2. THE QUEST FOR LUNAR THEORY

Why was lunar theory so actively pursued? There were theoretical as well as practical reasons why astronomers, and especially those working in the middle of the eighteenth century, were highly interested in accurate methods to compute lunar positions for specified times. The theoretical ones had to do with the general progress of science in the 18th century (and later); on the practical side, lunar theory contributed to navigation at sea.

2.1 SCIENTIFIC SIGNIFICANCE

In his *Philosophiae Naturalis Principia Mathematica* (1687), Isaac Newton had laid down the general laws of motion of bodies, whether they are on earth or in the heavens. Newton introduced a new unified way of understanding motion, in particular that of celestial bodies.[1] He showed that the basic properties of the planetary elliptic orbits, which Kepler had found, were a consequence of the force of gravitation, which varies with the inverse square of the distance between attracting bodies. He recognized that the same gravitation was responsible for many other observed phenomena, such as the motion of comets, the precession of the earth's rotational axis, and the tides of the oceans. He also realized that the moon's orbit was affected not only by the gravitation of the earth around which it rotates, but also by the gravitation of the sun.

Although Newton could calculate several properties of the motion of the moon successfully from his laws of motion and the principle of gravitation, the moon seemed to deviate from its predicted path in some other aspects. For example, severe difficulties obstructed the theoretical derivation of the correct average advance of the line of apsides of the lunar orbit, i.e., the imaginary line that connects its apogee and perigee. To his dismay, Newton could get only half the observed value (about 40° per year) out of his calculations.

This discrepancy was serious. Newton's principle of gravitation was not readily accepted by everyone, partly because it implied the odd phenomenon of 'action at a distance' (masses exerting their gravitational forces instantly in every part of the universe, apparently without a material contact). Therefore, the principle of gravitation was put to the test: if it were really universally applicable, then it should explain all the motions of the heavenly bodies quantitatively. The problem of the apsidal line movement cast doubts on the theory.

1 Currently, the most accessible edition is [Newton et al., 1999].

In the 1740's this problem was attacked by some of the ablest mathematicians of the time, including Leonhard Euler, Alexis Clairaut, and Jean le Rond d'Alembert. Their mathematical techniques were more powerful than Newton's, but nonetheless they, too, could account for only half the observed advance of the apsidal line. They proposed several causes for the discrepancy, including a law of gravitational attraction slightly differing from the inverse-square relation proposed by Newton. Eventually, though, Clairaut discovered in 1749 that some higher-order terms in the calculations, which he had supposed to be negligible, played havoc. With these terms taken into account, Clairaut computed a mean apsidal motion sufficiently close to the observed value. Clairaut's success contributed decisively to the general acceptance of the inverse-square law of gravitation, while his initial failure held a warning against a too easy neglect of supposedly very small terms.[2]

Outside of lunar theory, a similar difficulty was found in the movement of the planets Jupiter and Saturn, the two most massive planets in the solar system. From comparison of Tycho's observations and observations handed down by Ptolemy, Kepler had noted that the motions of these planets showed irregularities, the nature of which could not confidently be ascertained. Newton could calculate that these massive planets had a significant influence on each other's orbits, but it was beyond him to explain the phenomenon quantitatively. In 1748, this problem was made the subject of the annual contest of the *Académie des Sciences* of Paris. Euler's contribution[3] won the competition, yet he had failed to bring a satisfactory solution, and the Paris academy staged the same subject for the contests of 1750 and 1752.

The importance of Euler's memoir lay in the new analytical techniques that he applied in it, including trigonometric series and an attempt to fit parameters to observations—we will encounter his memoir again on several occasions. Only in 1785 did Laplace show that the observed phenomena were a consequence of gravitation. He found that the two planets are subject to perturbations of nearly $50'$ and $20'$, and a period of about 880 years, as a consequence of the nearly $5:2$ ratio of their orbital periods. To discover this long-periodic effect, Laplace had developed tools to select potentially sizable terms out of the infinite number of terms that arise in the trigonometric series.[4]

Another related topic which is outside our scope but which nonetheless deserves to be mentioned here, is the moon's secular acceleration. In 1693, Edmond Halley (1656–1742) compared lunar eclipses of recent, medieval Arabic, and classical Babylonian time, and discovered that the moon's mean motion had been gradually increasing. Dunthorne in 1749 fixed the acceleration at $10''$ per squared century, Mayer first put it at $7''$ but later revised his value to $9''$.[5] At that time, The phenomenon defied explanation through the theory of gravitation, and many as-

2 An extensive investigation of this episode is in [Waff, 1975]; Clairaut published his results in [Clairaut, 1752a].

3 [Euler, 1749a].

4 Wilson discusses the history of this problem (and much more) in [Wilson, 1985].

5 Mayer reported on his research in [Mayer, 1753b, pp. 388-390]; his calculations are scattered throughout Cod. $\mu_{41}^{\sharp}$.

tronomers considered it an effect of the aether, a thin hypothetical substance that was supposed to fill the universe.

It was Laplace again who showed in 1787 that the secular acceleration was actually a perturbation of very long period, caused by a periodical change in the eccentricity of the earth's orbit through the actions of the planets Jupiter and Venus. His success to account theoretically for the full observed effect lasted until John Couch Adams in 1854 discovered errors in Laplace's calculation, the correction of which caused half of the computed acceleration to disappear—and half of the observed amount to be unexplained. It is now believed that the missing half is brought about by tidal friction, which causes the earth to spin down, and consequently angular momentum to be transferred from the earth to the moon.[6]

Adams's discovery of Laplace's error holds a warning to lunar theorists that is parallel to Clairaut's discovery of the true precession of the line of apsides. The problems posed by the motion of the moon, and by the mutual influences of Jupiter and Saturn on each other, have opened the research on the 'three-body problem', which has defied mathematicians in search of a closed-form solution except in a few special cases. Indeed, Bruns and Poincaré have demonstrated that the general problem of three bodies does not allow the formulation of a sufficient number of relations between the coordinates and velocities of the bodies, in order for it to be solvable in closed form.[7]

2.2 APPLICATION TO NAVIGATION

From a more pragmatic point of view, a good lunar theory was urgently needed in oceanic navigation and geography for the determination of geographical longitude. Whereas navigators as well as geographers had sufficient means to find their geographical latitude, sufficiently accurate methods of finding longitude were lacking. In fact, navigators had absolutely *no* way of knowing their longitude when out of sight of well-surveyed coasts (of which there were but few) except by keeping an account of courses and distances traversed; with the instruments and procedures of the time, and the scant knowledge of ocean currents, the quality of this so-called *dead reckoning* was not so good that one would like to entrust a whole cargo's worth to it.

My own experience in small boat navigation tells me that an allowance for an error margin of 10% of recorded distance is not overly precautious—in 20th century circumstances, that is. A seventeenth-century navigator would perhaps do well to allow for twice that error unless he was sure that there were no appreciable inaccuracies in his sea charts. Thus, to a party proceeding from Batavia to the Cape of

6 Actually, the picture is a bit more complicated than that. The slowing of the earth's rotation increases the angular momentum of the moon via an increased radius of its orbit; Kepler's third law then implies a *decrease* of the moon's angular velocity, not an increase. But the moon's motion currently *appears* to accelerate relative to the increasing length of the day [Verbunt, 2002].

7 Cf. [Brown, 1896, p. 27].

Good Hope, which is a distance of some 5000 nautical miles covered in perhaps two months, the suggested margin might have amounted to a week sailing.

This systematic inability of the navigators of the 16th to 18th centuries to fix one of their coordinates on the earth became a concern of their kings and patrons. Substantial rewards were put in prospect, successively in Spain, the Netherlands, France, and England for solutions to the problem, neatly following the torch of sea power. Finding the longitude stood on a par with squaring the circle, according to many people. However, the longitude problem was overstressed: the attention that it received seems to be in imbalance compared to the lack of attention for some other potential causes of shipwreck.[8]

Probably the most well-known of these rewards was stipulated by British Parliament as the *Act 12 Queen Anne, Cap. XV*, colloquially known as the *Longitude Act* of 1714. Their Act had certainly been inspired by the wrecking of Sir Clowdisley Shovell's fleet on the Scilly Islands and the consequent loss of the lives of 2000 able-bodied seamen. This was one of the most dramatic disasters in British sea history; it was, however, caused by a navigational error in *latitude*, rather than *longitude*.[9]

The Act promised a reward of £20,000 for the discoverer of a method of finding longitude that was accurate to $\frac{1}{2}^\circ$ of a great circle, i.e., to 30 nautical miles; lower rewards of £15,000 and £10,000 were to be granted for an accuracy of $\frac{2}{3}^\circ$ and 1°, respectively. The method had to be practicable and useful, both aspects to be demonstrated on a voyage to the West-Indies and back. A commission, which became known as the *Board of Longitude*, and consisting of representatives of the Admiralty, Parliament, and scientists (including the Astronomer Royal and the Savilian and Lucasian professors of mathematics) was to oversee the execution of the Act. Many proposals were presented to members of the Board by various individuals, but the plans were almost always considered as impracticable, not useful, or even nonsensical, long before it was deemed necessary to bother the complete Board.

At least as influential as this British Act was the testament of the French politician Rouillé de Meslay of 1715, stipulating that the *Académie des Sciences* in Paris should organize an annual prize contest. The subject of the contest would alternately address the motions of celestial objects, and navigation (in particular longitude finding). These contests formed a significant impulse for the work of Euler, Clairaut, and others who contributed their essays in response to the prize questions. Between 1720 and 1792, the lunar orbit and other, related, problems in celestial mechanics repeatedly provided subjects for the challenge.

In the course of about two and a half centuries (from the beginning of the 16th to the middle of the 18th), many sensible and insensible solutions to the longitude problem had been proposed. The four most seriously pursued candidate solutions

8 [Davids, 1985, p. 86]; and: '[. . .] in an age when the world was not yet properly charted the fact that one did not know the longitude of one's destination meant that the longitude of one's own position was less important than it would be today. [. . .] The dire results of a lack of means of observing the longitude have been overstressed by many writers.' [May, 1973, p. 28–29].

9 [May, 1973, pp. 27–8].

employed either *timekeepers*, *lunar distances*, the *satellites of Jupiter*, or the *variation of the compass*.

The first three of these aimed at providing the local time on a (possibly distant) reference meridian: the timekeeper method by setting a reliable watch to that time, the lunar distance method by the relatively swift motion of the moon (to be further explained below), and Jupiter's satellites by their predictable crossings of Jupiter's visible disc or its shadow cone. The fourth method depended on the usefulness of a relation between magnetic variation and geographic longitude, but the relation is rather weak, even non-existent in some regions of the globe, and moreover slowly changing with time. Cartographers could—contrary to navigators—afford to wait for particularly favourable celestial events such as lunar eclipses or occultations of stars by the moon, to determine longitude. I will now briefly explain the principle of lunar distances, in order to clarify the relation between longitude determination and Mayer's work on lunar motion.[10]

The method of lunar distances exploits the relatively swift movement of the moon, which completes its orbit in about 29.5 days with respect to the sun, or in 27.3 days with respect to the stars. In other words, the moon moves roughly its own diameter ($30'$) per hour relative to the stars. Therefore, if an observer can make an accurate measurement of the place of the moon relative to another celestial body (whose position is accurately known), and if the observer has accurate knowledge of the movement of the moon with time, than he can combine the two and derive the time of his observation. Metaphorically, it is as if the moon is the hand of a celestial clock, with the stars and sun forming the dial. The somewhat puzzling expression of 'taking a lunar *distance*' should therefore be understood in the sense of measuring the angular separation of the moon from another celestial body.

Quite a number of difficulties make the lunar distance method awkward to apply. Many of these are directly or indirectly connected with the inherent inaccuracy of the method: the moon revolves approximately 30 times as slowly around the earth as the earth about its own axis. Therefore, errors in the obtained lunar distance (i.e., after the measurement and after all necessary computations have been made) get magnified by the same factor of 30 in the computed longitude. Thus, if the lunar distance is off by $2'$, the longitude will have an error of 1°, the lowest accuracy limit set by the Longitude Act. The contemporary enthusiastic reactions to Mayer's lunar tables were due to the fact that they were the first to predict lunar positions to within a $2'$ error margin; in other words, they were the first tables that did not consume already more than the allowable error margin even before the observer had taken his measurement.

The principles of the lunar distance and timekeeper methods had been published by, respectively, Johann Werner in 1514 and by Gemma Frisius in 1530. More than

10 For further reading on the history of longitude and lunar distances, see [Andrewes, 1996], [Howse, 1980], [Cotter, 1968], and [Marguet, 1931]. The best manuals on the method, in true 19th-century style, are contained in [Graff, 1914], [Jordan, 1885], [Chauvenet, 1863], and [Bohnenberger, 1795]. A refreshing modern approach is in [Stark, 1995], and an increasing number of modern-day enthusiast's can be reached over the internet.

two centuries later *both* methods were brought to fruition by, respectively, the lunar tables of Tobias Mayer (presented to the Board of Longitude in 1755), and the timekeepers of John Harrison (of which the fourth and last was presented to the same Board in 1760).

I have often pondered the near-coincidental perception and the near-coincidental realization of these methods, separated by such a long time-span of about 240 years. The difficulty with the moon was its seemingly irregular motion, which had to be modelled mathematically; the difficulty with timekeepers was also their irregularity, which in this case had to be harnessed by mechanical innovations. Werner and Frisius had conceived their ideas while Europe was expanding its horizon over the edges of the oceans, yet it took a dozen generations to expand the necessary technological abilities.

As a side remark, a particular role on the longitude stage was reserved for Edmond Halley. Halley was most instrumental in encouraging Newton to write *Principia Mathematica*, and subsequently in publishing that book, because he recognized the invaluable assets—if not foundations—that it was to offer to astronomy. But Halley had also spent several years as a sea captain: he had been surveying the geomagnetic field in the Atlantic Ocean (with an eye on its supposed potential for longitude finding), the coasts and tides of the English Channel, and the treacherous sandbanks of the Thames Estuary. When in 1675 the Royal Observatory was founded in Greenwich with the explicit object of improving astronomy for the sake of navigation and in particular longitude finding, the nineteen-year old Halley was involved. Decades later, in 1720, he succeeded Flamsteed as the Astronomer Royal. Spending almost the final quarter of his long life in that position, he must have had an extraordinarily diversified view of both the theoretical and practical aspects of longitude finding.

2.3 ASTRONOMICAL PREREQUISITES

The affiliation between mathematics and astronomy is long and intense, but nowadays not every mathematician is versed in astronomical terminology. This section is intended as a reference to those who feel uneasy (or perhaps even lost) amidst the ecliptic, nodes, anomalies and inclinations. It will guide them through a little background in coordinates, elliptic orbits, and equAtions of the lunar orbit.

2.3.1 Celestial coordinates

A terrestrial observer looking at the night sky, perceives the stars and the planets as if they are attached to a giant sphere of which he is the centre: distances cannot be discerned without advanced techniques, only directions can. This metaphor of the celestial sphere is an extremely powerful one, and will guide us in this section.

Figure 2.1 shows the celestial sphere with the comparatively extremely small earth in its centre. The partially drawn line $P_N P_S$ is the earth's polar axis extended,

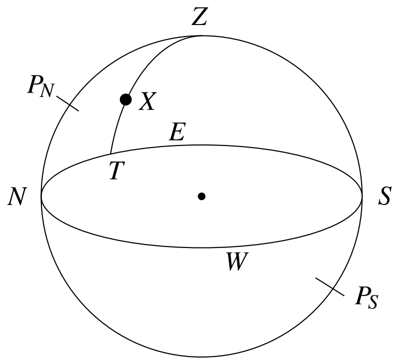

Figure 2.1: The celestial sphere, showing the celestial poles P_N and P_S as well as the observer's zenith Z and horizon $NESW$. The celestial object X has altitude TX and azimuth NT.

around which the sphere, with celestial body X attached to it, appears to rotate. Z is the zenith directly overhead of the observer, and $NESW$ is the horizon; these do not take part in the apparent rotation. The meridian circle ZP_NNP_SS intersects the horizon in N and S, which define the observer's directions north and south. The great circle arc ZX extended intersects the horizon perpendicularly in T; the altitude of X is the arc XT, and the azimuth of X is the arc NT, which must be labelled E or W to indicate if it is measured in easterly or westerly direction. The azimuth is not defined if X coincides with Z.

Thus, the immediately observable coordinates of a celestial body are its altitude and azimuth, which are related to the observer's horizon and meridian. The coordinates of each body change constantly (except when coincident with one of the poles) and it seems as if the celestial sphere rotates around a fixed axis. In reality, it is the earth that rotates.

Clearly, the celestial sphere must be provided with a coordinate system independent of the rotation of the earth and of the location of the terrestrial observer. As such, we will be concerned mostly with the geocentric ecliptic coordinate system. It is geocentric, meaning that the centre of the earth is taken as its origin. From this point of view, the earth stands still while the sphere rotates daily around it.

Taking the earth at rest in the centre of the celestial sphere, we distinguish between two different apparent motions of the sun. It moves with the apparent daily rotation of the celestial sphere, which is accountable for its rising and setting; besides, it appears to orbit around the earth once a year in the opposite direction. The plane in which the annual motion takes place, is called the plane of the ecliptic. The ecliptic is defined as the great circle formed by the intersection of this plane with the celestial sphere.

Similar to the ecliptic, we can think of the plane that contains the terrestrial equator, and intersect it with the celestial sphere to obtain the celestial equator, or just equator for short. It is a great circle just as the ecliptic. These two great circles

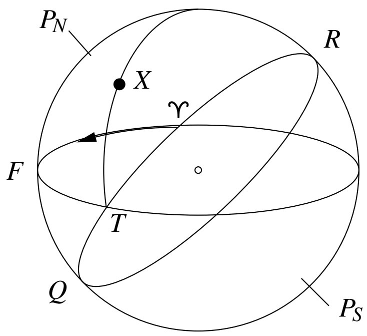

Figure 2.2: Ecliptic coordinates on the celestial sphere. $Q\Upsilon RQ$ is the celestial equator, the partially drawn line $P_N P_S$ is the earth's polar axis extended. $E\Upsilon FTE$ is the ecliptic, with the arc XT perpendicular to it. The longitude of object X is the arc ΥFT, its latitude is the arc TX.

are tilted with respect to one another, forming an angle of about 23.5°, called the obliquity of the ecliptic. Twice a year the sun crosses the celestial equator: on or about 21 March, it crosses to the northern hemisphere (which is located above its terrestrial namesake), and it returns to the southern hemisphere about 22 September. At these instances, the sun occupies the intersection points of the ecliptic and the equator; these are termed the vernal and autumnal equinox, respectively, referring to the fact that on those dates the lengths of day and night are equal everywhere on earth.

Now we come to define the ecliptic coordinates. In the ecliptic, the vernal equinox provides a point of reference, commonly marked by the symbol Υ. Consider (figure 2.2) a celestial object X on the sphere (not one of the ecliptic poles). Let T be the point on the ecliptic closest to X, hence $XT\Upsilon$ forms a right angle. The coordinates of X are then its ecliptic longitude ΥT, and its ecliptic latitude TX. The longitude is measured from Υ in the direction of the annual motion of the sun, from 0° to 360°. Traditionally, it was customary to divide the ecliptic in twelve so-called signs of 30° each. A longitude of, say, $9^s1°$ equals 271°, and $3^s22°30'$ equals 112.5°. Latitude is measured from 0° on the ecliptic to 90°, north or south, the hemispheres being labelled according to the terrestrial poles.

The physics of gravitation oblige us to make some refinements to this otherwise mathematical coordinate system. First and foremost, the rotating earth is not absolutely spherical. It is more like an oblate spheroid, with excess mass at its bulging equator. Solar and lunar gravitation exert a net moment on this bulge trying to pull it into the plane of the ecliptic. Since the earth behaves as a spinning top, this moment makes it precess, i.e., the rotational axis rotates itself slowly around a mean position perpendicular to the ecliptic. As a result, the vernal equinox is not a fixed point, but it precesses backwards along the ecliptic at the rate of approximately $50''$ per year relative to fixed space. Moreover, the moon's orbit is inclined to the ecliptic, and

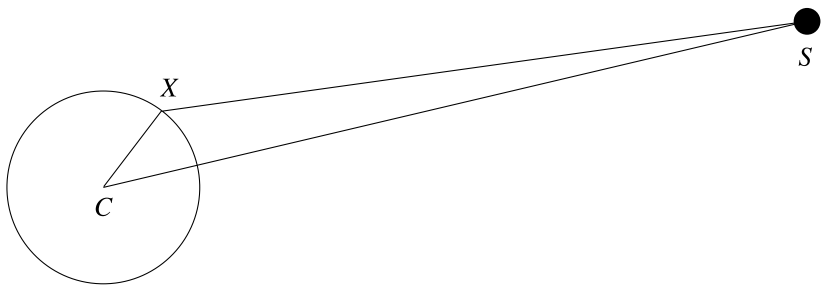

Figure 2.3: Parallax.

the lunar gravitational pull on the bulging terrestrial equator causes an additional wobble known as nutation, affecting the position of the vernal equinox periodically with period 18.6 year and amplitude $17''$. Precession was discovered by Hipparchus in the second century BC, nutation by James Bradley in 1747.

We have thus far disregarded distances. That is perfectly reasonable for our current purposes as long as only stars are concerned, because even from across two diametrically opposite positions in the earth's orbit (temporarily taking up the heliocentric view), shifts in stellar positions are at most $0.3''$.[11]. But for bodies in the solar system such as the sun and moon, we need to take their distances into account, in particular in relation to the size of the earth. This brings us to the subject of parallax.

Consider an observer X standing on the earth with centre C. The direction of the celestial object S as observed by X differs from the direction as it would have been observed from C, the origin of our coordinate system, unless S happens to be exactly overhead in the observer's zenith. The difference is the parallax, and it depends on the distance SC of the body, the radius CX of the earth, and the observed altitude of the body, $\angle CXS - 90^\circ$. It is largest (*ceteris paribus*) when CX is perpendicular to SX, i.e., when the body is in the horizon of the observer; in that case, it is termed the horizontal parallax. The oblateness of the earth makes it necessary to define the equatorial horizontal parallax, which is the horizontal parallax for an observer located on the equator, and to compute the horizontal parallax at the latitude of X assuming that the earth is an ellipsoid.

With the observer thus ruled out, and the radius of the earth considered constant, the horizontal parallax depends only on the distance of the object from the earth's centre, as expressed by the relation $\sin P = \frac{CX}{CS}$, where P is the horizontal parallax. Therefore P can be considered as a coordinate instead of radial distance.

There is also another way to look at horizontal parallax: observe that P denotes the apparent radius of the earth as seen from S. The lunar parallax, i.e., the hori-

11 Bradley had been searching in vain for this stellar parallax, and discovered aberration and nutation instead (aberration is an apparent deflection of light rays, resulting from the finite speeds of light and of the earth, which we will have no need to consider). Stellar parallax was first detected in 1838 by Bessel.

zontal parallax of the moon, is about $57'$, the solar parallax is a little bit less than $9''$. These parallaxes should *not* be compared to the (annual) stellar parallax of $0.3''$ mentioned earlier, without considering that the base line in the former is the earth's radius, in the latter however the radius of the earth's orbit.

2.3.2 Unperturbed orbits

Although we have hardly any interest in planets within the scope of this thesis, the ideas in this section will be explained with reference to planetary orbits. These ideas will be applied in the next section to the theory of the moon with only little changes. They were historically more readily applied in the planetary setting, and they belong to the generally desirable background knowledge about celestial phenomena.[12]

In first approximation, the motion of planets around the sun (and of the moon around the earth) can be treated as a two-body problem, disregarding all other masses. Numerous textbooks exist that teach how to formulate the two-body problem as a set of differential equations, and how these can be solved.[13] The solution is generally formulated in the form of the three Keplerian laws: that the planet revolves in an elliptic orbit, with the sun at one of the foci; that for each planet, the sun-planet radius describes equal areas in equal time intervals; and that for all planets, the ratio of the squares of their periods of revolution to the cubes of the semi-major axes of their orbits is constant.

Thus, let S in figure 2.4 be the sun and X a planet. According to Kepler's first law, the planet's orbit is an ellipse $AXX'PA$. Suppose that the planet moves from X to X' in a time interval τ, and that it takes a time T to complete its orbit and return to X. The second law dictates that the area of sector SXX' is to the area of the ellipse as τ is to T. According to the third law, $\frac{T^2}{(PC)^3} =$ constant. Actually, this constant depends on the masses of the sun and planet, but nearly all the mass of the solar system is concentrated in the sun; hence the ratio is nearly the same for all planets (but not for planetary satellites such as the moon), and Kepler's third law is true only approximately.

Nowadays, these three rules are considered the greatest contributions of Kepler to astronomy. Newton showed in *Principia Mathematica* that they had a common cause in the law of gravitation. The same Keplerian laws hold for the motion of the moon around the earth, at least in first approximation, but the moon will be further considered in the next section.

In the case of most planets, the approximation by way of an elliptic orbit is so good, and perturbations so small, that astronomers have adopted a comprehensive apparatus to describe the dimensions and orientation of the ellipse with respect to the plane of the ecliptic. The terminology of this apparatus, which we will next introduce, has for a large part been inherited and adapted from the Greek astronomers

12 Moreover, Kepler's third law would make no sense if the discussion were restricted to the lunar orbit.

13 For example [Moulton, 1902], [Brouwer and Clemence, 1961], [Roy, 1978].

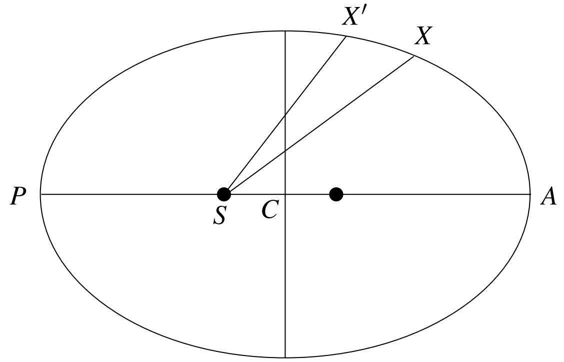

Figure 2.4: Elliptic orbit.

who lived with other models of planetary motion, where circles formed the basic building blocks instead of ellipses.

The location of one of the foci of the ellipse is the central body: for planets it is the sun. The dimensions of the ellipse (see figure 2.4) are characterized by its semi-major axis $PC = AC$ and eccentricity $SC : PC$. To know the orientation of the ellipse in space, one needs to know the plane of the ellipse as well as the orientation of its major axis within this plane.

The plane of the ellipse is specified by two parameters: its angle of inclination with respect to the plane of the ecliptic, and the direction of the line of intersection of the plane of the ecliptic and the orbit. This line is specified as the so-called longitude of the *ascending node*, which is the longitude of the point where the planet in its orbit crosses the ecliptic from south to north.

To orientate the ellipse in this plane, it is sufficient to specify the position of one of the extremities of the major axis AP (collectively called the apsides): either the longitude of perihelion P (the position in the orbit closest to the sun) or of aphelion A (where it is furthest away from the sun). In Mayer's time it was customary to specify the aphelion position for planets and the apogee position for the moon, but the perihelion position for comets, because comets could impossibly be observed near aphelion. The longitude of aphelion is measured from the vernal equinox along the ecliptic to the ascending node and then along the orbit to the aphelion. It consists of two consecutive arcs in two planes.

The parameters semi-major axis, eccentricity, inclination, ascending node, and longitude of aphelion are five of the so-called *orbital elements*. The sixth and last orbital element is the position of the planet in its orbit at a specified time.

The second law of Kepler, also known as the area law, implies that the motion of the planet is not uniform. Instead, it is swiftest at the perihelion and slowest at the aphelion. To arrive at the position of the planet at any specific instance in time, it is natural to relate it to one of the apsides first, and then to use the orbital elements to transform the position in the orbit to a position in ecliptic coordinates.

To ease the computation of the position of the planet in its orbit, one first assumes an imaginary planet that moves with uniform angular velocity and that coincides with the real planet in the aphelion (and, by symmetry, also in the perihelion). The position of this imaginary planet with respect to the aphelion is a linear function of time, and the difference between the imaginary and real planet follows from the area law and the properties of the ellipse.

In astronomical terminology, the imaginary planet represents the *mean motion* and the real planet the *true motion*. The angle ASX in figure 2.4 is termed the *true anomaly* if X is the real planet or the *mean anomaly* if X is the imaginary planet. The mean anomaly v can be computed as $v = 360° \frac{t}{T}$, where t is the time elapsed since aphelion passage and T the orbital period of the planet. The difference between the true anomaly θ and the mean anomaly is given by the *equAtion of centre*:

$$\theta - v = -\left(2 - \tfrac{1}{4}e^2\right)e\sin v + \tfrac{5}{4}e^2\sin 2v - \tfrac{13}{12}e^3\sin 3v - \tfrac{103}{96}e^4\sin 4v + \ldots, \qquad (2.1)$$

where e is the eccentricity of the orbit. Remember the typographical convention to invoke the astronomical context of the word 'equation', meaning a periodic correction of mean motion; in addition, note that one equAtion may consist of several terms of a trigonometric series. This particular example is a series approximation suitable for computations; besides, it is the only equAtion of longitude that applies to unperturbed motion.

2.3.3 The lunar orbit

The moon's motion, however, is notably less regular than the motions of the planets. The usual Keplerian laws of elliptical motion fail to account properly for the moon's apparent erratic behaviour, as a diligent naked-eye observer may detect. But Newton's theory of gravitation can equally well account for the perturbations of lunar motion as for the more regular planetary motions. The same force of gravity which keeps the planets in orbit around the sun, keeps the moon in orbit around the earth. In the course of its orbit, the distance between the moon and the sun varies, and consequently the gravitational force between sun and moon varies during a month by about 0.5% from its mean value, which is enough to cause considerable trouble.

The moon's motion is not so ill-behaving, though, that the basic model of an elliptical orbit need be abandoned altogether, it only needs some refinement. Basically the same discussion for unperturbed orbits applies to the moon as well, with the understanding that the earth is the central body instead of the sun, therefore the moon's elliptic path has the earth at a focus and the extremities of the apsidal line are termed the apogee and perigee. The concepts associated with an elliptical orbit, such as equAtion of centre, eccentricity, anomaly, etc. can be maintained.

The maximum equAtion of centre of the moon is about $6\frac{1}{3}°$, answering to an eccentricity of 0.055. An ellipse with the same eccentricity and with the diameter of a bicycle tyre exhibits a difference between its major and minor axes that would fit in the tread groove.[14] This illustrates how close the lunar orbit actually comes to a circle. Perturbations in parallax (as a measure of distance) are considerably less observable than perturbations in longitude.

The solar perturbations manifest themselves most visibly as slow changes in the orientation of the idealized elliptical lunar orbit. The rotation of the apsidal line has been mentioned earlier; the apogee completes a revolution in 3232 days. Additionally, the line of nodes travels backwards around the ecliptic in 6798 days, or in other words, the orbital plane of the moon rotates once around the axis of the ecliptic in that time. These periods are with respect to the reference point in the ecliptic, viz. the vernal equinox ♈.

The moon itself completes its orbit with respect to the equinox in 27.32 days on average, this is the length of the so-called tropical month. With the length of the year about 365.25 days and some basic arithmetic, it follows that the moon returns to the same position with respect to the sun in 29.53 days, this is the so-called synodic month after which the moon repeats its phases. Also, it returns to the same place with respect to its apogee in 27.55 days (this is called the anomalistic month), and to its ascending node in 27.21 days (the draconic month). All these periods are mean values only; deviations occur due to perturbations.

A striking near-commensurability arises between these periods, which is of particular importance to the prediction of eclipses (eclipses occur when the sun and moon are sufficiently close to the nodal line of the lunar orbit). It so happens that 223 synodic months, 239 anomalistic months, and 242 draconic months all are very nearly $6585\frac{1}{3}$ days, that is 18 years and $10\frac{1}{3}$ or $11\frac{1}{3}$ days, depending on the number of leap days. This period is known as the Chaldaean period, or Saros.[15] After a Saros the geometry of the earth-moon-sun system is very nearly repeated. This implies that if an eclipse occurs at a certain date, then an eclipse will occur again under very similar circumstances after one Saros has elapsed. Moreover, the solar perturbations on the moon's orbit are nearly cancelled out over this period, so that the lunar orbit is almost periodic over one Saros.[16]

The Saros is not only useful for eclipse predictions. For instance, Halley envisaged that an arbitrarily right or wrong set of lunar tables could be used to predict lunar positions with fidelity provided that its error be known for a date exactly one (or perhaps two or more) Saros intervals in the past. For this reason Halley set out to fulfil an entire Saros cycle of lunar position observations at the age of 65 after he was appointed the Astronomer Royal in 1720. Reversely, Mayer used the Saros periodicity whenever he met an exceptionally large difference between an observed and a calculated position of the moon, in order to verify whether the observation or

14 Let a be the semi-major axis, b the semi-minor axis, and $e = 0.055$ the eccentricity. These are related by $a^2 = b^2 + e^2a^2$; take $2a = 28$in (711mm) and verify that $2b = 27.96$in. Alternatively, rework the relation into $2(a-b) = 2\frac{e^2a^2}{a+b} \approx e^2a \approx 0.04$in, or 1mm.

15 The name 'Saros' was given to this time-span by Edmond Halley in 1691. Although the Babylonians were familiar with it and used it for eclipse predictions, they did not use that name for it.

16 [Roy, 1978, p. 285]; [Perozzi et al., 1991].

the calculation was faulty. Illustrative is also his *Catalogus Eclipsium Lunae*, a systematic comparison of lunar eclipse observations and predictions ordered according to the Saros principle.[17]

Thus far, we have seen that the moon's orbit can be approximated as an ellipse that slowly changes its orientation. The moon's position in such an orbit can be described by a set of *mean positions* and *mean motions*, which together specify the positions of the mean moon, mean apogee, and mean node as linear functions of time, with the adjective *mean* to signal that periodic terms are not taken into account. We have also met with the secular acceleration, which is actually an equAtion of such an extremely long period that its periodic nature is not immediately apparent. Next, the various eqAtions are taken into account, of which the lunar equAtion of centre is the largest.

We now turn to a brief historic overview of some other equAtions of the motion of the moon. As far as longitude is concerned, all equAtions consist of (sums of) terms of the form $c_k \sin k\alpha$ $(k \in \mathbb{N})$, where α is known as the argument and the c_k as coefficients.

Hipparchus (ca. 150 BC), building on the work of those before him, had developed a rather accurate model for solar and lunar motion. In his model, sun and moon went about the earth at constant speed in eccentric circular orbits, i.e., in circles whose centres did not coincide with the centre of the earth. The model incorporated the inclination of the lunar orbit to the ecliptic, the mean motions of the lunar apsidal and nodal lines, and his greatest discovery, the precession of the equinoxes. His model could predict lunar and solar eclipses reasonably well, but his observations showed discrepancies for lunar positions away from the phases of the New and Full Moon.

Ptolemy, ca. AD 150, constructed an elaborate theory[18] to account for the most prominent discrepancy occurring in the quadrants, i.e., near the first and last quarters. Ptolemy's model is successful in predicting lunar longitudes, but has a very unsatisfactory consequence of varying the earth-moon distance and hence also the apparent lunar diameter by a factor of nearly two. This was recognized and corrected independently by Ibn al-Shāṭir and Copernicus. Centuries later, the equAtion that Ptolemy had addressed in order to correct the position in the quadrants was named *evection* by Bulliau. It is an equAtion which Jeremiah Horrocks in 1638 successfully modelled as a variable eccentricity of an elliptical lunar orbit. This will be the subject of section 6.4. Evection, being the second largest of the lunar equAtions, may amount to $1\frac{1}{3}^{\circ}$. In the Mayerian notation that we adhere to, its argument is denoted as $2\omega - p$.

Further irregularities were discovered by Tycho Brahe shortly before 1600. The third equAtion in size and chronology is the *variation*, which attains a maximum of

17 The *Catalogus* was appended to [Mayer, 1754].

18 'Lunar theory', or theory in general, has long been the name for what we would prefer to call a (kinematic) 'model'. In the words of Francis Baily, in the 17th century it meant 'rules or formulae for constructing diagrams and tables that would represent the celestial motions and observations with accuracy' [Baily, 1835, p. 690], quoted in [Newton, 1975, p. 3].

some $\frac{2}{3}^\circ = 40'$. Its argument is 2ω, which is twice the angular separation of the sun and moon, hence it vanishes at the syzygies and quadratures.[19] One of the greatest successes that Isaac Newton achieved in *Principia* was his complete accounting for the variation on the basis of gravitation. Variation is principally caused by the veering of the solar direction of pull while the moon completes its orbit around the earth; in particular, it is independent of the lunar eccentricity.

Tycho and Johannes Kepler independently discovered the *annual equAtion*, which Newton later understood as being caused by the variable distance of the sun from the earth-moon system; it is brought about by the eccentricity of the earth's orbit. It reaches slightly more than $11'$ and has the solar anomaly ς as argument. Tycho also discovered the two prime equAtions of the lunar node and inclination. Kepler's was the first lunar theory to incorporate the idea of an elliptic orbit, his theory was however not very successful.

All these equAtions had been discovered observationally. Newton was the first to deduce new equAtions from theory, truly an outstanding feat in itself. With the refinement of mathematical methods since the eighteenth century, more and more new equAtions were derived, and the practice of baptizing equAtions was soon dispensed with. Most notable exception is the *parallactic equAtion* caused by the variable moon-sun distance over the course of a month.

This parallactic equAtion depends on argument ω, making for a convenient combination with the above mentioned variation when tabulated. By its nature, this equAtion provides a link between the earth-moon distance and the moon-sun distance, and hence between the lunar and solar parallaxes. The lunar parallax is quite large, and by the middle of the 18th century it was feasible to measure it almost directly. The Abbé Nicholas Louis de Lacaille was then temporarily observing at the Cape of Good Hope, a location well to the south of Europe which teemed with astronomers (in particular Lalande who was sent to Berlin, nearly on the same meridian), so that a sufficiently long baseline was available.

The solar parallax, on the contrary, happened to be one of the most sought after, yet elusive, constants in astronomy. Its value was crucial not so much because our knowledge of the size of the solar system depended on it, but because its erroneous value propagated into the orbital elements of all the observed solar system bodies. Two transits of Venus before the sun, first in 1761 and again in 1769, were carefully observed from various places on earth in order to deduce a value for the solar parallax. Unfortunately this subject, of the highest interest in the history of science, is beyond the scope of the current thesis. Let it suffice to remark that the link between the two parallaxes provided by the parallactic equAtion was signalled by Leonhard Euler, and exploited by Tobias Mayer, who obtained, in 1755, a value for the solar parallax of $7.8''$. His result is quite good against the yield of $8.5''$ to $10.5''$ produced after the Venus transit expeditions. The modern value of solar parallax is approximately $8.8''$.

19 Syzygy: the collective appellation of Full Moon and New Moon; quadrature: *idem* for First and Last Quarter.

To end this crash introduction to lunar theory, it may be well to point out that the difficulty of lunar theory lay not only in the complexity of the mathematics of the three-body problem, but also in the application of observed positions. It had to be taken into account that observations contain measurement errors, some of which could be dealt with approximately (such as atmospheric refraction), while others were inherently unknown (such as random observational errors); astronomers were among the first to realize this and to develop methods to deal with them. Moreover, it is no easy matter to determine for instance the position of the lunar apogee from observations: not only because the radius vector changes its length slowly within narrow limits, but also because all the other small inequAlities are superimposed on the elliptic motion. 'A major obstacle... lay in the multitude of small inequAlities, which they had no way of beginning to discern,' wrote Curtis Wilson.[20]

We can see mathematics at work in several ways. First of all, there is the mathematics of circles and ellipses which are used to construct kinematic models of motion. Then there is also the mathematics of Newton's *Principia*, later also the mathematics of differential equations and trigonometric series, not to mention Hamiltonian dynamics, forming the tools of *physical astronomy*, or *celestial mechanics* as it came to be known since Laplace. Finally, there is the mathematics of the 'combination of observations,' of inferring appropriate values for model parameters from data. The affiliation between astronomy and mathematics is indeed long and intense.

20 [Wilson, 1989a, p. 196].