2) - (24x) + 4(24) \\ &= x^4 + 24x^2 - 22x + 101. \end{aligned}$$

In order to make the fullest use of this method, it is desirable to keep in mind the following series expansions from elementary algebra:

$$\frac{1}{1-r} = 1 + r + r^2 + r^3 + \cdots \quad \text{and} \quad \frac{1}{1+r} = 1 - r + r^2 - r^3 + \cdots.$$

In this context we are only interested in these formulas as "formal" series expansions, and have no need to concern ourselves with their convergence behavior.

Example 6. Find a particular solution of $y''' + y'' + y' + y = x^5 - 2x^2 + x$.

Solution. We have $(D^3 + D^2 + D + 1)\, y = x^5 - 2x^2 + x$, so

$$\begin{aligned} y &= \frac{1}{1 + D + D^2 + D^3}(x^5 - 2x^3 + x) \\ &= \frac{1}{1 - D^4}(1 - D)(x^5 - 2x^2 + x) \\ &= \frac{1}{1 - D^4}\left[(x^5 - 2x^2 + x) - (5x^4 - 4x + 1)\right] \\ &= (1 + D^4 + D^8 + \cdots)[x^5 - 5x^4 - 2x^2 + 5x - 1] \\ &= (x^5 - 5x^4 - 2x^2 + 5x - 1) + (120x - 120) \\ &= x^3 - 5x^4 - 2x^2 + 125x - 121. \end{aligned}$$

The remarkable thing about the procedures illustrated in these examples is that they actually work!

METHOD 4: THE EXPONENTIAL SHIFT RULE. As we know, exponential functions behave in a special way under differentiation. This fact enables us to simplify our work whenever $f(x)$ contains a factor of the form e^{kx}. Thus, if $f(x)=e^{kx}g(x)$, we begin by noticing that

$$\begin{aligned}(D-r)f(x) &= (D-r)e^{kx}g(x)\\ &= e^{kx}Dg(x)+ke^{kx}g(x)-re^{kx}g(x)\\ &= e^{kx}(D+k-r)g(x).\end{aligned}$$

By applying this formula to the successive factors $D-r_1$, $D-r_2$, $D-r_n$, we see that for the polynomial operator $p(D)$,

$$p(D)e^{kx}g(x)=e^{kx}p(D+k)g(x). \tag{10}$$

This says that we can move the factor e^{kx} to the left of the operator $p(D)$ if we replace D by $D+k$ in the operator.

The same property is valid for the inverse operator $1/p(D)$, that is,

$$\frac{1}{p(D)}e^{kx}g(x)=e^{kx}\frac{1}{p(D+k)}g(x). \tag{11}$$

To see this, we simply apply $p(D)$ to the right side and use (10):

$$p(D)e^{kx}\frac{1}{p(D-k)}g(x)=e^{kx}p(D+k)\frac{1}{p(D+k)}g(x)=e^{kx}g(x).$$

Properties (10) and (11) are called the *exponential shift rule*. They are useful in moving exponential functions out of the way of operators.

Example 7. Solve the problem in Example 1 by this method.

Solution. We have $(D^2-3D+2)y=xe^x$, so

$$\begin{aligned}y &= \frac{1}{D^2-3D+2}xe^x = e^x\frac{1}{(D+1)^2-3(D+1)+2}x\\ &= e^x\frac{1}{D^2-D}x=-e^x\frac{1}{D}\frac{1}{1-D}x\\ &= -e^x\left(\frac{1}{D}+1+D+D^2+\cdots\right)x\\ &= -e^x\left(\frac{1}{2}x^2+x+1\right),\end{aligned}$$

as we have already seen in Examples 1 and 3.

Interested readers will find additional material on the methods of this section in the "Historical Introduction" to H. S. Carslaw and J. C. Jaeger, *Operational Methods In Applied Mathematics*, Dover, New York, 1963; and in E. Stephens, *The Elementary Theory of Operational Mathematics*, McGraw-Hill, New York, 1937.

Problems

1. Find a particular solution of $y'' - 4y = e^{2x}$ by using each of Methods 1 and 2.

2. Find a particular solution of $y'' - y = x^2 e^{2x}$ by using each of Methods 1, 2, and 4.

In Problems 3 to 6, find a particular solution by using Method 1.

3. $y'' + 4y' + 4y = 10x^3 e^{-2x}$.

4. $y'' - 2y' + y = e^x$.

5. $y'' - y = e^{-x}$.

6. $y'' - 2y' - 3y = 6e^{5x}$.

In Problems 7 to 15, find a particular solution by using Method 3.

7. $y'' - y' + y = x^3 - 3x^2 + 1$.

8. $y''' - 2y' + y = 2x^3 - 3x^2 + 4x + 5$.

9. $4y'' + y = x^4$.

10. $y^{(5)} - y''' = x^2$.

11. $y^{(6)} - y = x^{10}$.

12. $y'' + y' - y = 3x - x^4$.

13. $y'' + y = x^4$.

14. $y''' - y'' = 12x - 2$.

15. $y''' + y'' = 9x^2 - 2x + 1$.

In Problems 16 to 18, find a particular solution by using Method 4.

16. $y'' - 4y' + 3y = x^3 e^{2x}$.

17. $y'' - 7y' + 12y = e^{2x}(x^3 - 5x^2)$.

18. $y'' + 2y' + y = 2x^2 e^{-2x} + 3e^{2x}$.

In Problems 19 to 24, find a particular solution by any method.

19. $y''' - 8y = 16x^2$.

20. $y^{(4)} - y = 1 - x^3$.

21. $y''' - \frac{1}{4} y' = x$.

22. $y^{(4)} = x^{-3}$.

23. $y''' - y'' + y' = x + 1$.

24. $y''' + 2y'' = x$.

25. Use the exponential shift rule to find the general solution of each of the following equations:

(a) $(D - 2)^3 y = e^{2x}$ [hint: multiply by e^{-2x} and use (10)];

(b) $(D + 1)^3 y = 12e^{-x}$;

(c) $(D - 2)^2 y = e^{2x} \sin x$.

26. Consider the nth order homogeneous equation $p(D)y = 0$.

(a) If a polynomial $q(r)$ is a factor of the auxiliary polynomial $p(r)$, show that any solution of the differential equation $q(D)y = 0$ is also a solution of $p(D)y = 0$.

(b) If r_1 is a root of multiplicity k of the auxiliary equation $p(r) = 0$, show that any solution of $(D - r_1)^k y = 0$ is also a solution of $p(D)y = 0$.

(c) Use the exponential shift rule to show that $(D - r_1)^k y = 0$ has

$$y = (c_1 + c_2 x + c_3 x^2 + \cdots + c_k x^{k-1}) e^{r_1 x}$$

as its general solution. *Hint*: $(D - r_1)^k y = 0$ is equivalent to $e^{r_1 x} D^k (e^{-r_1 x} y) = 0$.

Appendix A. Euler

Leonhard Euler (1707–1783) was Switzerland's foremost scientist and one of the three greatest mathematicians of modern times (the other two being Gauss and Riemann).

He was perhaps the most prolific author of all time in any field. From 1727 to 1783 his writings poured out in a seemingly endless flood, constantly adding knowledge to every known branch of pure and applied mathematics, and also to many that were not known until he created them. He averaged about 800 printed pages a year throughout his long life, and yet he almost always had something worthwhile to say and never seems long-winded. The publication of his complete works was started in 1911, and the end is not in sight. This edition was planned to include 887 titles in 72 volumes, but since that time extensive new deposits of previously unknown manuscripts have been unearthed, and it is now estimated that more than 100 large volumes will be required for completion of the project. Euler evidently wrote mathematics with the ease and fluency of a skilled speaker discoursing on subjects with which he is intimately familiar. His writings are models of relaxed clarity. He never condensed, and he reveled in the rich abundance of his ideas and the vast scope of his interests. The French physicist Arago, in speaking of Euler's incomparable mathematical facility, remarked that "He calculated without apparent effort, as men breathe, or as eagles sustain themselves in the wind." He suffered total blindness during the last 17 years of his life, but with the aid of his powerful memory and fertile imagination, and with helpers to write his books and scientific papers from dictation, he actually increased his already prodigious output of work.

Euler was a native of Basel and a student of John Bernoulli at the University, but he soon outstripped his teacher. His working life was spent as a member of the Academies of Science at Berlin and St. Petersburg, and most of his papers were published in the journals of these organizations. His business was mathematical research, and he knew his business. He was also a man of broad culture, well versed in the classical languages and literatures (he knew the *Aeneid* by heart), many modern languages, physiology, medicine, botany, geography, and the entire body of physical science as it was known in his time. However, he had little talent for metaphysics or disputation, and came out second best in many good-natured verbal encounters with Voltaire at the court of Frederick the Great. His personal life was as placid and uneventful as is possible for a man with 13 children.

Though he was not himself a teacher, Euler has had a deeper influence on the teaching of mathematics than any other man. This came about chiefly through his three great treatises: *Introductio in Analysin Infinitorum* (1748); *Institutiones Calculi Differentialis* (1755); and *Institutiones Calculi Integralis* (1768–1794). There is considerable truth in the old saying that all elementary and advanced calculus textbooks since 1748 are essentially copies of Euler or copies of copies of Euler.[14] These works summed up and codified the discoveries of his predecessors, and are full of Euler's own ideas. He extended and perfected plane and solid analytic geometry, introduced the analytic approach to trigonometry, and was responsible for the modern treatment of the functions $\log x$ and e^x. He created a consistent theory of logarithms of negative and imaginary numbers, and discovered that $\log x$ has an infinite number of values. It was through his work that the symbols e, π, and $i\,(=\sqrt{-1})$ became common currency for all mathematicians, and it was he who linked them together in the astonishing relation $e^{\pi i}=-1$. This is merely a special case (put $\theta=\pi$) of his famous formula $e^{i\theta}=\cos\theta+i\sin\theta$, which connects the exponential and trigonometric functions and is absolutely indispensable in higher analysis.[15] Among his other contributions to standard mathematical

[14] See C. B. Boyer, "The Foremost Textbook of Modern Times," *Am. Math. Monthly*, Vol. 58, pp. 223–226, 1951.

[15] An even more astonishing consequence of his formula is the fact that an imaginary power of an imaginary number can be real, in particular $i^i=e^{-\pi/2}$; for if we put $\theta=\pi/2$, we obtain $e^{\pi i/2}=i$, so

$$i^i=(e^{\pi i/2})^i=e^{\pi i^2/2}=e^{-\pi/2}.$$

Euler further showed that i^i has infinitely many values, of which this calculation produces only one.

notation were sin x, cos x, the use of $f(x)$ for an unspecified function, and the use of Σ for summation.[16] Good notations are important, but the ideas behind them are what really count, and in this respect Euler's fertility was almost beyond belief. He preferred concrete special problems to the general theories in vogue today, and his unique insight into the connections between apparently unrelated formulas blazed many trails into new areas of mathematics which he left for his successors to cultivate.

He was the first and greatest master of infinite series, infinite products, and continued fractions, and his works are crammed with striking discoveries in these fields. James Bernoulli (John's older brother) found the sums of several infinite series, but he was not able to find the sum of the reciprocals of the squares, $1+\frac{1}{4}+\frac{1}{9}+\frac{1}{16}+\ldots$. He wrote, "If someone should succeed in finding this sum, and will tell me about it, I shall be much obliged to him." In 1736, long after James's death, Euler made the wonderful discovery that

$$1+\frac{1}{4}+\frac{1}{9}+\frac{1}{16}+\cdots=\frac{\pi^2}{6}.$$

He also found the sums of the reciprocals of the fourth and sixth powers,

$$1+\frac{1}{2^4}+\frac{1}{3^4}+\cdots=1+\frac{1}{16}+\frac{1}{81}+\cdots=\frac{\pi^4}{90}$$

and

$$1+\frac{1}{2^6}+\frac{1}{3^6}+\cdots=1+\frac{1}{64}+\frac{1}{729}+\cdots=\frac{\pi^6}{945}.$$

When John heard about these feats, he wrote, "If only my brother were alive now."[17] Few would believe that these formulas are related—as they are—to Wallis's infinite product (1656),

$$\frac{\pi}{2}=\frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdots.$$

Euler was the first to explain this in a satisfactory way, in terms of his infinite product expansion of the sine,

$$\frac{\sin x}{x}=\left(1-\frac{x^2}{\pi^2}\right)\left(1-\frac{x^2}{4\pi^2}\right)\left(1-\frac{x^2}{9\pi^2}\right)\cdots.$$

[16] See F. Cajori, *A History of Mathematical Notations*, Open Court, Chicago, 1929.

[17] The world is still waiting—more than 200 years later—for someone to discover the sum of the reciprocals of the cubes.

Wallis's product is also related to Brouncker's remarkable continued fraction,

$$\frac{\pi}{4} = \cfrac{1}{1+\cfrac{1^2}{2+\cfrac{3^2}{2+\cfrac{5^2}{2+\cfrac{7^2}{2+\cdots}}}}},$$

which became understandable only in the context of Euler's extensive researches in this field.

His work in all departments of analysis strongly influenced the further development of this subject through the next two centuries. He contributed many important ideas to differential equations, including substantial parts of the theory of second order linear equations and the method of solution by power series. He gave the first systematic discussion of the calculus of variations, which he founded on his basic differential equation for a minimizing curve. He introduced the number now known as *Euler's constant*,

$$\gamma = \lim_{n\to\infty}\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}-\log n\right) = 0.5772\ldots,$$

which is the most important special number in mathematics after π and e. He discovered the integral defining the gamma function,

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt,$$

which is often the first of the so-called higher transcendental functions that students meet beyond the level of calculus, and he developed many of its applications and special properties. He also worked with Fourier series, encountered the Bessel functions in his study of the vibrations of a stretched circular membrane, and applied Laplace transforms to solve differential equations—all before Fourier, Bessel, and Laplace were born. Even though Euler died about 200 years ago, he lives everywhere in analysis.

E. T. Bell, the well-known historian of mathematics, observed that "One of the most remarkable features of Euler's universal genius was its equal strength in both of the main currents of mathematics, the continuous and the discrete." In the realm of the discrete, he was one of the originators of modern number theory and made many far-reaching contributions to this subject throughout his life. In addition, the origins of topology—one of the dominant forces in modern mathematics—lie in his solution of the Königsberg bridge problem and his formula $V-E+F=2$ connecting the numbers of vertices, edges, and faces of a simple polyhedron. In the following paragraphs, we briefly describe some of his activities in these fields.

In number theory, Euler drew much of his inspiration from the challenging marginal notes left by Fermat in his copy of the works of Diophantus. He gave the first published proofs of both Fermat's theorem and Fermat's two squares theorem. He later generalized the first of these classic results by introducing the Euler ϕ function; his proof of the second cost him 7 years of intermittent effort. In addition, he proved that every positive integer is a sum of four squares and investigated the law of quadratic reciprocity.

Some of his most interesting work was connected with the sequence of prime numbers, that is, with those integers $p > 1$ whose only positive divisors are 1 and p. His use of the divergence of the harmonic series $1+\frac{1}{2}+\frac{1}{3}+\cdots$ to prove Euclid's theorem that there are infinitely many primes is so simple and ingenious that we venture to give it here. Suppose that there are only N primes, say $p_1, p_2, \ldots, p_N$. Then each integer $n > 1$ is uniquely expressible in the form $n = p_1^{a_1} p_2^{a_2} \cdots p_N^{a_N}$. If a is the largest of these exponents, then it is easy to see that

$$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n} \leq \left(1+\frac{1}{p_1}+\frac{1}{p_1^2}+\cdots+\frac{1}{p_1^a}\right)$$
$$\times\left(1+\frac{1}{p_2}+\frac{1}{p_2^2}+\cdots+\frac{1}{p_2^a}\right)\cdots\left(1+\frac{1}{p_N}+\frac{1}{p_N^2}+\cdots+\frac{1}{p_N^a}\right),$$

by multiplying out the factors on the right. But the simple formula $1+x+x^2+\cdots=1/(1-x)$, which is valid for $|x| < 1$, shows that the factors in the above product are less than the numbers

$$\frac{1}{1-1/p_1}, \frac{1}{1-1/p_2}, \ldots, \frac{1}{1-1/p_N},$$

so

$$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n} < \frac{p_1}{p_1-1}\,\frac{p_2}{p_2-1}\cdots\frac{p_N}{p_N-1}$$

for every n. This contradicts the divergence of the harmonic series and shows that there cannot exist only a finite number of primes. He also proved that the series

$$\frac{1}{2}+\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{11}+\frac{1}{13}+\frac{1}{17}+\cdots$$

of the reciprocals of the primes diverges, and discovered the following wonderful identity: if $s > 1$, then

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \frac{1}{1-1/p^s},$$

where the expression on the right denotes the product of the numbers $(1 - p^{-s})^{-1}$ for all primes p. We shall return to this identity later, in our note on Riemann in Appendix E in Chapter 5.

He also initiated the theory of partitions, a little-known branch of number theory that turned out much later to have applications in statistical mechanics and the kinetic theory of gases. A typical problem of this subject is to determine the number $p(n)$ of ways in which a given positive integer n can be expressed as a sum of positive integers, and if possible to discover some properties of this function. For example, 4 can be partitioned into $4=3+1=2+2=2+1+1=1+1+1+1$, so $p(4)=5$, and similarly $p(5)=7$ and $p(6)=11$. It is clear that $p(n)$ increases very rapidly with n, so rapidly, in fact, that[18]

$$p(200) = 3{,}972{,}999{,}029{,}388.$$

Euler began his investigations by noticing (only geniuses notice such things) that $p(n)$ is the coefficient of x^n when the function $[(1 - x)(1 - x^2)(1 - x^3) \cdots]^{-1}$ is expanded in a power series:

$$\frac{1}{(1-x)(1-x^2)(1-x^3)\ldots} = 1 + p(1)x + p(2)x^2 + p(3)x^3 + \cdots.$$

By building on this foundation, he derived many other remarkable identities related to a variety of problems about partitions.[19]

[18] This evaluation required a month's work by a skilled computer in 1918. His motive was to check an approximate formula for $p(n)$, namely

$$p(n) \cong \frac{1}{4n\sqrt{3}} e^{\pi\sqrt{2n/3}}$$

(the error was extremely small).

[19] See Chapter XIX of G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, 1938; or Chapters 12–14 of G. E. Andrews, *Number Theory*, W. B. Saunders, San Francisco, 1971. These treatments are "elementary" in the technical sense that they do not use the high-powered machinery of advanced analysis, but nevertheless they are far from simple. For students who wish to experience some of Euler's most interesting work in number theory at first hand, and in a context not requiring much previous knowledge, we recommend Chapter VI of G. Polya's fine book, *Induction and Analogy in Mathematics*, Princeton University Press, 1954.

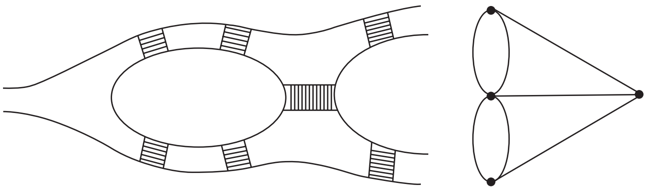

FIGURE 33
The Königsberg bridges.

The Königsberg bridge problem originated as a pastime of Sunday strollers in the town of Königsberg (now Kaliningrad) in what was formerly East Prussia. There were seven bridges across the river that flows through the town (see Figure 33). The residents used to enjoy walking from one bank to the islands and then to the other bank and back again, and the conviction was widely held that it is impossible to do this by crossing all seven bridges without crossing any bridge more than once. Euler analyzed the problem by examining the schematic diagram given on the right in the figure, in which the land areas are represented by points and the bridges by lines connecting these points. The points are called vertices, and a vertex is said to be odd or even according as the number of lines leading to it is odd or even. In modern terminology, the entire configuration is called a *graph*, and a path through the graph that traverses every line but no line more than once is called an *Euler path*. An Euler path need not end at the vertex where it began, but if it does, it is called an *Euler circuit*. By the use of combinatorial reasoning, Euler arrived at the following theorems about any such graph: (1) there are an even number of odd vertices; (2) if there are no odd vertices, there is an Euler circuit starting at any point; (3) if there are two odd vertices, there is no Euler circuit, but there is an Euler path starting at one odd vertex and ending at the other; (4) if there are more than two odd vertices, there are no Euler paths.[20] The graph of the Königsberg bridges has four odd vertices, and therefore, by the last theorem, has no Euler paths.[21] The branch of mathematics that has developed from these ideas is known as *graph theory*; it has applications to chemical bonding, economics, psychosociology, the properties of networks of roads and railroads, and other subjects.

A polyhedron is a solid whose surface consists of a number of polygonal faces, and a regular polyhedron has faces that are regular polygons. As we know, there exists a regular polygon with n sides for each positive integer

[20] Euler's original paper of 1736 is interesting to read and easy to understand; it can be found on pp. 573–580 of J. R. Newman (ed), *The World of Mathematics*, Simon and Schuster, New York, 1956.

[21] It is easy to see—without appealing to any theorems— that this graph contains no Euler circuit, for if there were such a circuit, it would have to enter each vertex as many times as it leaves it, and therefore every vertex would have to be even. Similar reasoning shows also that if there were an Euler path that is not a circuit, there would be two odd vertices.

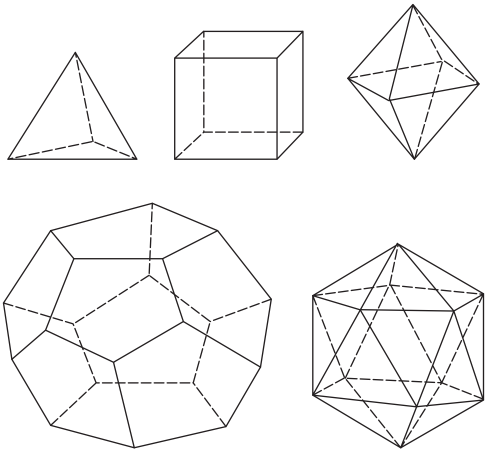

FIGURE 34
Regular polyhedra.

$n = 3, 4, 5, \ldots$, and they even have special names—equilateral triangle, square, regular pentagon, etc. However, it is a curious fact—and has been known since the time of the ancient Greeks—that there are only five regular polyhedra, those shown in Figure 34, with names given in the table below.

The Greeks studied these figures assiduously, but it remained for Euler to discover the simplest of their common properties: If V, E and F denote the numbers of vertices, edges, and faces of any one of them, then in every case we have

$$V - E + F = 2.$$

This fact is known as *Euler's formula for polyhedra*, and it is easy to verify from the data summarized in the following table.

	V	E	F
Tetrahedron	4	6	4
Cube	8	12	6
Octahedron	6	12	8
Dodecahedron	20	30	12
Icosahedron	12	30	20

This formula is also valid for any irregular polyhedron as long as it is *simple*—which means that it has no "holes" in it, so that its surface can be deformed continuously into the surface of a sphere. Figure 35 shows two simple irregular polyhedra for which $V - E + F = 6 - 10 + 6 = 2$ and $V - E + F = 6 - 9 + 5 = 2$. However, Euler's formula must be extended to

$$V - E + F = 2 - 2p$$

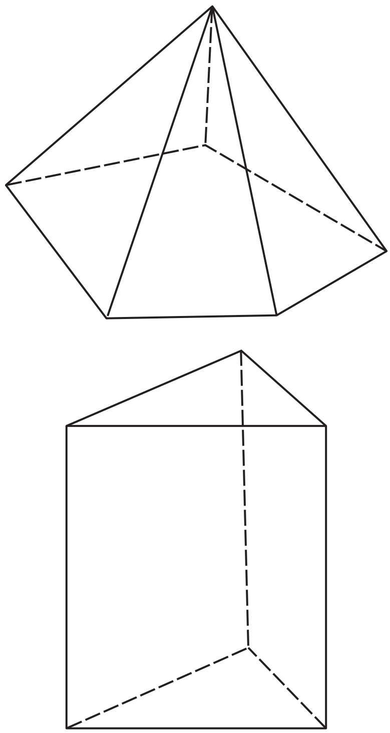

FIGURE 35

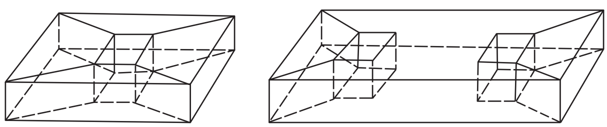

FIGURE 36

in the case of a polyhedron with p holes (a simple polyhedron is one for which $p=0$). Figure 36 illustrates the cases $p=1$ and $p=2$; here we have $V-E+F=16-32+16=0$ when $p=1$, and $V-E+F=24-44+18=-2$ when $p=2$. The significance of these ideas can best be understood by imagining a polyhedron to be a hollow figure with a surface made of thin rubber, and inflating it until it becomes smooth. We no longer have flat faces and straight edges, but instead a map on the surface consisting of curved regions, their boundaries, and points where boundaries meet. The number $V-E+F$ has the same value for all maps on our surface, and is called the *Euler characteristic* of this surface. The number p is called the *genus* of the surface. These two numbers, and the relation between them given by the equation $V-E+F=2-2p$, are evidently unchanged when the surface is continuously deformed by stretching or bending. Intrinsic geometric properties of this kind—which have little connection with the type of geometry concerned with lengths, angles, and areas—are called *topological*. The serious study of such topological properties has greatly increased during the past century, and has furnished valuable insights to many branches of mathematics and science.[22]

The distinction between pure and applied mathematics did not exist in Euler's day, and for him the entire physical universe was a convenient object whose diverse phenomena offered scope for his methods of analysis. The foundations of classical mechanics had been laid down by Newton, but Euler was the principal architect. In his treatise of 1736 he was the first to explicitly introduce the concept of a mass-point or particle, and he was also the first to study the acceleration of a particle moving along any curve and to use the notion of a vector in connection with velocity and acceleration. His continued successes in mathematical physics were so numerous, and his influence was so pervasive, that most of his discoveries are not credited to him at all and are taken for granted by physicists as part of the natural order of things. However, we do have Euler's equations of motion for the rotation of a rigid body, Euler's hydrodynamic equation for the flow of an ideal incompressible fluid, Euler's law for the bending of elastic beams, and Euler's critical load in the theory of the buckling of columns. On several occasions the thread of his scientific thought led him to ideas his contemporaries were not ready to assimilate. For example, he foresaw the phenomenon of radiation pressure, which is crucial for the modern theory of the stability of stars, more than a century before Maxwell rediscovered it in his own work on electromagnetism.

Euler was the Shakespeare of mathematics—universal, richly detailed, and inexhaustible.[23]

Appendix B. Newton

Most people are acquainted in some degree with the name and reputation of Isaac Newton (1642–1727), for his universal fame as the discoverer of the law of gravitation has continued undiminished over the two and a half centuries since his death. It is less well known, however, that in the immense sweep of his vast achievements he virtually created modern physical science, and in consequence has had a deeper influence on the direction of civilized life than the rise and fall of nations. Those in a position to judge have been

[22] Proofs of Euler's formula and its extension are given on pp. 236–240 and 256–259 of R. Courant and H. Robbins, *What Is Mathematics?*, Oxford University Press, 1941. See also G. Polya, op. cit., pp. 35–43.

[23] For further information, see C. Truesdell, "Leonhard Euler, Supreme Geometer (1707–1783)," in *Studies in Eighteenth-Century Culture*, Case Western Reserve University Press, 1972. Also, the November 1983 issue of *Mathematics Magazine* is wholly devoted to Euler and his work.

unanimous in considering him one of the very few supreme intellects that the human race has produced.

Newton was born to a farm family in the village of Woolsthorpe in northern England. Little is known of his early years, and his undergraduate life at Cambridge seems to have been outwardly undistinguished. In 1665 an outbreak of the plague caused the universities to close, and Newton returned to his home in the country, where he remained until 1667. There, in 2 years of rustic solitude—from age 22 to 24—his creative genius burst forth in a flood of discoveries unmatched in the history of human thought: the binomial series for negative and fractional exponents; differential and integral calculus; universal gravitation as the key to the mechanism of the solar system; and the resolution of sunlight into the visual spectrum by means of a prism, with its implications for understanding the colors of the rainbow and the nature of light in general. In his old age he reminisced as follows about this miraculous period of his youth: "In those days I was in the prime of my age for invention and minded Mathematicks and Philosophy [i.e., science] more than at any time since."[24]

Newton was always an inward and secretive man, and for the most part kept his monumental discoveries to himself. He had no itch to publish, and most of his great works had to be dragged out of him by the cajolery and persistence of his friends. Nevertheless, his unique ability was so evident to his teacher, Isaac Barrow, that in 1669 Barrow resigned his professorship in favor of his pupil (an unheard-of event in academic life), and Newton settled down at Cambridge for the next 27 years. His mathematical discoveries were never really published in connected form; they became known in a limited way almost by accident, through conversations and replies to questions put to him in correspondence. He seems to have regarded his mathematics mainly as a fruitful tool for the study of scientific problems, and of comparatively little interest in itself. Meanwhile, Leibniz in Germany had also invented calculus independently; and by his active correspondence with the Bernoullis and the later work of Euler, leadership in the new analysis passed to the Continent, where it remained for 200 years.[25]

Not much is known about Newton's life at Cambridge in the early years of his professorship, but it is certain that optics and the construction of telescopes were among his main interests. He experimented with many

[24] The full text of this autobiographical statement (probably written sometime in the period 1714–1720) is given on pp. 291–292 of I. Bernard Cohen, *Introduction to Newton's 'Principia*,' Harvard University Press, 1971. The present writer owns a photograph of the original document.

[25] It is interesting to read Newton's correspondence with Leibniz (via Oldenburg) in 1676 and 1677 (see *The Correspondence of Isaac Newton*, Cambridge University Press, 1959–1976, 6 volumes so far). In Items 165, 172, 188, and 209, Newton discusses his binomial series but conceals in anagrams his ideas about calculus and differential equations, while Leibniz freely reveals his own version of calculus. Item 190 is also of considerable interest, for in it Newton records what is probably the earliest statement and proof of the Fundamental Theorem of Calculus.

techniques for grinding lenses (using tools which he made himself), and about 1670 built the first reflecting telescope, the earliest ancestor of the great instruments in use today at Mount Palomar and throughout the world. The pertinence and simplicity of his prismatic analysis of sunlight have always marked this early work as one of the timeless classics of experimental science. But this was only the beginning, for he went further and further in penetrating the mysteries of light, and all his efforts in this direction continued to display experimental genius of the highest order. He published some of his discoveries, but they were greeted with such contentious stupidity by the leading scientists of the day that he retired back into his shell with a strengthened resolve to work thereafter for his own satisfaction alone. Twenty years later he unburdened himself to Leibniz in the following words: "As for the phenomena of colours.. . I conceive myself to have discovered the surest explanation, but I refrain from publishing books for fear that disputes and controversies may be raised against me by ignoramuses."[26]

In the late 1670s Newton lapsed into one of his periodic fits of distaste for science, and directed his energies into other channels. As yet he had published nothing about dynamics or gravity, and the many discoveries he had already made in these areas lay unheeded in his desk. At last, however, under the skillful prodding of the astronomer Edmund Halley (of Halley's Comet), he turned his mind once again to these problems and began to write his greatest work, the *Principia*.[27]

It all seems to have started in 1684 with three men in deep conversation in a London inn—Halley, and his friends Christopher Wren and Robert Hooke. By thinking about Kepler's third law of planetary motion, Halley had come to the conclusion that the attractive gravitational force holding the planets in their orbits was probably inversely proportional to the square of the distance from the sun.[28] However, he was unable to do anything more with the idea than formulate it as a conjecture. As he later wrote (in 1686):

> I met with Sir Christopher Wren and Mr. Hooke, and falling in discourse about it, Mr. Hooke affirmed that upon that principle all the Laws of the celestiall motions were to be demonstrated, and that he himself had

[26] *Correspondence*, Item 427.

[27] The full title is *Philosophiae Naturalis Principia Mathematica (Mathematical Principles of Natural Philosophy)*.

[28] At that time this was quite easy to prove under the simplifying assumption—which contradicts Kepler's other two laws—that each planet moves with constant speed v in a circular orbit of radius r. [Proof: In 1673 Huygens had shown, in effect, that the acceleration a of such a planet is given by $a = v^2/r$. If T is the periodic time, then

$$a = \frac{(2\pi r/T)^2}{r} = \frac{4\pi^2}{r^2} \cdot \frac{r^3}{T^2}.$$

By Kepler's third law, T^2 is proportional to r^3, so r^3/T^2 is constant, and a is therefore inversely proportional to r^2. If we now suppose that the attractive force F is proportional to the acceleration, then it follows that F is also inversely proportional to r^2.]

> done it. I declared the ill success of my attempts; and Sir Christopher, to encourage the Inquiry, said that he would give Mr. Hooke or me two months' time to bring him a convincing demonstration therof, and besides the honour, he of us that did it, should have from him a present of a book of 40 shillings. Mr. Hooke then said that he had it, but that he would conceale it for some time, that others triing and failing, might know how to value it, when he should make it publick; however, I remember Sir Christopher was little satisfied that he could do it, and tho Mr. Hooke then promised to show it him, I do not yet find that in that particular he has been as good as his word.[29]

It seems clear that Halley and Wren considered Hooke's assertions to be merely empty boasts. A few months later Halley found an opportunity to visit Newton in Cambridge, and put the question to him: "What would be the curve described by the planets on the supposition that gravity diminishes as the square of the distance?" Newton answered immediately, "An ellipse." Struck with joy and amazement, Halley asked him how he knew that. "Why," said Newton, "I have calculated it." Not guessed, or surmised, or conjectured, but *calculated*. Halley wanted to see the calculations at once, but Newton was unable to find the papers. It is interesting to speculate on Halley's emotions when he realized that the age-old problem of how the solar system works had at last been solved—but that the solver hadn't bothered to tell anybody and had even lost his notes. Newton promised to write out the theorems and proofs again and send them to Halley, which he did. In the course of fulfilling his promise he rekindled his own interest in the subject, and went on, and greatly broadened the scope of his researches.[30]

In his scientific efforts Newton somewhat resembled a live volcano, with long periods of quiescence punctuated from time to time by massive eruptions of almost superhuman activity. The *Principia* was written in 18 incredible months of total concentration, and when it was published in 1687 it was immediately recognized as one of the supreme achievements of the human mind. It is still universally considered to be the greatest contribution to science ever made by one man. In it he laid down the basic principles of theoretical mechanics and fluid dynamics; gave the first mathematical treatment of wave motion; deduced Kepler's laws from the inverse square law of gravitation, and explained the orbits of comets; calculated the masses of the earth, the sun, and the planets with satellites; accounted for the flattened shape of the earth, and used this to explain the precession of the equinoxes; and founded the theory of tides. These are only a few of the splendors of this prodigious work.[31] The *Principia* has always been a difficult book to read, for the

---

[29] *Correspondence*, Item 289.

[30] For additional details and the sources of our information about these events, see Cohen, *op. cit*., pp. 47–54.

[31] A valuable outline of the contents of the *Principia* is given in Chapter VI of W. W. Rouse Ball, *An Essay on Newton's Principia* (first published in 1893; reprinted in 1972 by Johnson Reprint Corp, New York).

style has an inhuman quality of icy remoteness, which perhaps is appropriate to the grandeur of the theme. Also, the densely packed mathematics consists almost entirely of classical geometry, which was little cultivated then and is less so now.[32] In his dynamics and celestial mechanics, Newton achieved the victory for which Copernicus, Kepler, and Galileo had prepared the way. This victory was so complete that the work of the greatest scientists in these fields over the next two centuries amounted to little more than footnotes to his colossal synthesis. It is also worth remembering in this context that the science of spectroscopy, which more than any other has been responsible for extending astronomical knowledge beyond the solar system to the universe at large, had its origin in Newton's spectral analysis of sunlight.

After the mighty surge of genius that went into the creation of the *Principia*, Newton again turned away from science. However, in a famous letter to Bentley in 1692, he offered the first solid speculations on how the universe of stars might have developed out of a primordial featureless cloud of cosmic dust:

> It seems to me, that if the matter of our Sun and Planets and all the matter in the Universe was evenly scattered throughout all the heavens, and every particle has an innate gravity towards all the rest … some of it would convene into one mass and some into another, so as to make an infinite number of great masses scattered at great distances from one to another throughout all that infinite space. And thus might the Sun and Fixt stars be formed, supposing the matter were of a lucid nature.[33]

This was the beginning of scientific cosmology, and later led, through the ideas of Thomas Wright, Kant, Herschel, and their successors, to the elaborate and convincing theory of the nature and origin of the universe provided by late twentieth century astronomy.

In 1693 Newton suffered a severe mental illness accompanied by delusions, deep melancholy, and fears of persecution. He complained that he could not sleep, and said that he lacked his "former consistency of mind." He lashed out with wild accusations in shocking letters to his friends Samuel Pepys and John Locke. Pepys was informed that their friendship was over and that Newton would see him no more; Locke was charged with trying to entangle him with women and with being a "Hobbist" (a follower of Hobbes, i.e., an atheist and materialist).[34] Both men feared for Newton's sanity. They responded with careful concern and wise humanity, and the crisis passed.

[32] The nineteenth century British philosopher Whewell has a vivid remark about this: "Nobody since Newton has been able to use geometrical methods to the same extent for the like purposes; and as we read the *Principia* we feel as when we are in an ancient armoury where the weapons are of gigantic size; and as we look at them we marvel what manner of man he was who could use as a weapon what we can scarcely lift as a burden."

[33] *Correspondence*, Item 398.

[34] *Correspondence*, Items 420, 421, and 426.

In 1696 Newton left Cambridge for London to become Warden (and soon Master) of the Mint, and during the remainder of his long life he entered a little into society and even began to enjoy his unique position at the pinnacle of scientific fame. These changes in his interests and surroundings did not reflect any decrease in his unrivaled intellectual powers. For example, late one afternoon, at the end of a hard day at the Mint, he learned of a now-famous problem that the Swiss scientist John Bernoulli had posed as a challenge "to the most acute mathematicians of the entire world." The problem can be stated as follows: Suppose two nails are driven at random into a wall, and let the upper nail be connected to the lower by a wire in the shape of a smooth curve. What is the shape of the wire down which a bead will slide (without friction) under the influence of gravity so as to pass from the upper nail to the lower in the least possible time? This is Bernoulli's *brachistochrone* ("shortest time") *problem*. Newton recognized it at once as a challenge to himself from the Continental mathematicians; and in spite of being out of the habit of scientific thought, he summoned his resources and solved it that evening before going to bed. His solution was published anonymously, and when Bernoulli saw it, he wryly remarked, "I recognize the lion by his claw."

Of much greater significance for science was the publication of his *Opticks* in 1704. In this book he drew together and extended his early work on light and color. As an appendix he added his famous Queries, or speculations on areas of science that lay beyond his grasp in the future. In part the Queries relate to his lifelong preoccupation with chemistry (or alchemy, as it was then called). He formed many tentative but exceedingly careful conclusions—always founded on experiment—about the probable nature of matter; and though the testing of his speculations about atoms (and even nuclei) had to await the refined experimental work of the late nineteenth and early twentieth centuries, he has been proven absolutely correct in the main outlines of his ideas.[35] So, in this field of science too, in the prodigious reach and accuracy of his scientific imagination, he passed far beyond not only his contemporaries but also many generations of his successors. In addition, we quote two astonishing remarks from Queries 1 and 30, respectively: "Do Not Bodies act upon Light at a distance, and by their action bend its Rays?" and "Are not gross Bodies and Light convertible into one another?" It seems as clear as words can be that Newton is here conjecturing the gravitational bending of light and the equivalence of mass and energy, which are prime consequences of the theory of relativity. The former phenomenon was first observed during the total solar eclipse of May 1919, and the latter is now known to underlie the energy generated by the sun and the stars. On other occasions as well he seems to have known, in some mysterious intuitive way, far more than he was ever willing or able to justify, as in this cryptic sentence in a letter to a friend: "It's plain to me by the fountain I draw it from, though I

[35] See S. I. Vavilov, "Newton and the Atomic Theory," in *Newton Tercentenary Celebrations*, Cambridge University Press, 1947.

will not undertake to prove it to others."[36] Whatever the nature of this "fountain" may have been, it undoubtedly depended on his extraordinary powers of concentration. When asked how he made his discoveries, he said, "I keep the subject constantly before me and wait till the first dawnings open little by little into the full light." This sounds simple enough, but everyone with experience in science or mathematics knows how very difficult it is to hold a problem continuously in mind for more than a few seconds or a few minutes. One's attention flags; the problem repeatedly slips away and repeatedly has to be dragged back by an effort of will. From the accounts of witnesses, Newton seems to have been capable of almost effortless sustained concentration on his problems for hours and days and weeks, with even the need for occasional food and sleep scarcely interrupting the steady squeezing grip of his mind.

In 1695 Newton received a letter from his Oxford mathematical friend John Wallis, containing news that cast a cloud over the rest of his life. Writing about Newton's early mathematical discoveries, Wallis warned him that in Holland "your Notions" are known as "Leibniz's *Calculus Differentialis*," and he urged Newton to take steps to protect his reputation.[37] At that time the relations between Newton and Leibniz were still cordial and mutually respectful. However, Wallis's letters soon curdled the atmosphere, and initiated the most prolonged, bitter, and damaging of all scientific quarrels: the famous (or infamous) Newton–Leibniz priority controversy over the invention of calculus.

It is now well established that each man developed his own form of calculus independently of the other, that Newton was first by 8 or 10 years but did not publish his ideas, and that Leibniz's papers of 1684 and 1686 were the earliest publications on the subject. However, what are now perceived as simple facts were not nearly so clear at the time. There were ominous minor rumblings for years after Wallis's letters, as the storm gathered:

> What began as mild innuendoes rapidly escalated into blunt charges of plagiarism on both sides. Egged on by followers anxious to win a reputation under his auspices, Newton allowed himself to be drawn into the centre of the fray; and, once his temper was aroused by accusations of dishonesty, his anger was beyond constraint. Leibniz's conduct of the controversy was not pleasant, and yet it paled beside that of Newton. Although he never appeared in public, Newton wrote most of the pieces that appeared in his defense, publishing them under the names of his young men, who never demurred. As president of the Royal Society, he appointed an "impartial" committee to investigate the issue, secretly wrote the report officially published by the society [in 1712], and reviewed it anonymously in the *Philosophical Transactions*. Even Leibniz's death could not allay Newton's wrath, and he continued to pursue the

[36] *Correspondence*, Item 193.
[37] *Correspondence*, Items 498 and 503.

> enemy beyond the grave. The battle with Leibniz, the irrepressible need to efface the charge of dishonesty, dominated the final 25 years of Newton's life. Almost any paper on any subject from those years is apt to be interrupted by a furious paragraph against the German philosopher, as he honed the instruments of his fury ever more keenly.[38]

All this was bad enough, but the disastrous effect of the controversy on British science and mathematics was much more serious. It became a matter of patriotic loyalty for the British to use Newton's geometrical methods and clumsy calculus notations, and to look down their noses at the upstart work being done on the Continent. However, Leibniz's analytical methods proved to be far more fruitful and effective, and it was his followers who were the moving spirits in the richest period of development in mathematical history. What has been called "the Great Sulk" continued; for the British, the work of the Bernoullis, Euler, Lagrange, Laplace, Gauss, and Riemann remained a closed book; and British mathematics sank into a coma of impotence and irrelevancy that lasted through most of the eighteenth and nineteenth centuries.

Newton has often been thought of and described as the ultimate rationalist, the embodiment of the Age of Reason. His conventional image is that of a worthy but dull absent-minded professor in a foolish powdered wig. But nothing could be further from the truth. This is not the place to discuss or attempt to analyze his psychotic flaming rages; or his monstrous vengeful hatreds that were unquenched by the death of his enemies and continued at full strength to the end of his own life; or the 58 sins he listed in the private confession he wrote in 1662; or his secretiveness and shrinking insecurity; or his peculiar relations with women, especially with his mother, who he thought had abandoned him at the age of 3. And what are we to make of the bushels of unpublished manuscripts (millions of words and thousands of hours of thought!) that reflect his secret lifelong studies of ancient chronology, early Christian doctrine, and the prophecies of Daniel and St. John? Newton's desire to know had little in common with the smug rationalism of the eighteenth century; on the contrary, it was a form of desperate self-preservation against the dark forces that he felt pressing in around him.[39] As an original thinker in science and mathematics he was a stupendous genius whose impact on the world can be seen by everyone; but as a man he was so strange in every way that normal people can scarcely begin to understand him. It is perhaps most accurate to think of him in medieval terms—as a consecrated, solitary, intuitive mystic for whom science and mathematics were means of reading the riddle of the universe.

---

[38] Richard S. Westfall, in the *Encyclopaedia Britannica*.

[39] The best effort is Frank E. Manuel's excellent book, *A Portrait of Isaac Newton*, Harvard University Press, 1968.