1
Physics from Archimedes to Rutherford

1.1
Mathematics and Physics in Antiquity

In the 3rd millennium BC, the descendants of the Sumerians in Mesopotamia already used a sexagesimal (base 60) place-value numeral system of symbols to represent numbers, in which the position of a symbol determines its value. They also knew how to solve *rhetorical* algebraic equations (written in words, rather than symbols). By the 2nd millenium BC, Mesopotamians had compiled tables of numbers representing the length of the sides of a right triangle.[1] Their astronomers catalogued the motion of the stars and planets, as well as the occurrence of lunar eclipses. The day was divided into 24 hours, each hour into 60 minutes, each minute into 60 seconds.

A decisive step in the development of mathematics was made by the Egyptians, who introduced special signs (*hieratic numerals*) for the numbers 1 to 9 and multiples of powers of 10. They devised methods of multiplication and division which only involved addition. However, their decimal number system did not include a zero symbol, nor did it use the principle of place value.

The roots of modern science grew in the city-states of Ancient Greece scattered across the Eastern Mediterranean. The Greeks established physics as a science, greatly advanced astronomy, and made seminal contributions to mathematics, including the idea of formal mathematical proof, the basic rules of geometry, discoveries in number theory, and the rudiments of symbolic algebra and calculus.

A remarkable testament to their ingenuity is the *Antikythera mechanism*, the oldest known complex scientific calculator, which was discovered in a wreck off the Greek island of Antikythera and is dated to about 100 BC (see Fig. 1.1). The mechanism, whose complexity is comparable to that of 18th century clocks, could calculate the positions of celestial bodies [1].

1) A general theorem stating the relationship of those lengths is attributed to Pythagoras of Samos (c. 570–490 BC), one of the early Greek philosophers (“lovers of wisdom”). He is also credited with discovering that the harmonic intervals correspond to unique whole number proportions.

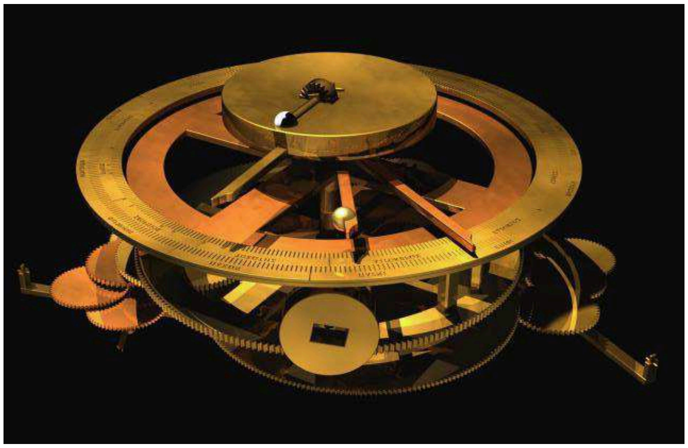

Fig. 1.1 Computer-generated re-creation of the Antikythera Mechanism. Credit: Tony Freeth/Antikythera Mechanism Research Project.

EUCLID of Alexandria (325–265 BC) is credited with establishing *Euclidean geometry*, a mathematical system used by Greek mathematicians and astronomers to describe the heavens. Euclid's *Elements* is the earliest known systematic and rigorous exposition of geometry — "*the first grandiose building of mathematical architecture*" [2]. Euclid presented mathematical concepts in logical order, starting with the most basic of assumptions, and using them to form a series of propositions and conclusions of increasing complexity. His work is as relevant today as it was in ancient times! However, there is a small ambiguity in the foundation of the system (the *fifth postulate*). This was explored by mathematicians in the 19th century, particularly by Bernhard Riemann (1826–1866) who laid foundations of a new form of non-Euclidean geometry which provides the mathematical basis for general relativity.

ARCHIMEDES of Syracuse (287–212 BC) is the greatest mathematician and physicist of antiquity. By consistently resorting to both mathematics and experiment to solve specific problems in physics, he established physics as a science. Archimedes used a method familiar from the *differential calculus* to construct a tangent to any given point of a curve, and devised general methods for finding areas of curvilinear plane figures and volumes bounded by curved surfaces, thus anticipating the *integral calculus* [3]. He showed that the surface of a sphere is four times that of the largest circle it contains, and that the volume of a sphere is two-thirds the volume of a circumscribed cylinder. Archimedes also devised a method of calculating the value of π to any desired degree of accuracy. On a recently deciphered palimpsest, Archimedes proposes that two *infinite sets* have the same size because the elements in them can be put in a one-to-one correspondence [4]. Such sets are said to have the same *cardinality* (this concept was introduced in the 19th century).

Archimedes formulated some of the fundamental laws of mechanics (e.g., the *law of the lever*), and laid the foundations of hydrostatics. *Archimedes' law* in hydrostatics, which is described in his treatise *On Floating Bodies*, states that *a body immersed in a fluid experiences a buoyant force equal to the weight of the fluid it displaces*. Since water is practically incompressible, an immersed body would displace an amount of water equal to its own volume. By dividing the mass of the body by the volume of water displaced, the density of the body could be determined.

ARISTARCHUS of Samos (310–230 BC) proposed the *heliocentric model* of the solar system described in Archimedes' book *The Sand Reckoner* [3] and in Plutarch's treatise *Face on the Moon*. Aristarchus knew that the Moon shines by reflected light. He computed the ratio of the distances to the Moon and the Sun by measuring the angle between them when the Moon is exactly half-illuminated by the Sun. Thus he found out that the Earth-Sun distance is large compared with the Earth-Moon distance. The Sun and the Moon have nearly equal apparent sizes, so he concluded that the diameter of the Sun was much larger than the diameter of the Moon, for their apparent diameters must be inversely proportional to their distances from the Earth. Since the Earth is much farther away from the Sun than from the Moon, the shadow cast by the Earth on the Moon during a lunar eclipse is roughly the same size as the Earth itself. To determine the Moon's distance from the Earth, Aristarchus timed a lunar eclipse and used $2\pi R/2r = \tau/t$ to obtain R/r; here R is the radius of Moon's orbit around the Earth, τ is the corresponding orbital period, r is the radius of the Earth, and t is the time it takes the mid-point of the Moon to traverse the shadow of width $2r$. Aristarchus reasoned that, since the Sun is much larger than the Earth, our planet most probably orbited the Sun. Therefore, he argued, the Earth must rotate on its axis in order to explain the apparent motion of the stars. Seventeen centuries later, Nicolaus Copernicus (1473–1543) published the epoch-making treatise *De Revolutionibus Orbium Celestium*, in which he described his heliocentric model of the universe.

ERASTOTHENES of Cyrene (276–197 BC) determined the size of the Earth by assuming that the Sun is so far away that its rays are essentially parallel. He knew that at noon on June 21 (the summer solstice), the Sun would illuminate the bottom of a deep vertical water well in the city of Syene (now Aswan on the Nile in Egypt), which means that the Sun was then at the zenith in the sky. At the same time, he measured the length of a shadow cast by an obelisk in Alexandria, located north of Syene. He used this length and the known height of the obelisk to infer that the direction to the Sun at Alexandria differed from that at Syene by an angle of seven degrees. In other words, the known distance between Syene and Alexandria was 7/360 of the circumference of the Earth. Expressed in modern units of length, Eratosthenes calculated a radius of 6100 km, amazingly close to the actual mean value of 6371 km.

HIPPARCHUS of Nicaea (190–120 BC) is the greatest of the ancient Greek astronomers. He developed *trigonometry*, discovered the precession of the equinoxes, calculated the length of the solar year and catalogued the positions of almost one thousand stars to a precision of one degree. Hipparchus classified the visible stars in six groups, called *magnitudes*, according to their apparent brightness: he divided a fixed time interval after sunset in six equal periods, and assigned the same magnitude to the stars that became visible to the naked eye during each period.

DIOPHANTUS of Alexandria (3rd century AD), often referred to as the "father of algebra", is best known for his *Arithmetica*, a series of books on algebraic equations and the theory of numbers (the Greek word for number is *arithmos*). Diophantus was the first Greek mathematician to recognize fractions as numbers. He also took a fundamental step from *rhetorical* algebra to modern *symbolic* algebra, wherein symbols are used for the unknown and for algebraic powers and operations.

The civilization of Ancient Greece spread across the African and Asian territories of the empire created by Alexander the Great (356–323 BC). The Greek scientific ideas were adopted and enriched by Indian and Islamic mathematicians, and were later reintroduced into Europe by Arab scholars, as well as through the work of Leonardo Fibonacci of Pisa (1175–1250). Arguably the single most important contribution of post-Hellenic scholars to modern mathematics is succinctly described by Pierre-Simon Laplace (1749–1827):

"*The ingenious method of expressing every possible number using a set of ten symbols* (*each symbol having a place value and an absolute value*) *emerged in India. The idea seems so simple nowadays that its significance and profound importance is no longer appreciated. Its simplicity lies in the way it facilitated calculation and placed arithmetic foremost among useful inventions*" [5].[2]

A textbook by Indian mathematician Brahmagupta (598–668) is considered the earliest work to treat zero as a number in its own right. Among the first to use zero as a place holder in positional base notation was Persian mathematician Muhammad al-Khwarizmi (780–850). The word *algebra* is derived from the name of one of the basic operations with equations (*al-jabr*) described in a book he wrote around 820 AD. Omar Khayyam (1048–1131), Persian astronomer, mathematician and poet, contributed to calendar reform and discovered a geometrical method of solving cubic equations by intersecting a parabola with a circle.

2) The three essential features of this number system — (1) nine signs and the concept of *zero*, (2) a place-value system, and (3) a decimal base — may have already been known to Chinese mathematicians. The Maya of Central America also discovered the zero digit, but used a *vigesimal* (base 20) number system [2].

1.2
Renaissance under Galileo, Kepler and Descartes

The age of modern physics was inaugurated by Galileo Galilei (1564–1642) through his use of geometry in the study of motion. Galileo discovered that:

- *Free objects propagate in straight lines at constant speed* (Newton's first law).
- *The trajectory of an object in a gravitational field does not depend on its mass.*

Galileo's *Discourses Regarding Two New Sciences* [6], which laid the foundations of statics and kinematics, is his ultimate scientific 'testament'. The book describes Galileo's discoveries concerning uniformly accelerated systems, free fall under gravity, the motion of pendulums, as well as the law of inertia.

By letting balls of different weights roll down a slope (see Fig. 1.2) — thereby slowing down their fall so that it could be timed[3] — Galileo showed that the balls experienced uniform acceleration [7]. For such a motion he established the *time-squared law* $d = at^2/2$, where d is the distance travelled and a is the acceleration. Galileo used a water clock to measure time intervals quite accurately, and found that the same law holds for freely falling objects, which implies that they experience a constant acceleration g due to gravity. He separated the motion of a projectile into a vertical component $y = gt^2/2$ and a horizontal component $x = v_x t$, where v_x is the constant velocity along the x-axis, and found that $y = const \cdot x^2$, i.e., that its trajectory is a parabola [8].

In his *Dialogue Concerning the Two Chief World Systems: Ptolemaic & Copernican*, Galileo asserts that the uniform motion of one *inertial frame* (the coordinate system in which Newton's first law, the *law of inertia*, is valid) relative to another cannot be detected by purely mechanical tests performed in each frame (*Galilean relativity*).

In 1608 Galileo built the first telescopes that could be used for astronomical observations (see Fig. 1.3). From his observations of the movement of sunspots, Galileo deduced that the Sun rotates around its axis. In his book *Starry Messenger*, published in 1610, Galileo claims to have observed that our galaxy, the *Milky Way*, is composed of myriads of stars. He also discovered Jupiter's satellites and the phases of the planet Venus. This last discovery was interpreted as evidence in favour of the Copernican heliocentric model of the solar system. Around 1640 Galileo invented the *pendulum clock*.

The basic measurement technique in astronomy is to determine the angles to a celestial object from two different places, and then deduce the distance to the object from known distance between the observation points (called the

3) Galileo placed little frets on a grooved slope, and adjusted their positions so that the clicks, produced when the ball passed over each fret, occurred at equal time intervals. Thus, the distances were direct measures of average speeds. He found that the speed increased as the odd numbers $1, 3, 5, 7, \ldots$ in equal time intervals.

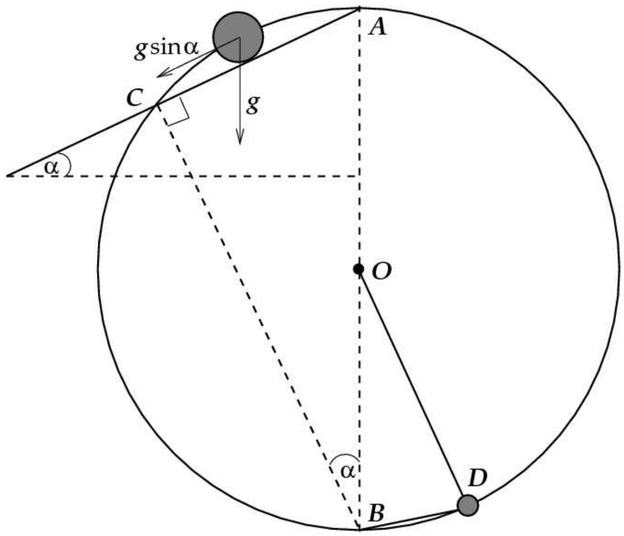

Fig. 1.2 *Galileo's theorem* states that a body will fall freely down the diameter AB of a circle in the same amount of time that it will fall along any chord AC of the circle. The theorem follows from the geometrical relation $AC/AB = sin\alpha$ and the *time-squared law* $d = at^2/2$ for the distances traversed. Based on this Galileo inferred that the period of a pendulum, τ, is independent of the amplitude of its swing, since the time to travel along any chord drawn to B will be the same as the time it takes the body to fall freely down twice the length OD of the pendulum [7]. Galileo also observed that τ depended on the pendulum's length, but not weight.

baseline). Ptolemy of Alexandria (90-168 AD) made a remarkably accurate estimate of the distance between the Earth and the Moon by measuring the angles of sight to the latter at different times, which is equivalent to making measurements at two different places because of the Earth's daily rotation.

Until the invention of the telescope, the instruments constructed by Tycho Brahe (1546–1601) produced the most accurate measurements of the positions of the planets (see Fig. 1.4). After Galileo's pioneering application of telescopes in astronomy, the ability to measure angles of sight precisely[4] gradually improved to the point that, when combined with a newfound precision in measuring locations on the Earth's surface, it enabled astronomers to determine, in 1672, the distance from the Earth to the Sun. This was a great breakthrough, for one could then use the diameter of the Earth's orbit around the Sun as the new baseline, thereby considerably enhancing the difference between the angles of sight. A century and a half later, Thomas Henderson (1798–1844) and Friedrich Bessel (1784–1846) determined the distances to our nearest stellar neighbours.

4) The astronomical parallax of Sirius, a nearby star, is 0.38 arcsec, about 5×10^{-4} the angular diameter of the Moon. That is the apparent size of a coin at a distance of a few kilometers!

Fig. 1.3 One of the telescopes for astronomical observations built by Galileo.

In 1609, after a decade of painstaking investigation using Tycho Brahe's astronomical data, Johannes Kepler (1571–1630) announced [9] the following two laws of planetary motion:

(1) *The orbit of Mars is an ellipse with the Sun at one focus*;

(2) *The planet's orbital velocity changes in such a way that the line joining Mars to the Sun covers equal areas of the ellipse in equal times*.

Kepler's third law of planetary motion [10], established in 1619, states that:

(3) *The square of the orbital period of a planet is proportional to the cube of its mean distance from the Sun*.

Kepler's momentous discovery regarding planetary orbits had to wait for two great advances in mathematics, *analytic geometry* and the *calculus*, before a self-consistent theory of celestial mechanics could be formulated. Analytic geometry took a definite shape in 1637 through the work of René Descartes (1596–1650), who fused algebra and geometry to create an essential foundation for the development of the differential and integral calculus. Descartes' treatise *La Géométrie* is a culmination of a century-long effort, in particular by Francois Viéte (1540–1603), to create modern *symbolic algebra*. Analytic ge-

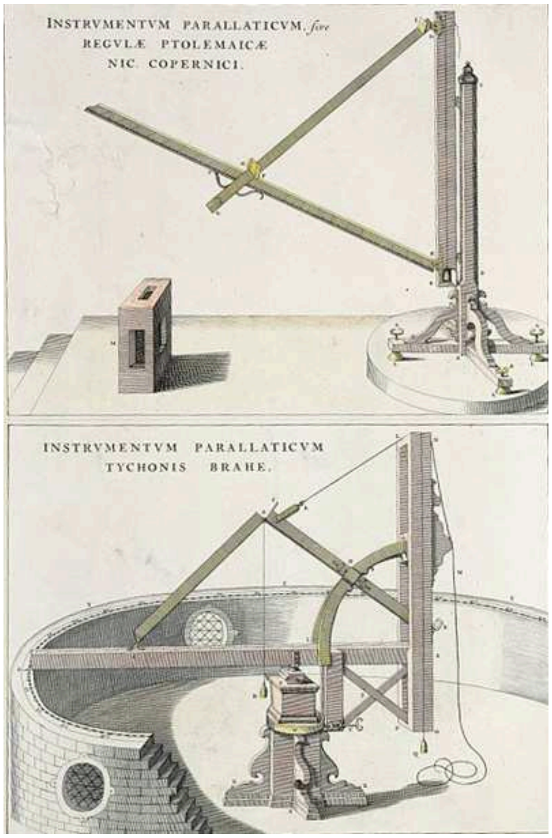

Fig. 1.4 These copper-plate engravings by Joan Blaeu were revised from wood-cuts originally published in Brahe's book *Astronomiae Instauratae Mecanica* (1598). "*This parallatic instrument, which is also called the Ptolemaic rulers, . . . is for observations of the zenith distances of the stars, in the way Ptolemy used to do it with the moon, particularly to find its maximum latitude . . . A few years earlier I had built some other rulers . . . In doing so I followed a particular method, which was partly an imitation of the ancient principle of Hipparchus . . . The use of the instrument depends on the fact that it will measure the altitude . . . as well as azimuths . . .*" (pages 28–31 in *Astronomiae*).

ometry was invented independently by Pierre de Fermat (1601–1665), who was also a founder, with Blaise Pascal (1623–1662), of the *theory of probability*. Descartes' *cartesian coordinates* made possible the drawing of tangents to curves of any kind. Fermat's method of finding the maximum or minimum value of any algebraic function involved the modern form of differentiation. "*Kepler was the first to introduce the idea of infinity into geometry and to note that the increment of a variable was evanescent for values of the variable in the immediate neighbourhood of a maximum or minimum*" [11].

1.3 Isaac Newton and his *Principia Mathematica*

The formulation of *classical mechanics* by Isaac Newton (1642–1727) represents a giant milestone in the history of science. In his masterpiece *Philosophiae Naturalis Principia Mathematica* [12], published in 1687, Newton shows that physical phenomena can be explained by a simple set of laws expressed in mathematical form. In particular, he demonstrates that the laws of planetary motion discovered by Johannes Kepler can be accounted for by the same law of gravitation as the motion of terrestrial objects in free fall.

In the years 1664–1666 Newton brought to fruition the ideas of his predecessors, most notably John Wallis (1616–1703) and Isaac Barrow (1630–1677), by developing the *differential calculus* ("the method of fluxions") and the *integral calculus* ("the inverse method of fluxions"). His most crucial insight was that integration and differentiation are inverse operations (the *fundamental theorem of calculus*) [13]. This development followed Newton's discovery that $(P + PQ)^{m/n} = P^{m/n} + (m/n)AQ + (m - n)BQ/2n + (m - 2n)CQ/3n + \ldots$, known as the *general binomial theorem*, which enabled him to find infinite series for many algebraic functions. Newton also invented the *calculus of variatons*.

"*The same year* [1666] *I began to think of gravity extending to the Orb of the Moon, and having found out how to estimate the force with which* [*a*] *globe revolving within a sphere presses the surface of the sphere, from Kepler's Rule of the periodical times of the Planets . . . I deduced that the forces which keep the Planets in their Orbs must* [*be*] *reciprocally as the squares of their distances from the centers about which they revolve; and thereby compared the force requisite to keep the Moon in her Orb with the force of gravity at the surface of the earth, and found them answer pretty nearly*" [14].

In 1679, Newton established that Kepler's areal law was a consequence of centripetal forces. He also showed that if the orbit is an ellipse under the action of central forces then the radial dependence of the force is the inverse square with the distance from the center. These propositions were stated in his 1684 tract *De Motu Corporum in Gyrum* (The Motion of Revolving Bodies) [14].

Newton made a crucial step on the path to his *Principia* when he proved, in 1685, that the gravitational attraction of a spherically symmetric body can be determined by placing all its mass at its center. Newton's "superb theorem" was crucial for an exact solution of the problem of finding the orbit of a planet under the attraction of the Sun (neglecting the attractions between planets).

In the first section of the *Principia* Newton presents his definitions of mass, "quantity of motion" (momentum), and three types of forces: inertial, impressed and centripetal. Following these definitions is a *Scholium* on absolute time, space and motion. Both absolute time and absolute location are quantities that cannot themselves be observed, but instead have to be inferred from measurements of the corresponding relative quantities; such measurements are always provisional. Motion in the solar system, e.g., is referred to the

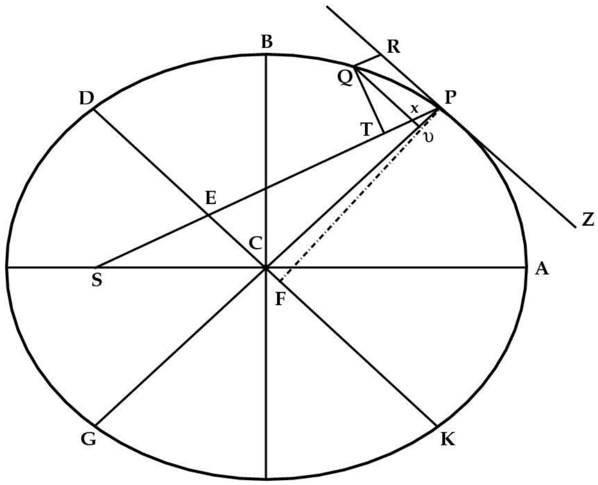

Fig. 1.5 Diagram used in the proof of *Proposition* XI in Book I of Newton's *Principia*. Assuming that the ellipse in the diagram represents an arbitrary trajectory, *Proposition* VI in Book I can also be proved using the same drawing.

fixed stars, and sidereal time is provisionally taken as the preferred approximation to absolute time. "*The causes which distinguish true motions from relative motions are the forces impressed upon bodies to generate motion*". Newton argued that the concave surface of the water in a rotating bucket is created by absolute rotation (acceleration) with respect to absolute space.

The next section contains Newton's laws of motion:

Law 1: *Every body perseveres in its state of rest, or of uniform motion in a straight line, unless it is compelled to change that state by forces impressed*;

Law 2: *A change in motion is proportional to the motive force impressed, and takes place along the straight line in which that force is impressed*;[5]

Law 3: *To every action there is always opposed an equal reaction; or the mutual actions of two bodies upon each other are always equal and directed to contrary parts*.

The main body of the *Principia* is divided into three books. Book I opens with an exposition of the underlying mathematics, a geometrical form of infinitesimal calculus based on the "method of ultimate ratios" of "evanescent quantities" (infinitesimals) [16]. Such quantities are represented by lines and arcs of curves of arbitrary small length, a procedure that had been used by Archimedes for calculating lengths and areas encompassed by curves.

In the rest of the *Principia* Newton analyses the motion of bodies in non-resisting (Book I) and resisting (Book II) media under the action of centripetal

5) The familiar $F = m\mathfrak{a}$ form of Newton's second law can be traced back to Jacob Hermann's *Phoronomia* [15].

forces — which is the central theme of his masterpiece — and applies the results to orbiting bodies, projectiles, pendulums and free-fall under gravity. Newton's law of universal gravitation, which states that *all matter attracts all other matter with a force proportional to the product of their masses and inversely proportional to the square of the distance between them*, is deduced from the astronomical phenomena discussed in Book III. Newton utilises this law to explain tides and their variations, the precession of the Earth's axis, the motion of the Moon as perturbed by the gravity of the Sun, and the eccentric orbits of comets.[6]

Proposition VI in Book I establishes Newton's law of motion under *centripetal forces* and along an arbitrary trajectory, such that equal areas are swept in equal times with respect to the acceleration center, in accord with *Proposition* I in Book I. In Andrew Motte's translation (1729), *Proposition* VI reads: "*In a space void of resistance, if a body revolves in any orbit along an immoveable centre, and in the least time describes any arc just then nascent; and the versed sine of that arc is supposed to be drawn bisecting the chord, and produced passing through the centre of force; the centripetal force in the middle of the arc will be as the versed sine directly and the square of the time inversely.*"

Assume, for the moment, that the ellipse in the diagram in Fig. 1.5 represents an *arbitrary* trajectory. A particle in orbital motion can be thought of as falling freely toward the acceleration center. The central force at P is then proportional to the displacement from the tangent QR over a short increment of time, divided by the square of this time: $\mathrm{F} \propto \mathfrak{a} \propto x/t^2$ (Galileo's time-squared law). Since the time is proportional to the area swept out, which in the limit $\mathrm{Q} \to \mathrm{P}$ is the triangular area $\mathrm{SP} \cdot \mathrm{QT}/2$, the centripetal force must vary along the trajectory as $1/\mathrm{SP}^2$ times $\mathrm{QR}/\mathrm{QT}^2$ (where QR is also called "versed sine"). This proves Newton's *Proposition* VI.

Proposition XI Problem VI reads: "*If a body revolves in an ellipsis; it is required to find the law of the centripetal force tending to the focus of the ellipsis.*" In the ellipse in Fig. 1.5, Newton draws two *conjugate diameters* DK and PG, with DK parallel to the tangent RPZ. From Q he draws three lines: QR parallel to the focal radius SP, QT perpendicular to SP, and Qx completing the parallelogram QxPR. He then extends Qx until it meets PG at v, and draws PF pependicular to DK. Newton's analysis makes use of the following lemmas: (1) $\mathrm{PE} = \mathrm{AC}$; (2) *All parallelograms circumscribed about any conjugate diameters of an ellipse have equal area*; (3) *In an ellipse, the squares of the ordinates of any conjugate diameter are proportional to the rectangles under the segments which they make on the diameter*.

6) Gottfried Leibniz was the first to express Newton's theory of orbital motion in the form of a differential equation [17]. Leonhard Euler synthesized Leibniz's differential calculus with Newton's method of fluxions into *mathematical analysis*.

The first lemma is Newton's and the other two are attributed to Apollonius of Perga (262–190 BC), who is famous for developing the theory of *conic sections*.

Since QR is Px and (by Lemma 1) PE is AC, the similarity of the triangles PxV and PEC implies $\mathrm{QR} = (\mathrm{P}v \cdot \mathrm{AC})/\mathrm{PC}$. On the other hand, the similarity of the triangles QxT and PEF, combined with Newton's lemma, leads to $\mathrm{QT} = (\mathrm{Qx} \cdot \mathrm{PF})/\mathrm{AC} = (\mathrm{Qx} \cdot \mathrm{BC})/\mathrm{CD}$, where the second equality follows from Lemma 2 (which implies $\mathrm{PF} \cdot \mathrm{CD} = \mathrm{BC} \cdot \mathrm{AC}$). It is now straightforward to show that $\mathrm{QT}^2/\mathrm{QR} = (\mathrm{L}/2)\,(\mathrm{Qx}^2 \cdot \mathrm{PC})/(\mathrm{P}v \cdot \mathrm{CD}^2)$; here $\mathrm{L} \equiv 2\mathrm{BC}^2/\mathrm{AC}$ is the *latus rectum* of the ellipse. In the limit $\mathrm{Q} \to \mathrm{P}$, the ratio $\mathrm{Q}v/\mathrm{Qx} \to 1$. Furthermore, Lemma 3 implies $\mathrm{Q}v^2/(\mathrm{P}v \cdot v\mathrm{G}) = \mathrm{CD}^2/\mathrm{PC}^2$. Hence, one finally obtains $\mathrm{QT}^2/\mathrm{QR} = (\mathrm{L}/2)(v\mathrm{G}/\mathrm{PC}) \to \mathrm{L}$ because $v\mathrm{G} \to 2\mathrm{PC}$ as $\mathrm{Q} \to \mathrm{P}$.

This completes Newton's analysis for an elliptic orbit:

$$Centripetal\ force \propto \frac{1}{\mathrm{L} \cdot SP^2}$$

Since L is a constant, the centripetal force is reciprocal to the square of SP, the distance to the focus of the ellipse.[7]

To further illustrate the mathematical reasoning that characterizes the *Principia*, Newton's geometrical proof of *Proposition* LXX in Book I will be presented. The proposition states: "*If to every point of a spherical surface there tend equal centripetal forces decreasing as the square of the distances from those points, I say, that a corpuscule placed within that surface will not be attracted by those forces any way*" [14]. For the analogous case in electrostatics, Newton's proposition was tested by the null experiment of Henry Cavendish (see Fig. 1.10).

Newton's proof of *Proposition* LXX is based on Fig. 1.6. Consider a cone of infinitesimally small solid angle $\mathrm{d}\Omega$ that intersects a homogeneous spherical shell of matter in both directions. The intersection areas around the points Q and Q' are denoted by $\mathrm{d}S$ and $\mathrm{d}S'$, respectively. The apex of the cone is at the position of a test body placed inside the shell (point P). The forces of attraction at P, due to the source-surface areas $\mathrm{d}S$ and $\mathrm{d}S'$, are equal and opposite because: (1) the gravitational force decreases as the square of the distance, while the surface areas of the source, cut out by the cone of a given solid angle $\mathrm{d}\Omega$, increases as the square of the distance; and (2) the angles between the infinitesimally thin two-way cone and the normals to the spherical shell at both intersection points (radial lines OQ and OQ') are equal. Since the entire solid angle around the point P can be divided into such double cones, the resultant attraction is zero.

7) In the first ten sections of the *Principia* Newton speaks of "force", but he actually calculates *accelerations*. In his study of orbital motions, Newton does not comment on the *cause* of gravity; this issue is addressed in the *General Scolium* on the final pages of the *Principia*.

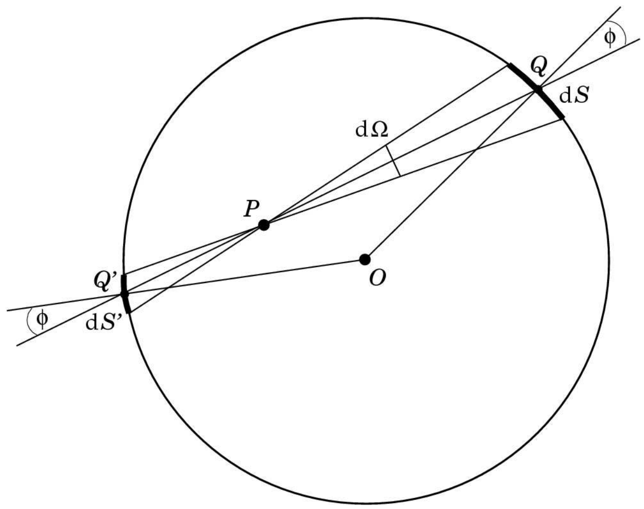

Fig. 1.6 Newton's proof of *Proposition* LXX in Book I. The point O is the center of a homogeneous spherical shell of matter, P is the position of a test body, $d\Omega$ is the solid angle of an infinitesimally thin two-way cone with P at its apex, dS and dS' are intersection areas (around Q and Q') of the two sides of the cone with the spherical shell, and ϕ is the angle between PQ and OQ, which is the same as the angle between PQ' and OQ'.

The proofs of Newton's propositions involve limits, derivatives, integrals, curved paths, acceleration, etc. In other words, the mathematics of the *Principia* is calculus disguised in the form of geometry, as stated earlier.

Newton realized that the motions of Jupiter's moons and of the Earth-Moon system in the Sun's gravitational field would be greatly altered if the observed equality between gravitational and inertial mass did not hold. He demonstrated this equality (known as the *weak equivalence principle*) to a precision of about one part in 10^3 using pendulums made from different materials. A generalization of the weak equivalence principle played a crucial role in the formulation of general relativity.

In the *General Scholium*, added at the end of the second edition of the *Principia*, Newton writes: "*I have not as yet been able to deduce from phenomena the reason for these properties of gravity, and I do not feign hypotheses. For whatever is not deduced from the phenomena must be called a hypothesis; and hypotheses . . . have no place in experimental philosophy. In this experimental philosophy, propositions are deduced from the phenomena and are made general by induction*".

Newton originated the science of spectroscopy by demonstrating that white light is composed of a spectrum of colours, and proposed a corpuscular theory of light. He invented and built a new type of telescope, the design of which is the prototype for all modern, large optical telescopes. In the opinion of Albert Einstein, Newton "*determined the course of western thought, research and practice like no one else before or since*" [18].

1.4 Development of Analytical Mechanics by Euler, Lagrange and Hamilton

Classical mechanics was reformulated by Joseph Lagrange (1736–1813) about a century after the publication of Newton's *Principia*. Whereas in Newtonian mechanics one deals with forces nowadays expressed in terms of vectors, in *Lagrangian mechanics* the dynamics of a system is governed by a scalar function $\mathcal{L}(q_i, \dot{q}_i, t)$, called the *Lagrangian* ($\dot{q}$ is the time derivative of q). The trajectory of the system is determined by solving, for each of the *generalized coordinates* q_i ($i = 1, \ldots, N$), the second-order differential *Euler-Lagrange equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i}\right) - \frac{\partial \mathcal{L}}{\partial q_i} = 0 \qquad \textit{Euler-Lagrange equation} \tag{1.1}$$

For a non-relativistic particle in one dimension, $\mathcal{L} \equiv \mathrm{E}_k - \mathrm{E}_p = m\dot{x}^2/2 - V(x)$; here E_k and E_p are the kinetic and potential energies, respectively, and $V(x)$ is a *potential function*. In this simple case, the Euler-Lagrange equation (1.1) yields *Newton's second law*: $\frac{\mathrm{d}}{\mathrm{d}t}(m\dot{x}) - \frac{\partial}{\partial x}(-V) = 0$, that is, $m\ddot{x} = F(x)$.

The Lagrangian formalism, which can easily be extended to describe systems outside the realm of Newtonian mechanics, is one of the mathematical cornerstones of modern theoretical physics. In the preface to his *Méchanique Analytique*, published in 1788, Lagrange proudly declares: "*The methods that I expound require neither constructions, nor geometrical or mechanical arguments, but only algebraic operations, subject to a regular and uniform course*" [19].

In an alternative formulation of classical mechanics, proposed in 1834 by William Rowan Hamilton (1805–1865), the *generalized coordinates* q_i and the *conjugate momenta* p_i appear on what is an essentially equal footing [20]:

$$\dot{q}_i = \frac{\partial \mathcal{H}}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i} \qquad \textit{Hamilton's equations} \tag{1.2}$$

The *Hamiltonian* $\mathcal{H}$ is the *Legendre transform* of the Lagrangian $\mathcal{L}$:

$$\mathcal{H}(q_i, p_i, t) \equiv \sum_i p_i \dot{q}_i - \mathcal{L}(q_i, \dot{q}_i, t) \quad \Leftrightarrow \quad \mathcal{L} = \mathbf{p} \cdot \dot{\mathbf{q}} - \mathcal{H}(\mathbf{q}, \mathbf{p}, t) \tag{1.3}$$

where $\sum_i^n p_i \dot{q}_i = p_1 \dot{q}_1 + p_2 \dot{q}_2 + \cdots + p_n \dot{q}_n \equiv \mathbf{p} \cdot \dot{\mathbf{q}}$ is the *scalar product* of two *vectors*. A non-relativistic particle in one dimension has the total energy $\mathcal{H}(x, p) \equiv \mathrm{E}_k + \mathrm{E}_p = p^2/2m + V(x)$. First-order *Hamilton's equations* (1.2) in this case yield $\dot{x} = p/m$ and $\dot{p} = -\partial V/\partial x$; hence, $m\ddot{x} = -\partial V/\partial x$.

The Hamiltonian formalism provides the most straightforward link between classical mechanics and quantum theory. In Paul Dirac's opinion: "*Hamilton seemed to have some remarkable insight into what was important — one of the most remarkable insights, I suppose, which a mathematician has ever had. He found a form of writing the equations of mechanics whose importance would be realized only after a hundred years*" [21].

The *calculus of variations* — originally invented by Newton while trying to determine the shape of a moving object that would guarantee the least possible resistance — traces its inception to 1696, the year when Johann Bernoulli (1667–1748) formulated the *brachistochrone problem* (see Subsection 1.4.1). The first major contributions to this venerable branch of mathematics were made by Leonhard Euler (1707–1783) and Joseph Lagrange. The development of the calculus of variations has served as a catalyst for major advances in various fields of mathematics, such as functional analysis and topology [22].

"*Once the laws of physical theory are expressed as differential equations, the possibilty of their reduction to a variational principle is evident from purely mathematical reasoning* ..." [23].

A *variational principle* is a general rule in physics expressed in terms of the calculus of variations; it states that some functional of dynamical variables (called *action*) is stationary with respect to small variations of these variables. The *action* S of a system is an integral of its Lagrangian:

$$S[q(t)] \equiv \int \mathcal{L}(q_i, \dot{q}_i, t)\mathrm{d}t \tag{1.4}$$

The Euler-Lagrange equations (1.1) follow from the variational principle

$$\delta S = 0 \qquad\qquad Hamilton's\ principle \tag{1.5}$$

This principle states that, out of all possible "paths" in *configuration space* $\{q_1 \cdots q_n\}$, a particle will travel along the path for which the value of the integral (1.5) is the same (to first order in δq) as that for any path which differs from it by an infinitesimal displacement [20].

It is straightforward to prove that the Euler-Lagrange equations follow from Hamilton's principle. The variation of the action reads

$$\delta S = \int_{t_1}^{t_2} \mathrm{d}t\, \delta\mathcal{L} = \int_{t_1}^{t_2} \mathrm{d}t \left\{ \frac{\partial \mathcal{L}}{\partial q_i(t)}\, \delta q_i(t) + \frac{\partial \mathcal{L}}{\partial \dot{q}_i(t)}\, \delta \dot{q}_i(t) \right\}$$

Writing $\delta\dot{q}_i(t) = \frac{\mathrm{d}}{\mathrm{d}t}\,\delta q_i$, the second term can be integrated by parts:

$$\int_{t_1}^{t_2} \mathrm{d}t\, \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \frac{\mathrm{d}}{\mathrm{d}t}\, \delta q_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i}\, \delta q_i \Bigg|_{t_1}^{t_2} - \int_{t_1}^{t_2} \mathrm{d}t\, \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathcal{L}}{\partial \dot{q}_i}\, \delta q_i$$

But $\delta q_i(t_1) = \delta q_i(t_2) = 0$, and so the first term on the right is zero. Thus,

$$\delta S = \int_{t_1}^{t_2} \left\{ \frac{\partial \mathcal{L}}{\partial q_i} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right\} \delta q_i \mathrm{d}t$$

which vanishes provided that (1.1) holds.

At first sight it looks as if Hamilton's principle ascribes to the particle a fair amount of 'intelligence': before it moves from one point in the configuration space to another, a particle 'calculates' the value of S for every possible path linking these points, and then follows the one for which S is a minimum. This is not what really happens: the particle merely obeys the Euler-Lagrange equations of motion (i.e., Newton's laws) at each point to minimize the action, which means that every subsection of the actual path must be a minimum. In other words, the particle somehow 'probes' the potential within an infinitesimal region of spacetime, and moves in the direction of greatest change.

In general, $\mathcal{L}$ does not have the form of the kinetic energy minus the potential energy. The actual form of the Lagrangian for any particular case is a pure guess! The quantity defined by $\tilde{p}_i \equiv \partial\mathcal{L}/\partial\dot{q}_i$ is called the *generalized momentum*, conjugate to q_i. As the name implies, $\tilde{p}_i$ is not always the familiar linear momentum of a particle. For instance, the conjugate momentum of an angular degree of freedom is an angular, rather than a linear, momentum.

If the Lagrangian of a system does not contain a coordinate q_i (although it may contain $\dot{q}_i$), then (1.1) implies that the generalized momentum conjugate to q_i is conserved. For a system that is *invariant*, e.g., under translation along a given direction, the corresponding linear momentum is conserved. Thus, momentum conservation is closely connected with the *symmetry properties* of the system.

The most compelling reason for studying Hamiltonian dynamics is to acquire a deeper insight into the structure of classical mechanics and its relation to other branches of physics. The relation between the *Hamilton-Jacobi theory* of classical dynamics (see below) and the short-wavelength limit of wave optics is a bridge between classical and quantum mechanics. The Hamilton-Jacobi theory was originally expounded by Hamilton in his second essay [20], and was further developed by Carl Gustav Jacobi (1804–1851) [24].

A transformation is said to be *canonical* if the new conjugate variables $P_i(q, p, t)$ and $Q_i(q, p, t)$ satisfy Hamilton's equations

$$\dot{P}_i = -\frac{\partial\widetilde{\mathcal{H}}}{\partial Q_i}, \qquad \dot{Q}_i = \frac{\partial\widetilde{\mathcal{H}}}{\partial P_i}$$

which implies that $\delta \int [\mathbf{p}\cdot\dot{\mathbf{q}} - \mathcal{H}(\mathbf{q}, \mathbf{p}, t)]\mathrm{d}t = 0 = \delta \int [\mathbf{P}\cdot\dot{\mathbf{Q}} - \widetilde{\mathcal{H}}(\mathbf{Q}, \mathbf{P}, t)]\mathrm{d}t$. For this to hold simultaneously, the integrands should be equal. Since a total time derivative of an arbitrary function $G(q, p, Q, P, t)$ can be added to either integrand without affecting the variations ($\int_{t_1}^{t_2} \mathrm{d}G = const.$), one can write

$$\sum p_i\dot{q}_i - \mathcal{H} = \sum P_i\dot{Q}_i - \widetilde{\mathcal{H}} + \frac{\mathrm{d}G}{\mathrm{d}t}$$

The *generating function* G is introduced in order to render the transformation non-trivial. Depending on which pair of canonical variables are considered as *independent*, canonical transformations can be classified into four basic types. Substituting, e.g, $G \equiv F(q, P, t) - \sum Q_i P_i$ in the last equation, one obtains

$$\sum_{i=1}^{n}\left[\left(p_i - \frac{\partial F}{\partial q_i}\right)\mathrm{d}q_i + \left(Q_i - \frac{\partial F}{\partial P_i}\right)\mathrm{d}P_i\right] + \left(\widetilde{\mathcal{H}} - \mathcal{H} - \frac{\partial F}{\partial t}\right)\mathrm{d}t = 0$$

Since all the (q, P) are independent, their coefficients must vanish independently. This yields the *canonical transformation equations*[8]

$$p_i = \frac{\partial F(q_i, P_i, t)}{\partial q_i}, \qquad Q_i = \frac{\partial F(q_i, P_i, t)}{\partial P_i}, \qquad \widetilde{\mathcal{H}} = \mathcal{H} + \frac{\partial F}{\partial t}$$

It is evident that (Q, P) will be *constants* if the new Hamiltonian $\widetilde{\mathcal{H}}$ vanishes. One then obtains the following first-order differential equation

$$\mathcal{H}\left(q_i, \frac{\partial F}{\partial q_i}, t\right) + \frac{\partial F}{\partial t} = 0 \qquad \textit{Hamilton-Jacobi equation} \qquad (1.6)$$

The function F, customarily denoted by S, is known as *Hamilton's principal function*. Since P_i is constant ($P_i \equiv \alpha_i$), the total time derivative of S reads

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \sum \frac{\partial S}{\partial q_i}\dot{q}_i + \frac{\partial S}{\partial t} = \sum p_i \dot{q}_i - \mathcal{H} \equiv \mathcal{L}$$

Therefore, S differs from the indefinite time integral of $\mathcal{L}$ only by a constant: $S = \int \mathcal{L}\mathrm{d}t + constant$. If $\mathcal{H}$ does not depend explicitly on time, one can write $S(q, \alpha, t) = W(q, \alpha) - \mathrm{E}t$, where $W(q, \alpha)$ is *Hamilton's characteristic function*.

1.4.1 The Brachistochrone: Gateway to the Calculus of Variations

In the June 1696 issue of the journal *Acta Eruditorum*, Johann Bernoulli posed the following mathematical problem:

- *Find the curve connecting two points, at different heights and not on the same vertical line, along which a body acted upon only by gravity will fall in the shortest possible time.*

Bernoulli named the curve *brachistochrone*, a term he coined from the Greek words *brachistos* (shortest) and *chronos* (time). The May 1697 issue of *Acta Eruditorum* contained four solutions to this problem, provided by the greatest mathematicians of the time: Isaac Newton, Gottfried Leibniz (1646–1716), Johann Bernoulli, and his elder brother Jacob Bernoulli (1654–1705). Their answers all agreed, but the methods of derivation differed considerably [25].

8) Consider a specific F, for instance $F(q, P, t) = \sum q_i P_i$. In this case it is easy to show that $P_i = p_i$, $Q_i = q_i$ and $\widetilde{\mathcal{H}} = \mathcal{H}$. Thus, the identity transformation is also a canonical transformation.

Johann Bernoulli's solution is of particular interest because it represents the earliest known example of the analogy between optics and mechanics [26]. The solution utilizes *Fermat's principle of least time*, which is equivalent to the optical law of refraction $\sin\alpha_1/v_1 = \sin\alpha_2/v_2$ (*Snell's law*); here α_1 and α_2 are the angles of incidence and refraction, respectively, and (v_1, v_2) are the velocities of light in the two media (see Section 1.5).

Consider a ray of light propagating through a sequence of horizontal, thin layers of decreasing optical density. As the speed of light increases, so does the angle of refraction; that is, the ray of light follows a curved path. This inspired Bernoulli to come up with the following optical-mechanical analogy: *The constantly increasing speed of a body sliding down a brachistochrone curve corresponds to a light ray propagating along a curved optical path through an optical medium of ever-decreasing optical density*.

For a very large number of infinitesimally thin layers, Snell's law can be expressed as $\sin\alpha/v = \kappa$, where κ is constant. The speed of a body that falls from rest starting at $y = 0$ is $v = gt$, and the distance of fall is $y = gt^2/2$ (Galileo's time-squared law), where g is the acceleration due to gravity; hence, $v = \sqrt{2gy}$. Combining this expression with Snell's law gives $y = k\sin^2\alpha$, where $k = 1/2g\kappa^2$. If v is the velocity at a point (x, y), then $y' = \mathrm{d}y/\mathrm{d}x$ and $\sin^2\alpha = 1/(1 + y'^2)$. From this result and $y = k\sin^2\alpha$ follows a *differential equation* describing the motion of the body: $y(1 + y'^2) = 2h$, where $h = k/2$.

The body's trajectory is a *cycloid* described by the parametric equations $x(\theta) = h(\theta - \sin\theta)$ and $y(\theta) = h(1 - \cos\theta)$, as can be readily verified by direct substitution in the differential equation derived above. A cycloid curve represents the trajectory of a point on the circumference of a wheel that rolls without slipping along the x-axis. To completely solve the brachistochrone problem, one has to find the right cycloid that connects the two given points. This task was first accomplished by Newton. His elegant solution (adopted by Johann Bernoulli) originally appeared in the journal *Philosophical Transactions*, and was later published in the May 1697 issue of *Acta Eruditorum* [25]. "*It is, however, interesting to speculate on why Newton gave no clue in his paper as to how he knew the curve was cycloidal* ..." [26].

Johann Bernoulli was greatly impressed that the solution of his brachistochrone problem was the same as that of the following *tautochrone problem* (*tautos* means identical): *Determine the curve for which the time taken by an object sliding without friction in uniform gravity to its lowest point is independent of its starting point*. The curve is a cycloid, and the time is equal to π times the square root of the radius over the acceleration of gravity. The tautochrone problem was solved geometrically in 1659 by Christiaan Huygens (1629–1695), who found that a cycloid was the curve along which a pendulum would have to descend to be exactly isochronous [25].[9]

9) The brachistochrone problem can also be solved geometrically [27].

The novelty of the brachistochrone problem lies in the fact that the quantity under consideration, the time of descent, depends on the *whole curve*. This problem stimulated the development of ideas and techniques that led to the branch of mathematics known as the *calculus of variations*.

The calculus of variations has many applications in physics. The laws of motion and of equilibrium are dominated by maximum and minimum principles. For instance, stable equilibrium of a mechanical system is attained if the system is arranged in such a way that its potential energy is a minimum. Euler and Lagrange were the first to develop "*more general methods for solving extremum problems in which the independent element was ... a whole curve or function or even a system of functions*" [28].

In 1744, Euler published a seminal treatise entitled *Methodus Inveniendi* [29], in which he presented a mathematical method for solving a general class of variational problems. In Appendix II of the treatise, Euler formulated his *principle of least action* (see Section 1.5). Given a function $y(x)$ satisfying the boundary conditions $y(x_A) = y_A,\ y(x_B) = y_B$ and $x_A < x_B$, he showed that

$$\frac{\partial f}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) = 0$$

is a necessary condition for the definite integral $I = \int_A^B f(y, y', x)\mathrm{d}x$ to be a minimum. To derive this *Euler-Lagrange equation*, Euler replaced the integral I by the sum $S = \sum_j f(y_j, z_j, x_j)(x_{j+1} - x_j)$, where $z_j \equiv (\Delta y/\Delta x)_{x=x_j}$, and asked for the stationary value of S. This problem can be solved by setting to zero the partial derivatives of S with respect to y_k. The resulting difference equation becomes a differential equation in the limit $\Delta x_j \to 0$ [26,30].

In 1755, Joseph Lagrange (born Giuseppe Lodovico Lagrangia) was only nineteen years old when he sent a letter [31] to Euler explaining how Euler's tedious geometrical reasoning could be replaced by a method that only requires a straightforward use of the principles of the differential and integral calculus. Lagrange's first published account of the calculus of variations [32] appeared in 1760. In the first essay [32] he formulated a new calculus based on the symbol δ (see Eq. (1.5)). This enabled him to solve the following general problem: *given an expression involving several variables and their derivatives, find the functional relation between these variables which renders the definite integral of the given expression an extremum*. The second essay [32] contains an extensive application of variational techniques to the principle of least action in dynamics. The publication of his *Méchanique Analytique* marks the culmination of Lagrange's early work on the foundations of the calculus of variations.

"*Euler's variational principles of physics, rediscovered and extended by* [*Hamilton*], *have proved to be among the most powerful tools in mechanics, optics and electrodynamics ... Recent developments in physics — relativity and quantum theory — are full of examples revealing the power of the calculus of variations*" [28].