Revisiting the mathematical synthesis of the laws of Kepler and Galileo leading to Newton’s law of universal gravitation

Wu-Yi Hsiang  
Department of Mathematics  
University of California, Berkeley

Eldar Straume  
Department of Mathematical Sciences  
Norwegian University of Science and Technology, Trondheim, Norway

August 29, 2014

Abstract

Newton’s deduction of the inverse square law from Kepler’s ellipse and area laws together with his “superb theorem” on the gravitation attraction of spherically symmetric bodies, are the major steps leading to the discovery of the law of universal gravitation (Principia, 1687). The goal of this article is to revisit some “well-known” events in the history of science, and moreover, to provide elementary and clean-cut proofs, still in the spirit of Newton, of these major advances. Being accessible to first year university students, the educational aspect of such a coherent presentation should not be overlooked.

Contents

1 Introduction 2

2 Prelude 3
   2.1 Astronomy and geometry of the antiquity up to the Renaissance 3
   2.2 Copernicus, Tycho de Brahe, and Kepler: The new astronomy . . . 4
   2.3 Galileo’s empirical laws on terrestrial gravity, as evidence for the inertia and force laws ......................... 6
   2.4 Equivalence between the area law and the action of a centripetal force ................................................. 6
1 Introduction

Towards the later half of the 16th century, the Renaissance of Greek civilization in Europe had paved the way for major advancements leading to the creation of modern science. In astronomy, the heliocentric theory of Copernicus and the systematic astronomical observational data of Tycho de Brahe led to the discovery of the magnificent empirical laws of Kepler on planetary motions. In physics, the empirical laws of Galileo on terrestrial gravity (i.e., free falling bodies and motions on inclined planes) laid the modern foundation of mechanics. Then, in the later half of the 17th century, the mathematical synthesis of those empirical laws led Isaac Newton (1643–1727) to the discovery of the law of universal gravitation, published in his monumental treatise *Philosophiae Naturalis Principia Mathematicae* (1687), divided into three books, often referred to as *Principia* for brevity.

Note that, at this juncture of scientific advancements, it is the mathematical analysis as well as the synthesis of empirical laws that play the unique role leading towards the underlying fundamental laws. Among them, the first major step is the deduction of the inverse square law from Kepler’s area law and ellipse law, which was first achieved by Newton most likely sometimes in 1680–81, and another most crucial step is the proof of the integration formula for the gravitation attraction of spherically symmetric bodies, nowadays often referred to as the “superb theorem”. In *Principia*, these major results are respectively stated as Proposition 11 and Proposition 71 of Book I, where Newton deals with the laws of motion in vacuum.

Certainly, Newton’s celebrated treatise is widely worshipped also today. However, his proofs of the two most crucial propositions are quite difficult to understand, even among university graduates. Therefore, a major goal of this article is to provide elementary and clean-cut proofs of these results.

In the Prelude we shall briefly recall the major historical events and the central figures contributing to the gradual understanding of planetary motions prior to Newton. Moreover, the significance of the area law is analyzed in
the spirit of Newton’s, but in modern terms of classical mechanics, for the convenience of the reader.

Section 3 is devoted to a review of the focal and analytic geometry of the ellipse, the type of curves playing such a predominant role in the work of Kepler and Newton. Of particular interest is the curvature formula, which Newton must have known, but for some reason appears only implicitly in his proof of the inverse square law. In Section 4 we shall give three simple different proofs of this law, allowing the modern reader to view Newton’s original and rather obscure proof with modern critical eyes. Moreover, Newton’s original and geometric proof of his “superb theorem” is not easy reading, which maybe explains the occurrence of many new proofs in the more recent literature. We believe, however, the proof given in Section 5 is the simplest one, in the spirit of Newton’s proof but based on one single new geometric idea.

Finally, for the sake of completeness, in Section 6 we shall solve what is called the Kepler problem in the modern terminology, but was in fact referred to as the “inverse problem” at the time of Newton and some time after, see [15]. We present two proofs, one of them is rather of standard type, reducing the integration to the solution of Binet’s equation, which in this case is a well known and simple second order ODE in elementary calculus with \( \sin(x) \) and \( \cos(x) \) as solutions. But there is an even simpler proof, where the integration problem is just finding the antiderivatives of these two functions.

One may contend that "simple" proofs of the above classical problems are nowadays found in a variety of calculus textbooks or elsewhere, so is there anything new at all? The original motivation for the present article came, however, from reading Chandrasekhar [3], which encouraged us to write a rather short but coherent presentation accessible for the "common reader", in a historical perspective and in the spirit of Newton’s original approach.

2 Prelude

2.1 Astronomy and geometry of the antiquity up to the Renaissance

The ancient civilization of Egypt and Babylon had already accumulated a wealth of astronomical and geometric knowledges that those great minds of Greek civilization such as Thales (ca. 624–547 BC), Pythagoras (ca. 569–475 BC) etc. gladly inherited, studied and deeply reflected upon. In particular, Pythagoras pioneered the philosophical belief that the basic structures of the universe are harmonious and based upon simple fundamental principles, while the way to understand them is by studying numbers, ratios, and shapes. Ever since then, his remarkable philosophical foresight still inspires generations after generations of rational minds.

Following such a pioneering philosophy, the Pythagoreans devoted their studies to geometry and astronomy and subsequently, geometry and astronomy became the two major sciences of the antiquity, developing hand in hand. For
example, those great geometers of antiquity such as Eudoxus (408–355 BC), Archimedes (287–212 BC), and Apollonius (ca. 262–190 BC), all had important contributions to astronomy, while those great astronomers of antiquity such as Aristarchus (ca. 310–230 BC), Hipparchus (190–120 BC), and Ptolemy (ca. 85–165 AD) all had excellent geometric expertise. We mention here three well-known treatises that can be regarded as the embodiment of the glory of scientific achievements of the antiquity, namely

- Euclid’s *Elements* (13 books)
- Apollonius: *Conics* (8 books)
- Ptolemy: *Almagest* (13 books)

### 2.2 Copernicus, Tycho de Brahe, and Kepler: The new astronomy

Today, it is a common knowledge that the Earth is just one of the planets circulating around the sun. But this common knowledge was, in fact, the monumental achievement of the new astronomy, culminating the successive life-long devotions of Copernicus (1473–1543), Tycho de Brahe (1546–1601), and Kepler (1571–1630). Kepler finally succeeded in solving the problem on planetary motions (see below) that had been puzzling the civilization of rational mind for many millenniums.

In the era of the Renaissance, Euclid’s *Elements* and Ptolemy’s *Almagest* were used as important textbooks on geometry and astronomy at major European universities. At Bologna University, Copernicus studied deeply Ptolemy’s *Almagest* as an assistant of astronomy professor Navara, and both of them were aware of *Almagest’s* many shortcomings and troublesome complexities. In 1514, inspired by the account of Archimedes on the heliocentric theory of Aristarchus, he composed his decisive *Commentary* (1515), outlining his own heliocentric theory which was finally completed as the book *De revolutionibus orbium coelestium* (1543). Nowadays, this is commonly regarded as the heralding salvo of the modern scientific revolution.

Note that a creditable astronomical theory must pass the test of accurate predictions of verifiable astronomical events, such as observable events on planetary motions. However, just a qualitatively sound heliocentric model of the solar system would hardly be accepted as a well-established theory of astronomy. Fortunately, almost like a divinely arranged “relay in astronomy”, the most diligent astronomical observer Tycho de Brahe, with generous financial help from the King of Denmark and Norway, made twenty years of superb astronomical observations at Uraniborg on the island of Hven, and subsequently, Johannes Kepler became his assistant (1600–1601), and moreover, succeeded him as Imperial Mathematician (of the Holy Roman Empire, in Prague) after the sudden death of Tycho de Brahe. With Kepler’s superb mathematical expertise and marvelous creativity, it took him 20 years of hard work and devotion
to finally succeed in solving the millennium puzzle of planetary motions, namely the following remarkable Kepler’s laws, which we may state as follows:

**Kepler’s first law (the ellipse law):** The planets move on elliptical orbits with the sun situated at one of the foci.

**Kepler’s second law (the area law):** The area per unit time sweeping across by the line interval joining the planet to the sun is a constant, as illustrated by Figure 1.

**Kepler’s third law (the period law):** The ratio between the cube of the major axis and the square of the period, namely \((2a)^3/T^2\), is the same constant for all planets.

The following are the major publications of Kepler on his new astronomy:

- Astronomia Nova (1609)
- 3 volumes of Epitome of Copernican astronomy (1618-1621)
- Harmonice Mundi (1619)
- Tabulae Rudolphinae (1627)

First of all, the predictions of the Rudolphine tables turned out to be hundred times more accurate than that of the others. Moreover, Kepler predicted the Mercury transit of Nov. 7, 1631 (which was observed in Paris by P. Gassendi), and the Venus transit of Dec. 7, 1631 (that could not be observed in Europe), while the next Venus transit would only occur after another 130 years. Here, we would like to mention the remarkable achievements of J. Horrocks (1618–1641). This brilliant young man was already fully in command of Kepler’s new astronomy at the age of 20, and after correcting Kepler’s tables, he realized...
that a transit of Venus would occur already on Nov. 24, 1639. His subsequent observation on the predicted date, which he reported in *Venus in Sole Visa*, was found in 1659, and this is a noticeable triumph of the new astronomy.

2.3 Galileo’s empirical laws on terrestrial gravity, as evidence for the inertia and force laws

During his years at the University of Pisa (1589–92), Galileo Galilei (1564–1642) wrote *De Motu*, a series of essays on the theory of motion (containing some mistakes, but was never published). Perhaps his most important new idea in *De Motu* is that one can test theories by conducting experiments, such as testing his theory on falling bodies using an inclined plane to vary the rate of descent.

In the years 1602–04 at Padua, he had returned to the dynamical study of terrestrial gravity by conducting experiments on the inclined plane and the pendulum. He had, by then, formulated the correct law of falling bodies and worked out that a projectile follows a parabolic path. However, these important results that laid the foundation of modern mechanics were only published 35 years later in *Discourses and mathematical demonstrations concerning the two new sciences*. Here one finds the origin of the law of inertia, in the sense that Galileo’s conception of inertia is tantamount to *Newton’s first law* of motion. Furthermore, Galileo’s experiments on falling bodies pointed toward the general force law (i.e. $F = ma$), namely *Newton’s second law*, which was certainly also known to Newton’s contemporaries Huygens, Halley, and Hooke. However, the notion of “force” was, to some extent, already present in the work of Archimedes on statics. On the other hand, *Newton’s third law* concerning mutually interacting forces, was a major innovation due to himself, which we shall return to in Section 7.

2.4 Equivalence between the area law and the action of a centripetal force

According to B. Cohen (cf. [5], pp.167–169), a decisive step on the path to universal gravity came in late 1679 and early 1680, when Robert Hooke (1635-1703) introduced Newton to a new way of analyzing motion along a curved trajectory, cf. Koyré [8]. Hooke had cleverly seen that the motion of an arbitrary body can be regarded as the combination of an inertia component and a centripetal component. But he was unable to express this in a more precise mathematical language. However, the possible influence of Hooke on Newton’s *Principia* still engages many scientists and historians, cf. e.g. Purrington [11].

The terms “centripetal force” and “radial force” will be used interchangeably. The very first proposition of the *Principia* develops the dynamical significance of the law of areas by proving the mathematical equivalence between the area law and the centripetality of the force, using Hooke’s technique. We include here a slight simplification of Newton’s proof in terms of modern terminology,
Theorem 2.1 Let \( \overrightarrow{OP} \) be the position vector of a point mass at \( P \) moving in a plane, and let \( \frac{dA}{dt} \) be the area swept out by \( \overrightarrow{OP} \) per unit time. Then \( \frac{dA}{dt} \) is a constant if and only if \( \overrightarrow{OP} \) and the acceleration vector \( \overrightarrow{a} \) are collinear, namely the force is centripetal.

Proof. The velocity vector \( \overrightarrow{v} = \frac{d}{dt} \overrightarrow{OP} \) and the acceleration vector \( \overrightarrow{a} = \frac{d^2}{dt^2} \overrightarrow{OP} \) lie in the plane of the motion, with unit normal vector \( \overrightarrow{n} \), say. Then we can write
\[
2 \frac{dA}{dt} = (\overrightarrow{OP} \times \overrightarrow{v}) \cdot \overrightarrow{n} \tag{1}
\]
On the other hand
\[
\frac{d}{dt}(\overrightarrow{OP} \times \overrightarrow{v}) = \overrightarrow{v} \times \overrightarrow{v} + \overrightarrow{OP} \times \overrightarrow{a} = \overrightarrow{OP} \times \overrightarrow{a}, \tag{2}
\]
and by differentiating both sides of equation (1), we conclude that \( \frac{dA}{dt} \) is a constant if and only if \( \overrightarrow{OP} \times \overrightarrow{a} = 0 \), which simply means \( \overrightarrow{OP} \) and \( \overrightarrow{a} \) are collinear.

Remark 2.2 By (2), centripetality of the force acting on \( P \) means the vector \( \overrightarrow{OP} \times \overrightarrow{v} \) is constant during the motion, and clearly it is also normal to the motion. In particular, centripetality implies the point moves in a plane. However, in the above theorem the meaning of “area swept out” needs no further explanation since the motion is by assumption confined to a fixed plane.

Remark 2.3 Let \((r, \theta)\) be polar coordinates centered at the point \( O \), hence \( r = |\overrightarrow{OP}| \) and \( \dot{\theta} = \frac{d\theta}{dt} \) is the angular velocity. Then the quantity \( \frac{dA}{dt} \) expresses as
\[
2 \frac{dA}{dt} = r^2 \dot{\theta} = k \tag{3}
\]
In particular, for a planet whose trajectory is the ellipse with the sun at the focal point \( O \), it follows from Kepler’s area law that the quantity (3) is the constant \( \frac{2\pi ab}{T} \).

Remark 2.4 (i) In Book 1 of Principia, Newton paid much attention to centripetal forces \( F(r) \), asking two natural questions, namely (i) for a given trajectory curve, what is the attracting force \( F(r) \), and conversely, (ii) for a given force law such as \( F(r) \sim r^n \), \( n = 1, -2, -3, -5 \), what curves are the corresponding trajectories? For a few decades the problems were, respectively, referred to as the “direct” and “inverse” (Kepler) problem (cf. Speiser [16]), which is rather peculiar since the terms “direct” and “inverse” later became switched, and this is also the modern terminology.

(ii) Making a leap forward to J.P.B. Binet (1786–1856), the Binet equation
\[
F(q^{-1}) = -mk^2q^2 \left( \frac{d^2q}{dt^2} + q \right), \quad q = 1/r, \tag{4}
\]
is providing a unifying approach to the central force problem. We shall illustrate
its usage by applying it both to the inverse Kepler problem (in Section 4) and
the Kepler problem (in Section 6). For studies of the inverse Kepler problem in
the physics literature, see for example Ram [12] and Sivardi [15].

In what follows, the notation $PQ$ is used both for the segment between
points $P$ and $Q$ and the length of the segment.

3 On the focal geometry of the ellipse

In Greek geometry, the shape of ellipses first occurred as the tilted plane sec-
tions of a circular cylinder, as indicated in Figure 3, while “ellipse” means
“non-circular” or distorted circle. However, the discovery of its remarkable
geometric characterization greatly excited the enthusiasm of studying such a
natural generalization of circular shapes, namely

**Theorem 3.1** An ellipse $\Gamma$ has two foci $\{F_1, F_2\}$ such that the sum of $PF_1$
and $PF_2$ is equal to a constant for all points $P$ on $\Gamma$.

**Proof.** Referring to Figure 3, $Z$ is a circular cylinder cut by a plane $\Pi$ and
$\Gamma = Z \cap \Pi$ is the plane section. Let $\Sigma_1$ (resp. $\Sigma_2$) be the spheres of the same
radius, inscribed and tangent to $Z$, which are tangent to $\Pi$ at $F_1$ (resp. $F_2$).
Then, for any $P \in \Gamma$, one has $PF_i = PQ_i$, and hence

$$PF_1 + PF_2 = PQ_1 + PQ_2 = Q_1Q_2 = \text{constant} \quad (5)$$

$\square$

3.1 The optical property of the ellipse

**Theorem 3.2** Let $P$ be a point on an ellipse $\Gamma$ with $\{F_1, F_2\}$ as the pair of foci.
Then, the tangent $T_P$ (resp. normal $N_P$) bisects the outer (resp. inner) angle
of $\Delta F_1P F_2$ at $P$.

**Proof.** Let $l$ be the bisector of the outer angle of $\Delta F_1P F_2$ at $P$, and $Q$ be
another point on $l$. As indicated in Figure 3, $F_2'$ is the reflection point of $F_2$
w.r.t. $l$. Then $QF_2 = QF_2'$ and hence

$$QF_1 + QF_2 = QF_1 + QF_2' > F_1F_2 = PF_1 + PF_2,$$

where the last identity follows by considering the angles at $P$, showing that $P$
must, in fact, lie on the line through $F_1$ and $F_2$. Thus, $Q$ must be outside of $\Gamma$,
meaning that $l = T_P$ (i.e. $l \cap \Gamma = \{P\}$). Now the statement about the normal
$N_P$ follows immediately. $\square$

As usual, we shall always denote the constant $PF_1 + PF_2$ of a given ellipse
$\Gamma$ by $2a$, the distance between $F_1$ and $F_2$ by $2c$, and $b = \sqrt{a^2 - c^2}$. The sign //
reads “is parallel to”. 
Figure 2: Ancient geometric proof of (5)
Figure 3: Illustration of the optical property of the ellipse
Corollary 3.3 Let $d_1$ (resp. $d_2$) be the distance between $F_1$ (resp. $F_2$) and a tangent line $T_P$. Then

$$d_1d_2 = b^2.$$  \hspace{1cm} (6)

**Proof.** Let $\{F'_1, F'_2\}$ be the reflection points of $\{F_1, F_2\}$ w.r.t. $T_P$, see Figure 4. Then $F_1F'_2$ and $F'_1F_2$ have length $2a$ and intersect at $P$. By the Pythagorean Theorem, applied to $\Delta F_1F'_2H$ and $\Delta F'_1F_2H$, one has

$$4a^2 = (d_1 + d_2)^2 + (F'_2H)^2$$
$$4c^2 = (d_1 - d_2)^2 + (F'_2H)^2$$

and the identity (6) follows from this. $\blacksquare$

Corollary 3.4 As indicated in Figure 3, if $K$ is the point on $PF_1$ so that $OK \parallel T_P$, then $PK = a$.

**Proof.** Let $E$ be the point on $P F_1$ so that $P E = P F_2$. Then $F_2E \perp V_P \Rightarrow F_2E \parallel T_P \parallel OK$ and hence $F_2K = EK$ and

$$2a = PF_1 + PF_2 = (PK + EK) + PE = 2PK.$$ $\blacksquare$

3.2 A remarkable formula for the curvature of the ellipse

**Theorem 3.5** As illustrated in Figure 3, set $\varepsilon$ to be the angle between $PF_1$ and $T_P$, and $\rho$ to be the radius of the osculating circle of $\Gamma$ at $P$ (i.e. the radius of curvature). Then

$$\kappa = \frac{1}{\rho} = \frac{a}{b^2} \sin^3 \varepsilon$$  \hspace{1cm} (7)

**Proof.** Let us begin with a pertinent fact on circular motions which led Hooke to grasp the dynamical significance of curvature. A circular motion with radius $\rho$ can be represented by

$$\begin{cases} x = \rho \cos \theta(t) \\ y = \rho \sin \theta(t) \end{cases} \text{ or } \overrightarrow{OP} = \rho \begin{pmatrix} \cos \theta(t) \\ \sin \theta(t) \end{pmatrix}$$

Thus, using Newton’s notation of $\dot{\theta} = \frac{d\theta}{dt}$, etc.,

$$\begin{align*}
\overrightarrow{v} &= \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \rho \dot{\theta} \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix}, \quad |\overrightarrow{v}|^2 = (\rho \dot{\theta})^2 \\
\overrightarrow{a} &= \begin{pmatrix} \ddot{x} \\ \ddot{y} \end{pmatrix} = \rho \ddot{\theta}^2 \begin{pmatrix} -\cos \theta \\ -\sin \theta \end{pmatrix} + \rho \dot{\theta} \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix}, \quad \overrightarrow{n} = \begin{pmatrix} -\cos \theta \\ -\sin \theta \end{pmatrix}
\end{align*}$$
Therefore
\[ \vec{a} \cdot \vec{n} = \rho \dot{\theta}^2 = \frac{|\vec{v}|^2}{\rho} \]
\[ \kappa = \frac{1}{\rho} = \frac{\vec{a} \cdot \vec{n}}{|\vec{v}|^2} = \frac{|\vec{v} \times \vec{a}|}{|\vec{v}|^3} = \frac{\dot{x} \dot{y} - \dot{y} \dot{x}}{(\dot{x}^2 + \dot{y}^2)^{3/2}} \]

Recall that the osculating circle at a point approximates a \((C^2\text{-smooth})\) curve up to second order accuracy. Therefore, one can apply the above formula for circular motions to the osculating circle at a point \(P\). Thus, the localization of the dynamics on such a curve at \(P\) is essentially the same as that of a corresponding motion on its osculating circle at \(P\), and hence
\[ \vec{a}_P \cdot \vec{n}_P = \frac{|\vec{v}_P|^2}{\rho}, \quad \kappa = \frac{\dot{x} \dot{y} - \dot{y} \dot{x}}{(\dot{x}^2 + \dot{y}^2)^{3/2}} \]
holds in general. This is the physical meaning of \(\vec{a}_P \cdot \vec{n}_P\), the normal component of the acceleration.

Next, let us use the simple dynamical representation of a given ellipse \(\Gamma\), namely
\[ x = a \cos t, \quad y = b \sin t, \]
to compute the curvature of \(\Gamma\) at \(P\), as follows:
\[ \vec{v} = \begin{pmatrix} -a \sin t \\ b \cos t \end{pmatrix}, \quad \vec{a} = \begin{pmatrix} -a \cos t \\ -b \sin t \end{pmatrix} = \vec{P}\vec{O} \]
As can be seen from Figure 3, the area of the parallelogram spanned by \(\vec{v}\) and \(\vec{a}\) is
\[ \text{Area}((\vec{v}, \vec{a})) = |\vec{v}||\vec{a} \cdot \vec{n}| = ab \]
and combined with (8) one has
\[ \frac{1}{\rho} |\vec{v}|^2 = \vec{a} \cdot \vec{n} \implies \frac{1}{\rho} = \frac{ab}{|\vec{v}|^2} \]
(9)
On the other hand, by Corollary 3.4 and Figure 3 one also has \(\vec{P}\vec{K} = a\), so
\[ \vec{a} \cdot \vec{n} = a \sin \varepsilon. \]
Therefore, by (9)
\[ ab = |\vec{v}|a \sin \varepsilon, \quad \text{i.e. } |\vec{v}| = \frac{b}{\sin \varepsilon}, \]
and formula (7) follows immediately from this.

3.3 The polar coordinate equation of an ellipse

**Theorem 3.6** Set \(\overline{F_1P} = r\) and \(\theta = \angle F_2F_1P\). Then the equation of the ellipse \(\Gamma\) is given by
\[ \frac{1}{r} = \frac{1}{b^2} (a - c \cos \theta) \]
(10)
Proof. \( PF_2 = 2a - r \), and by the cosine law applied to triangle \( \Delta F_2F_1P \)
\[
(2a - r)^2 = r^2 + 4c^2 - 4cr \cos \theta,
\]

and consequently
\[
4b^2 = 4a^2 - 4c^2 = 4r(a - c \cos \theta),
\]
which can be restated as in [10]. ■

4 On the derivation of the inverse square law as a consequence of Kepler’s area law and ellipse law

First of all, it follows readily from the area law, namely the quantity in (3) is a constant, that the acceleration vector \( \vec{a} \) is pointing towards \( F_1 \), see Theorem 2.1. Thus, what remains to prove is that the magnitude of \( \vec{a} \) should be inverse proportionate to the square of \( PF_1 \) as a consequence of the ellipse law. This is the monumental achievement of Newton which also led him to the great discovery of the universal gravitation law, see Section 7. However, his proof (cf. Proposition 11 of [9]) is rather difficult to understand.

The following are three much simpler proofs, each uses different aspects of the focal geometry of ellipses discussed in Section 3:

Proof. I (in the spirit of Greek geometry)

By the area law and the identity (6)
\[
d_1 \frac{1}{v} = \frac{2\pi ab}{T}, \quad d_1d_2 = b^2
\]

Therefore,
\[
|\vec{v}| = \frac{2\pi ab}{T d_1} = \frac{2\pi ab d_2}{T b^2} = \frac{\pi a}{b T} \frac{F_2F_2}{F_2F_2}
\]
\[
F_2F_2 = F_2F_1 + F_1F_2 = \left( -2c, 0 \right) + 2a \left( \cos \theta, \sin \theta \right)
\]

Hence (see Figure 3),
\[
\vec{v} = \frac{\pi a}{b T} \left( 0, -2c \right) + \frac{2\pi a^2}{b T} \left( -\sin \theta, \cos \theta \right)
\]

and by Remark 2.3,
\[
\vec{a} = \frac{d}{dt} \vec{v} = \frac{2\pi a^2 \theta}{b T} \left( -\cos \theta, -\sin \theta \right) = -\frac{\pi^2}{2} \frac{(2a)^3}{T^2} \frac{1}{r^2} \left( \cos \theta, \sin \theta \right)
\]

■
Proof. II (using the kinematic formula for curvature— the proof Hooke sorely missed?)

By the area law (Remark 2.3, Theorem 3.5, and (8)),

\[ r \left| \vec{v} \right| \sin \varepsilon = \frac{2\pi ab}{T}, \]
\[ |\vec{a}| \sin \varepsilon = \vec{a} \cdot \vec{n} = \frac{1}{\rho} \left| \vec{v} \right|^2, \]
\[ \frac{1}{\rho \sin^3 \varepsilon} = \frac{a}{b^2}, \]

and consequently

\[ |\vec{a}| = \frac{1}{\rho \sin \varepsilon} \left| \vec{v} \right|^2 = \frac{4\pi^2 a^3}{T^2 r^2}. \tag{11} \]

Remark 4.1 On page 110 of [3], S. Chandrasekhar states: “That Newton must have known this relation (cf. (7)) requires no argument!” In fact, a thorough analysis of the proof of Proposition 11 in [9] will reveal that its major portion is devoted to the proof of (7), and the inverse square law can then be deduced essentially in the same way as the short simple step of Proof II. But such a crucial role of curvature in his proof is, somehow, hidden in his presentation.

Proof. III (using analytic geometry)

For a planary motion with position vector

\[ \vec{OP} = r \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}, \tag{12} \]

the acceleration vector is

\[ \vec{a} = \frac{d^2}{dt^2} \vec{OP} = (\ddot{r} - r \dot{\theta}^2) \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} + (2r \dot{\theta} + r \ddot{\theta}) \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix}. \tag{13} \]

Now, assuming Kepler’s area law the force must be radial, and therefore the second component in (13) vanishes. Moreover, assuming the trajectory is an ellipse (or conic section, but not a circle), we need only show that \((\ddot{r} - r \dot{\theta}^2)\) is inverse proportional to \(r^2\). This will be achieved by differentiation of the polar coordinate equation (10).

At this point, however, it is instructive to derive the Binet equation and apply it to our situation, since the remaining calculations will be similar in both cases. Thus, setting \(q = 1/r\) as a new variable depending on \(\theta\), straightforward differentiation of \(q\) and elimination of \(\dot{\theta}\) by introducing the constant \(k = r^2 \dot{\theta}\) yield the identity

\[ \ddot{r} - r \ddot{\theta}^2 = -k^2 q^2 \left( \frac{d^2 q}{dq^2} + q \right), \tag{14} \]

which is just the Binet equation (11) divided by \(m\). Next, by differentiation of the ellipse equation (10)

\[ q = \frac{1}{b^2} (a - c \cos \theta), \]

14
we deduce the identity
\[ \frac{d^2 q}{dq^2} + q = \frac{a}{b^2} \]
which by substitution into (14) yields
\[ r - \frac{dr}{d\theta}^2 = \frac{k^2 a}{b^2} \frac{1}{r^3} = \frac{4\pi^2 a^3}{T^2} \frac{1}{r^2} \]
(15)

Remark 4.2 The proof given by Newton in [9] is historically the first proof of the inverse square law, a great historical event among the major advances of the civilization of rational mind. Therefore, a careful reading as well as a thorough understanding of the underlying pertinent ideas (or insights) of such a proof are, of course, highly desirable and of great significance towards one’s understanding of the history of science. For example, a reading of corresponding sections of [3] might be helpful for such an undertaking.

Remark 4.3 If one compares Newton’s proof and the above triple of proofs, one finds that the area law and the focal geometry of ellipses always play the major roles in each proof, while the differentiation of sine and cosine is the only needed analytical computation involved in each proof. In fact, the main differences between them lie in the ways of proper synthesis between the area law and the focal geometry of ellipses.

5 A crucial integration formula for the gravitation attraction of a spherically symmetric body

Newton’s proof (cf. Proposition 71 in Book I of [9]) for the following important integration formula, nowadays often referred to as the “superb theorem”, is again not easy to understand. Therefore, for the convenience of the reader, we include here an elementary simple proof (cf. [7]). In the very recent physics literature, see Schmid [13] for a similar but still different geometric proof.

Theorem 5.1 The total gravitation force acting on an outside particle by a body with spherically symmetric mass distribution is equal to that of a point mass of its total mass situated at its center.

Proof. First of all, the proof can be directly reduced to the case of a thin spherical shell with a uniform mass density \( \rho \) per area. The key idea of this elementary geometric proof is to use the subdivision of the spherical surface of radius \( R \) induced by the subdivision of the total solid angle (i.e. the unit sphere) centered at \( P' \) on \( OP \) with \( OP' \cdot OP = R^2 \), as illustrated in Figure 4.
Figure 4: Newton’s “superb theorem”
For any given point $Q$ on the $R$-sphere, $\Delta OPQ$ and $\Delta OQP'$ have the same angle at $O$, and moreover, their corresponding pairs of sides are in proportion, namely $P'P' \overline{PP'}$
\[
\frac{OP}{OQ} = \frac{OP}{OP'} = \frac{OQ}{OP'}.
\]
Therefore, $\Delta OPQ \sim \Delta OQP'$ and hence
\[
\angle OQP' = \angle OPQ \ (= \theta) \quad \text{and} \quad \frac{P'Q}{PQ} = \frac{R}{OP'}.
\]

Now, let $dA$ be the element of area on the $R$-sphere around $Q$ and $d\sigma$ be its corresponding element of area on the unit sphere. Then, as indicated in the magnified solid angle cone of $d\sigma$ in Figure 4, the corresponding area element on the $P'Q$-sphere centered at $P'$ is equal to $P'Q^2d\sigma$ on the one hand, but equals $dA\cos\theta$ on the other hand, because the dihedral angle between the tangent planes of $dA$ (resp. $P'Q^2d\sigma$) at $Q$ is equal to $\theta$. Consequently,
\[
dA\cos\theta = P'Q^2d\sigma.
\]

Note that the contribution of $d\overrightarrow{F}$ to the total composite force is equal to $|d\overrightarrow{F}|\cos\theta$, namely (with mass $m_1$ at $P$)
\[
|d\overrightarrow{F}|\cos\theta = G\frac{m_1\rho dA}{|PQ|^2}\cos\theta = Gm_1\rho \frac{P'Q^2}{PQ^2}d\sigma = Gm_1\rho \frac{R^2}{OP'^2}d\sigma.
\]
Therefore, the total gravitation force is given by
\[
\int |d\overrightarrow{F}|\cos\theta = Gm_1\rho \frac{R^2}{OP'^2}\int d\sigma = Gm_1\frac{4\pi R^2}{OP'^2}d\sigma = Gm_1\frac{m_2}{OP'^2}, \quad m_2 = 4\pi R^2\rho.
\]  

**Remark 5.2** (i) Newton was undoubtedly aware of the importance of the integration formula for the gravitational force of a spherically symmetric body, both for celestial gravity and for terrestrial gravity, and moreover, for the unification of both, thus enabling him to proclaim the law of universal gravitation.

(ii) In fact, he must have been working hard on it, ever since his success in proving the inverse square law some time in 1680–81. His letter of June 20, 1686, to Halley recorded his repeated failure up to around 1685, while his final success, in 1686, of proving such a wonderful simple formula is actually the crown-jewel of his glorious triumph — the law of universal gravitation (cf. [3], [9]).
6 A simple proof of the Kepler problem

For the convenience of the reader, we include here a simple proof on the solution of the Kepler problem, as follows:

**Theorem 6.1** Suppose that the acceleration of a motion is centripetal and inverse proportional to the square of $\overrightarrow{OP} = r$, namely for some constant $C > 0$,

$$ \overrightarrow{a} = \frac{C}{r^2} \left( -\cos \theta - \sin \theta \right). \quad (17) $$

Then the motion satisfies the area law and its orbit is a conic section with the center as one of its foci.

**Proof.** I (vector calculus involving only $\sin x$ and $\cos x$)

The centripetality property amounts to the area law (see Section 2.3), so there exists a constant $k$ such that

$$ 2 \frac{dA}{dt} = r^2 \dot{\theta} = k \quad (18) $$

Therefore, it follows directly from (17) and (18) that

$$ \frac{d}{d\theta} \overrightarrow{v}(\theta) = \frac{C}{k} \left( \frac{-\sin \theta}{\cos \theta} \right) $$

Hence, there exists a constant vector $\overrightarrow{\delta}$ such that

$$ \overrightarrow{v}(\theta) = \frac{C}{k} \left( \frac{-\sin \theta}{\cos \theta} \right) + \overrightarrow{\delta}. $$

Without loss of generality, we may assume that $\overrightarrow{v}(0)$ is pointing vertically, thus having $\overrightarrow{\delta} = \left( \delta, 0 \right)$.

Now, again using the area law, one has from (1) and (18)

$$ r \left| \begin{array}{cc} \cos \theta & \frac{C}{k} \left( -\sin \theta \right) \\ \sin \theta & \frac{C}{k} \cos \theta + \delta \end{array} \right| = k, $$

namely

$$ \frac{1}{r} = \frac{C}{k^2} \left( 1 + e \cos \theta \right), \quad e = \frac{k \delta}{C}, \quad (19) $$

which is exactly the polar coordinate equation (10) of a conic section with $\pm e$ as its eccentricity and $O$ as a focal point.

**Proof.** II (integration, starting from the Binet equation)

By assumption, the radial force is of type $F = -\alpha q^2$, $\alpha > 0$ constant, and $q = 1/r$. Therefore, by the Binet equation, for

$$ \frac{d^2q}{d\theta^2} + q = h > 0 \quad (constant) $$

$$ 18 $$
Since the general solution of this 2nd order ODE can be written as $q = A \cos(\theta + \theta_0) + h$, with the appropriate choice of the axis $\theta = 0$ the solution takes the form (19).

Remark 6.2 (i) For the Kepler problem, the constant $C$ in (17) is positive (i.e. in the case of attraction force). However, the proof also works as well for $C < 0$ (i.e. repulsive force), which is important in quantum mechanics for studying scattering.

(ii) Mathematically, the solution of the Kepler problem amounts to solve the second order ODE (17) in terms of the given initial data, especially the uniqueness without appealing to sophisticated theorems. Our first proof of Theorem 6.1 accomplishes such a task in two simple, elementary steps, namely, firstly obtaining the solution of $\vec{v}(\theta)$ in terms of the initial velocity by a direct application of the area law (i.e. consequence of the centripetality), and then obtaining the polar coordinate equation of the trajectory by another direct application of the area law.

Note that the area law is actually the dynamical manifestation of the rotational symmetry of the plane with respect to the center of the centripetality. Therefore, it is, of course, natural to use the polar coordinate system and compute $\frac{d}{d\theta} \vec{v}(\theta)$ in the proof. In retrospect, it is not only the simplest proof with perfect generality and the least of technicality, but it is also the most natural way of solving the Kepler problem.

(iii) We refer to [1], [3], [9], [16] for comparison of proofs of Theorem 6.1 as well as for the discussions of whether Newton actually proved it.

7 Concluding remarks

(i) In Book III of Principia, Newton presents his crowning achievements, namely a demonstration of the structure of the “system of the world”, derived from the basic principles that he had developed in Book I and Book II. Newton’s three laws of motion and the law of universal gravitation are for the first time seen to provide a unified quantitative explanation for a wide range of physical phenomena. In particular, they provide the foundation of celestial mechanics, and the first complete mathematical formulation of the classical n-body problem appears in Newton’s Principia. The law of universal gravitation is, in fact, the first and also one of the most important scientific discoveries in the entire history of sciences. An in-depth understanding of how it arises naturally from the mathematical analysis as well as synthesis of those empirical laws of Kepler and Galilei is not only instructional but also inspirational.

(ii) The law of universal gravitation reflects the physical principle expounded by Newton that all bodies interact gravitationally. But such a statement presupposes a deeper understanding of the force law $F = ma$, namely that two interacting bodies attract each other by equal forces and in opposite directions. This follows from Newton’s third law, which is his own insight, stating that
for every action there is an opposite reaction. In the case of gravitation this interaction is expressed by the basic and well known formula

$$F = G \frac{mM}{r^2}$$

(20)

for the mutual gravitation force between two point masses $m$ and $M$ separated by the distance $r$, where $G$ is the gravitational constant. By Newton’s “superb theorem” (see Section 5), the same formula holds for two bodies with spherically symmetric mass distributions and total masses $m$ and $M$, and $r$ is the distance between their centers. For many bodies, such as the planets circling the sun, the bodies attract one another and therefore they also perturb one another’s orbits. Still, as pointed out by Newton, the law of universal gravitation explains why the planets follow Kepler’s laws approximately and why they depart (as is also observed) from the laws in the way they do. Let us briefly recall the underlying reasoning.

First, consider a sun-planet system with masses $M$ and $m$, ignoring the other celestial bodies. By combining the force law $\vec{F} = m \vec{a}$ and formula (20), it may seem that one is led to equation (17), with $C = GM$, and thus the planet’s orbit will be a solution of the Kepler problem as stated in Theorem 6.1, namely an ellipse with the sun at one of the focal points. However, this reduction of the sun-planet problem to a one-body (or Kepler) problem centered at the sun is only approximately correct. As we would phrase it today, the validity of the Newtonian dynamics hinges upon using an inertial frame of reference, namely with the origin “at rest”.

How did Newton himself imagine the origin of an inertial frame could be chosen? The sun is much larger than the planets, but Newton was aware of the tiny motion of the sun due to the attraction of the planets. He estimated the Center of the World, namely the center of gravity of the whole solar system, to be very close to the sun, say within one solar diameter.

For a general two-body system, the common center of gravity is “at rest” if the interaction with other bodies is neglected. Thus, the position of one body determines the position of the other, and Newton argues correctly that the two-body problem again reduces to a one-body problem with radial attraction towards the center of gravity. So, both bodies follow Keplerian orbits with the latter point as a common focus.

However, whereas an exact solution of the two-body problem is one of the great triumphs of classical mechanics, the non-integrability of the $n$-body problem for $n \geq 3$, which is well known nowadays, was maybe suspected already by Newton when he wrote in his tract *De Motu* (1684): “— to define these motions by exact laws allowing of convenient calculation exceeds, unless I am mistaken, the force of the entire human intellect”.

(iii) The measurement of the gravitational constant $G$ in formula (20) has a long history; in fact, the formulation of gravity in terms of $G$ did not become standard until the late 19th century. The first successful experiment in the laboratory, by H. Cavendish (1731–1810) in 1798, aimed at measuring the mass
$M_e$ of the earth, or equivalently, the (average) density $\rho_e$ of the earth from the knowledge of the earth’s radius $R$. However, knowing the acceleration of gravity $g$ at the surface of the earth, measuring $\rho_e$ amounts to measuring $G$ due to the relations

$$G = \frac{gR^2}{M_e} = \frac{3g}{4\pi R\rho_e}.$$ 

The apparatus used by Cavendish was actually designed by the geologist J. Michell (1724–1793), who was a pioneer in seismology and did also important work in astronomy. Although Laplace (1796) is usually credited for being the first who described the concept of a black hole (condensed star), Michell argued in a 1784 paper how such an objects could be observed from its gravitational effect on nearby objects. However, Michell is best known for his invention, probably in the early 1780’s, of the torsion balance, which is the major device of the apparatus he built to measure the quantity $\rho_e$.

But Michell did not complete this project, and his equipment was taken over by Cavendish, who rebuilt the apparatus with some improvements, which enabled him to carry out measurements of the density of the earth with very high accuracy. We refer to his report [2], see also [14], [4]. The measurement of the universal constant $G$ had many remarkable consequences, for example, estimates of the mass of the earth, moon, sun, other planets, and massive black holes.

(iv) In 1785 Coulomb published his investigation of the electric force, using an apparatus involving a torsion balance. But, according to Cavendish, Michell had described his torsion balance device to him before 1785, so it seems that both Michell and Coulomb must be credited with the invention of the ingenious torsion balance. Like the gravitation force, the Coulomb force between charged particles is also of the inverse square type. In fact, here Newton’s “superb theorem” not only applies, but also plays a useful role in providing the empirical evidence as well as the measurement of the constant of proportionality, namely the electric force constant (or Coulomb’s constant) $k_e$.

References

[1] V.I. Arnol’d, *Huygens and Barrow, Newton and Hooke*, Birkhäuser Verlag, 1990.

[2] H. Cavendish, *Experiments to determine the Density of the Earth*, Philosophical Transactions of the Royal Society of London 88 (1798), pp. 469-526.

[3] S. Chandrasekhar, *Newton’s Principia for the Common Reader*, Oxford University Press, 1995.
[4] B.E. Clotfelter, The Cavendish experiment as Cavendish knew it, Amer. J. Phys. 55 (1987), 210.

[5] I.B. Cohen, Newton’s Discovery of Gravity, Scientific American, Vol. 244, March 1981 pp. 166-179.

[6] J. Horrocks, Opera posthuma (edited by J. Wallis), London 1673.

[7] W.Y. Hsiang, The Geometry of Spheres, Normat 1, 1994.

[8] A. Koyré, An Unpublished Letter of Robert Hooke to Isaac Newton, Isis, Vol. 43, No. 4 (Dec.,1952), pp. 312-337.

[9] I. Newton, Philosophiae Naturalis Principia Mathematica, London 1687. English translation by Andrew Motte, The Mathematical Principles of Natural Philosophy, 1729.

[10] P.J. Pugliese, Robert Hooke and the dynamics of motion in a curved path, in M. Hunter and S. Schaffer, ”Robert Hooke, New Studies” (Boydell, Woodbridge, 1989) pp. 181-205.

[11] R.D. Purrington, The First Professional Scientist. Robert Hooke and the Royal Society of London, Birkhäuser Verlag, 2009.

[12] B. Ram, On the deductive and inductive nature of Newton’s gravitation law, Amer. J. Phys. 45 (1977), 1223.

[13] C. Schmid, Newton’s superb theorem: An elementary geometric proof, Amer. J. Phys. 79 (2011), 536.

[14] M.H. Shamos, Great Experiments in Physics, Henry Holt &Co., New York, 1959, p.75, and Dover Publications, 1987.

[15] J. Sivardière, On the inverse Kepler problem, Eur. J. Phys. 6 (1985), 245-248.

[16] D. Speiser, The Kepler Problem from Newton to Johann Bernoulli, Archive for History of Exact Sciences, 50 (1996), pp. 103-116.
