

Chapter 3

Kepler's laws

So far we concerned ourselves with the motion of the stars – a reflex of the motion of the Earth. We considered also the motion of the Sun and the Moon, objects that move distinctly from stars, but also in a predictable and relatively simple way. In this chapter we will consider the five wandering stars whose motion seen from Earth, albeit also periodic, appears overwhelmingly complicated.

The question that will guide this chapter is a question that intrigued humanity for millenia: *how to predict the position of the planets?*

The solution of this problem is known as *Kepler's laws*, published by Johannes Kepler in his books *Astronomia Nova* (1609) and *Harmony of the World* (1619). Kepler's work was based on the meticulous observational data from Tycho Brahe, guided by the new heliocentric cosmology of Nicolaus Copernicus, and buoyed by the discoveries of his contemporary Galileo Galilei, with whom Kepler corresponded. Kepler's laws are the culmination of a century of thought that completely overhauled the previous cosmological model, and one of the most fascinating chapters in the history of astronomy.

But we are already getting ahead of ourselves. Let us dial the clock back a little, and understand how to extract the laws of planetary motion from the observations of the night sky.

3.1 The motion of the planets

The Sun and stars all move across the sky in a regular, predictable way. The planets, Mercury, Venus, Mars, Jupiter, and Saturn, however behave differently. They move mostly east to west (prograde motion) but would stop and revert they motion, moving west to east (retrograde motion) before stopping again and resuming the prograde motion. In addition, Mercury and Venus were only seen near sunrise and sunset.

The period of motion differs for each planets as well. If we define the period as the sidereal period, the time to return to the same position with respect to the fixed stars, then their ordering, from fastest to slowest should be Mercury (88 days), Venus (225 days), Mars (1.88 year), Jupiter (12 years), and Saturn (29 years).

Let us see what else we can learn from their motion in the sky. But first, we will setup some of the basic terminology to study planetary motion. The **elongation** is the angle between the planet and the Sun (Fig. 3.1).

We can follow the motion of each planet, and plot longitude and elongation versus time. These plots are shown for each planet, in Fig. 3.2–Fig. 3.6. The upper plot shows

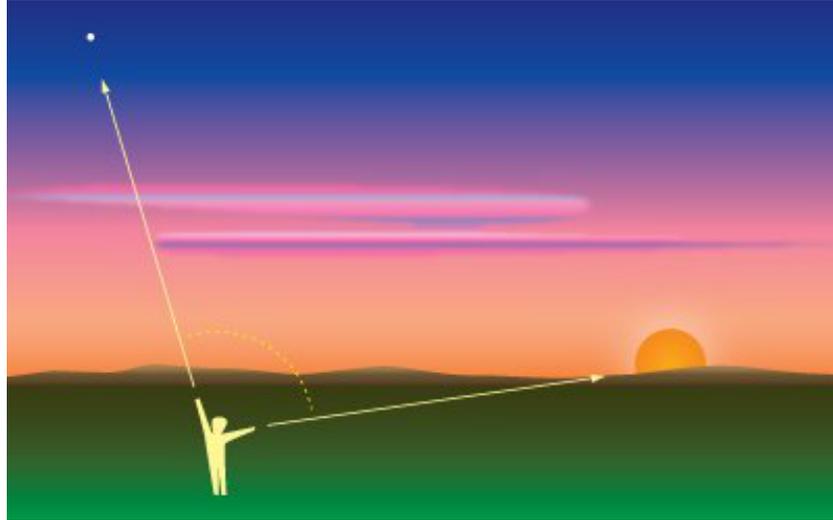


Figure 3.1: Elongation is the angle between the planet and the Sun.

the ecliptic longitude, of the planet (full line) and of the Sun (dashed line). The lower plot shows the elongation.

From the observations, we find that two planets (Mercury and Venus) keep a minimum and maximum elongation from the Sun. Mercury is never more than 28° away from the Sun, and Venus never more than 47° . Mars, Jupiter, and Saturn go all the way from -180° to 180° elongation. Clearly, there is something peculiar about Mercury and Venus.

These plots contain more information. Each planet seems to go forward in longitude, then stop and revert, going backward in longitude for a small amount, revert again, and resume the forward motion in longitude. They execute these motions in different periods.

Let us call the time to come back to the same longitude the *sidereal* period (since it means the same location with respect to the background stars), and the time to come back to the same elongation the *synodic* period (since it means the same location with respect to the Sun).

The plots show that the sidereal period of Mercury is the same as the Sun, 365.25 days. The synodic period is 116 days (Fig. 3.2). For Venus the sidereal period is again the same as the Sun, 365.25 days. The synodic period is 584 days (Fig. 3.3).

For Mars the sidereal period is 710 days, and the synodic period 780 days (Fig. 3.4). For Jupiter the sidereal period is 11.9 years, and the synodic period 400 days (Fig. 3.5). For Saturn the sidereal period is 29.5 years, and the synodic period 378 days (Fig. 3.6).

From a geocentric point of view, these different periods are the basis of Ptolemy's model. The planet is attached to a small circle whose center lies in the actual orbit around the Earth. The small circle, called an *epicycle*, occurs in the planet's synodic period. The orbit, called *deferent*, has the planet's sidereal period.

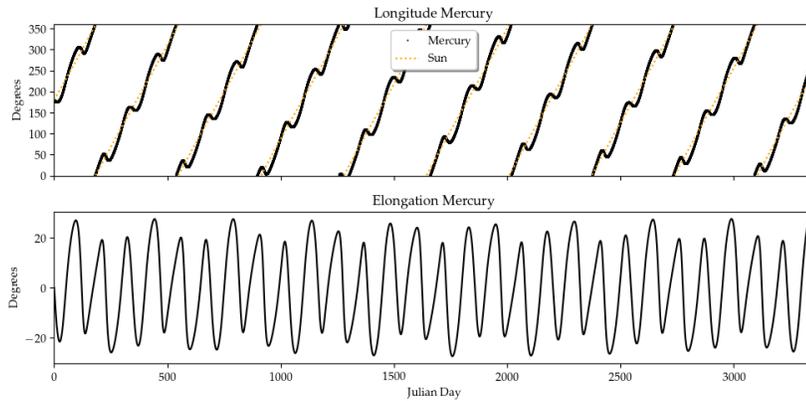


Figure 3.2: Longitude of Mercury, and elongation. Mercury never deviates much from the Sun, with maximum elongation between 18 and 28 degrees, average 23 degrees. Mercury does one revolution in the celestial sphere (360° in longitude) with the same period as the Sun. The period to return to the same elongation (synodic period) is 116 days.

3.1.1 Inferior planets

If we did not know anything about the order of the planets, it would make sense to order them based on their periods. Thus, the Moon is the fastest planet, going around the Earth in 28 days. The Sun, Mercury and Venus would come next, with their sidereal periods of 365 days, followed by Mars (710 days), Jupiter (12 years) and Saturn (29 yr).

Ordering the Sun, Mercury and Venus is not at all immediate judging on the observations alone. We are already used to the heliocentric system where Mercury and Venus are between the Sun and the Earth, but this has to be painstakingly derived from the observation, being not at all obvious. They follow the Sun, but how? Should Mercury and Venus be placed between the Earth and the Sun? Or between the Sun and Mars?

There is in fact a good reason for *not* placing them between the Earth and the Sun. If Mercury and Venus were between the Earth and the Sun, they would *transit* the disk of the Sun, and no such events had been recorded in classical antiquity. In lacking any parallax to determine their true distances, Ptolemy finds it more plausible¹ to have them placed between the Earth and the Sun simply because they behave differently than Mars, Jupiter, and Saturn, so it was natural to separate the two types of planets.

This led to their division as Mercury and Venus being referred to as *inferior planets*; Mars, Jupiter, and Saturn are *superior planets*.

The deferents and epicycles of inferior and superior planets are different: the inferior planets had deferent periods that followed the Sun (1 yr), whereas their epicycles had their sidereal periods. The opposite was true for the superior planet, having deferents with the sidereal period and epicycles that followed the Solar period.

As for the lack of transit, Ptolemy argues that maybe Venus and Mercury always

¹Almagest's book IX, pg 1.

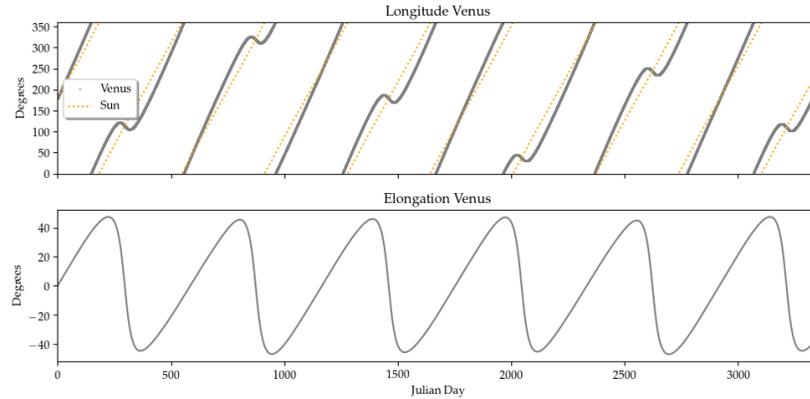


Figure 3.3: Longitude of Venus, and elongation. Like Mercury, Venus never deviates much from the Sun, with maximum elongation of 47 degrees. Venus does one revolution in the celestial sphere (360° in longitude) with the same period as the Sun. The period to return to the same elongation (synodic period) is 584 days.

pass above or below the Sun or that transits happen but the planets are too small and the Sun is too bright for any noticeable effect to occur with their instruments. We now know that the latter interpretation is correct.

In the Copernican system the arrangement of circles is far more harmonious. The inferior planets are between the Earth and the Sun. The maximum elongation happens because their orbits are smaller than Earth's orbit (Fig. 3.7). We define also the inferior and superior conjunction of the inferior planets as the places in the orbit where it is aligned with the Earth and the Sun. The maximum elongation happens when the line connecting the Earth to the planet tangents the planet's orbit. That is, the angle Sun-Planet-Earth is 90° .

That allows us to determine the distance to the planet through trigonometry.

According to the figure, the maximum elongation is one of the angles of a right triangle. The opposite side is the distance from the planet to the Sun d , and the hypotenuse the distance from the Sun to the Earth D . Solving it for Venus

$$\sin 47^\circ \approx 0.72 = \frac{d}{D} \quad (3.1)$$

That is, Venus distance to the Sun is 0.72 AU. For Mercury, the same procedure yields 0.39 AU.

In the Ptolemaic system, these numbers (0.72 and 0.39) would yield the ratio of the radii of the epicycle to the deferent of each planet (Fig. 3.9, left). So, the epicycle of Venus was gigantic, about 0.72 of the deferent. Mercury's epicycle was 0.39 of the deferent. These orbits had to fit between the Moon and the Sun.

Call V_d the radius of Venus' deferent and V_e the radius of Venus' epicycle. The minimum distance to the Earth is $V_{\min} = V_d - V_e = 0.18V_d$, the maximum is $V_{\max} = V_d + V_e = 1.72V_d$. For Mercury, its minimum distance is $M_{\min} = M_d - M_e = 0.61M_d$, the maximum is $M_{\max} = M_d + M_e = 1.39M_d$

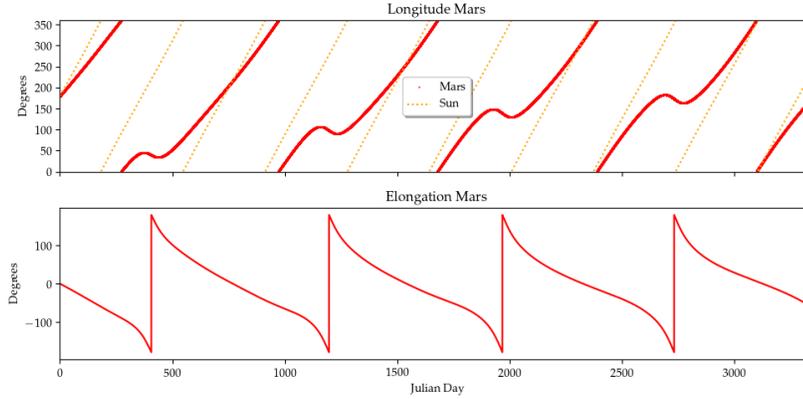


Figure 3.4: Longitude of Mars, and elongation. Mars has a full range of elongations, from 0° (conjunction) to 90° (quadrature) and 180° (opposition). Mars synodic period is 780 days.

Assume a very tight packing: Venus' epicycle touches the Sun's deferent and Mercury's epicycle (Fig. 3.9, right). That is,

$$V_{\max} = 1 \text{ AU} \quad (3.2)$$

$$V_{\min} = M_{\max} \quad (3.3)$$

Solving the system yields $V_d = 0.57 \text{ AU}$ and $M_d = 0.1 \text{ AU}$

With an epicycle-deferent ratio of 0.39, Mercury's perigee was at 0.06AU.

Now consider that the ancients measured the AU to be only 20 times the distance to the Moon (far from the actual value of $\times 400$). The Moon for them was at 0.05AU. Mercury was right next to the Moon. It had to be so close to the Earth that a parallax would be easily measurable.

Ptolemy was aware of this problem.

3.1.2 Superior planets

Superior planets can be seen at any elongation: they have only one conjunction with the Sun (0°), and when they are at 180° elongation they are said to be in *opposition*. When their elongation is 90° , they are in *quadrature*.

The Earth is an inferior planet for the superior planets. From the geometry of the orbit, when a superior planet is in quadrature, from the point of view of the planet the Earth is in maximum elongation.

Determining the size of a superior planet's orbit requires more work than for inferior planets. One way to do this is to record the amount of time t it takes for a superior planet to go from, e.g., opposition to the next quadrature (Fig. 3.10). Then the Earth will sweep out an angle

$$\alpha = \frac{2\pi}{T_\oplus} t \quad (3.4)$$

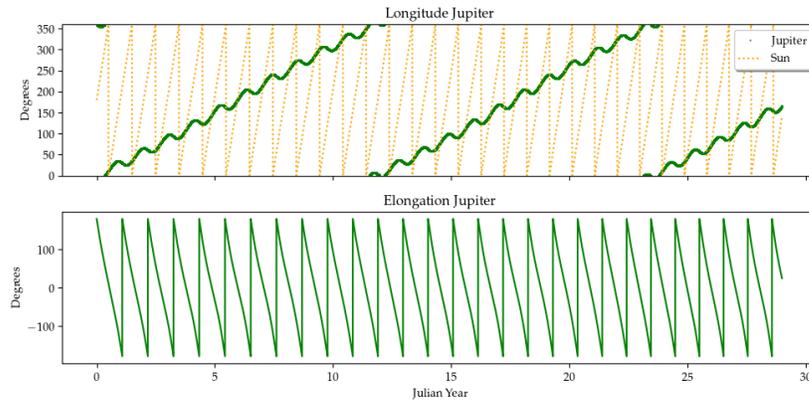


Figure 3.5: Longitude of Jupiter, and elongation. Jupiter has a full range of elongations, from 0° (conjunction) to quadrature (90°) and 180° (opposition). Jupiter sidereal period is 12 years and its synodic period is 400 days.

where T_\oplus is the orbital period of the Earth. If T_p is the period of the planet, then at the same time it sweeps out an angle

$$\beta = \frac{2\pi}{T_p} t \quad (3.5)$$

The difference of these two angles gives us the Earth-Sun-planet angle at quadrature. Thus, if d is the radius of the orbit of the superior planet, then

$$d = \frac{1}{\cos(\alpha - \beta)} \quad (3.6)$$

The formula valid at quadrature. Both angles are easily calculated knowing the orbital periods of the planet.

A variant of the method is to use two observations separated by one orbital period of the planet. Thus, the planet is at the same location in the orbit, but the Earth will have moved. This second method was used by Tycho Brahe and Johannes Kepler to determine the distance to Mars, in units of the Earth's distance.

Notice that both methods require knowing the orbital period of the planet.

3.2 Orbital Period

Notice that, in the Copernican system, the orbital period is *not* the same as the sidereal period as measured from Earth. Because the Earth moves, when the planet returns to the same position against the background stars, it is not necessarily at the same position in its own orbit.

We can tell the orbital period from the synodic period and Earth's period. Let us take the example of Mars, for instance, and count the synodic period between two oppositions, 780 days. At the end of 780 days, Earth has done 2 orbits plus 50 days, as it lapped Mars, that did between 1 and 2 orbits.

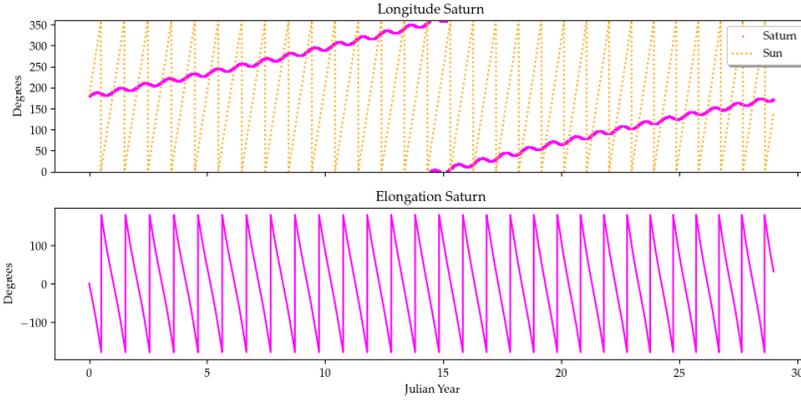


Figure 3.6: Longitude of Saturn, and elongation. Saturn has a full range of elongations, from 0° (conjunction) to quadrature (90°) and 180° (opposition). Saturn sidereal period is 29 years and its synodic period is 380 days.

At the end of 780 days, Earth has done extra $50/365$ past 2 orbits. Because Mars and Earth are now in opposition, Mars has done $50/365$ past 1 orbit. So,

$$780 = 2 + 50/365 T_{\text{Earth}} \quad (3.7)$$

$$780 = 1 + 50/365 T_{\text{Mars}} \quad (3.8)$$

Solving for T_{Mars}

$$T_{\text{Mars}} = \frac{780}{1 + 50/365} \quad (3.9)$$

yielding $T_{\text{Mars}} = 687$ days.

What we did can be better mathematized by realizing that this amounts to changing the reference frame to the Earth, thus transferring its angular velocity to the other body. That is, we can define the synodic frequency ω_{syn} between bodies 1 and 2 as

$$\omega_{\text{syn}} = \omega_1 - \omega_2 \quad (3.10)$$

And the synodic period is its inverse, $T_{\text{syn}} = 2\pi/\omega_{\text{syn}}$. Thus,

$$\frac{1}{T_{\text{syn}}^{\text{inf}}} = \frac{1}{T_1} - \frac{1}{T_2} \quad (3.11)$$

If the reference frame is *Earth*, then for the inferior planets

$$\frac{1}{T_{\text{syn}}} = \frac{1}{T_P} - \frac{1}{T_E} \quad (3.12)$$

And solving for their orbital period

$$T_P = \frac{T_E T_{\text{syn}}}{T_{\text{syn}} + T_E} \quad \text{for inferior planets} \quad (3.13)$$

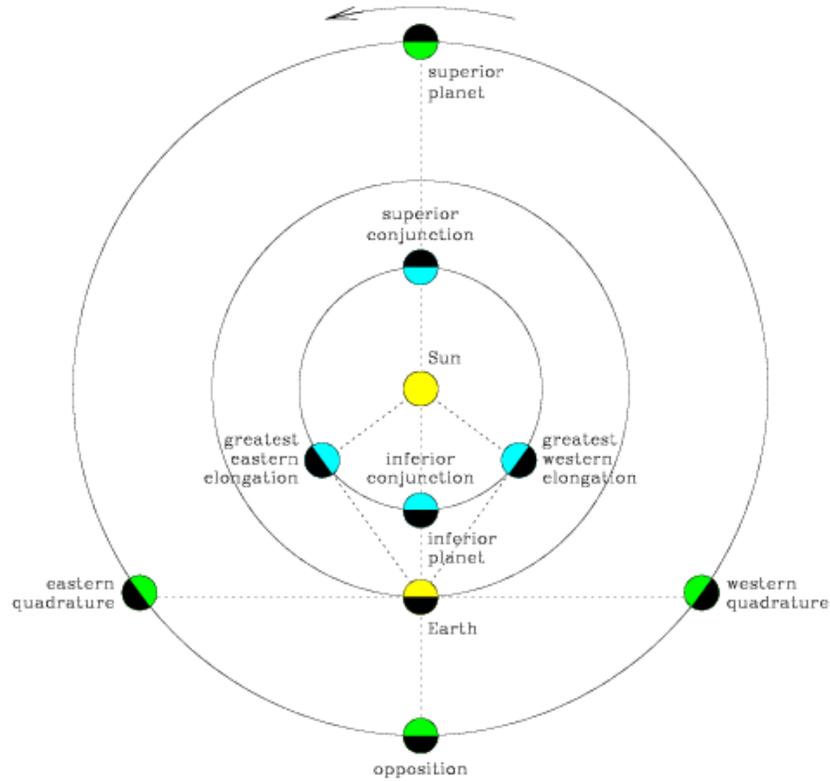


Figure 3.7: Special points in the orbits of inferior and superior planets.

For Mercury, of synodic period 116 days, this yields an orbital period of 88 days. For Venus, of synodic period 584 days, the orbital period should be 225 days.

For the superior planets now Earth is the fastest planet, so to keep T_{syn} positive,

$$\frac{1}{T_{\text{syn}}^{\text{sup}}} = \frac{1}{T_E} - \frac{1}{T_P} \quad (3.14)$$

Again solving for the orbital period

$$T_P = \frac{T_E T_{\text{syn}}}{T_{\text{syn}} - T_E} \quad \text{for superior planets} \quad (3.15)$$

For Mars, of synodic period 780 days, the orbital period is 678 days. For Jupiter, plugging $T_{\text{syn}} = 400$ days yields $T_P = 4204$ days, or 12 years. For Saturn, $T_{\text{syn}} = 378$ yields $T_P = 10829$, or 29.5 years.

Notice that the furthest the planet is, the closest the synodic period gets to the orbital period of the Earth. Also, the closest the sidereal period gets to the actual orbital period.

The time between opposition and quadrature for Mars is about 105 days. This yields, according to Eq. (3.6), a distance of 1.5 AU.

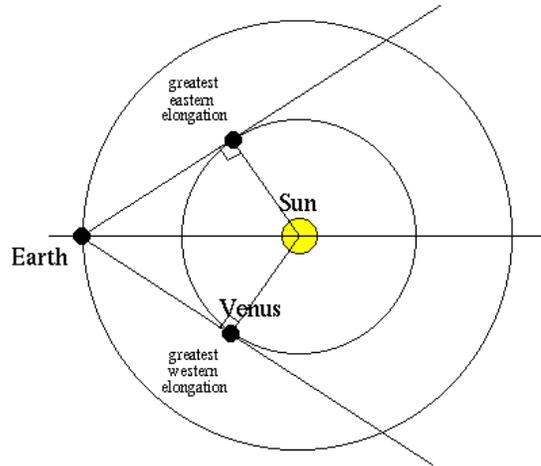


Figure 3.8: Maximum elongation of inferior planets happen at right angles, i.e., when from the point of view of the inferior planet the Earth is in quadrature.

Based on Tycho's and Kepler's method, we could also do two different observations of Mars, spaced by 687 days, Mars' orbital period. Because Mars would be at the same position in the orbit, we could thus triangulate the distance (Fig. 3.12). This was the method used by Kepler to find the shape of Mars' orbit (Fig. 3.12).

3.3 The orbit of Mars

With several sets of observations spaced by Mars' orbital period, the orbital path of Mars can be traced (Fig. 3.13). Centered on the Sun one cannot trace a circle that traces all positions of Mars. The orbit of Mars as a sort of off-centered circle is clear.

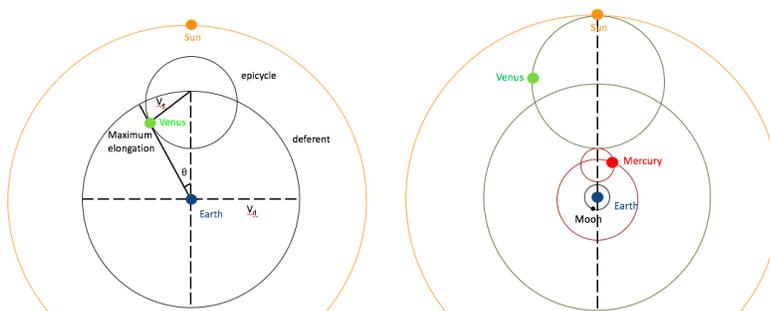


Figure 3.9: In the Ptolemaic system, the maximum elongation was the ratio between the radius of the epicycle and the radius of the deferent.

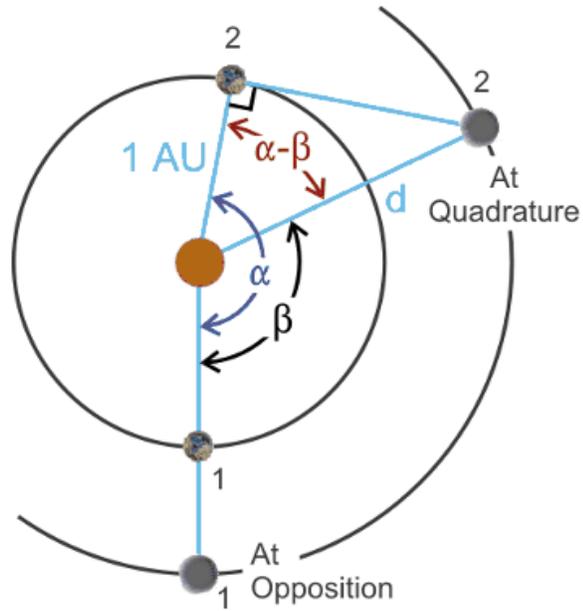


Figure 3.10: Distance determination for superior planets.

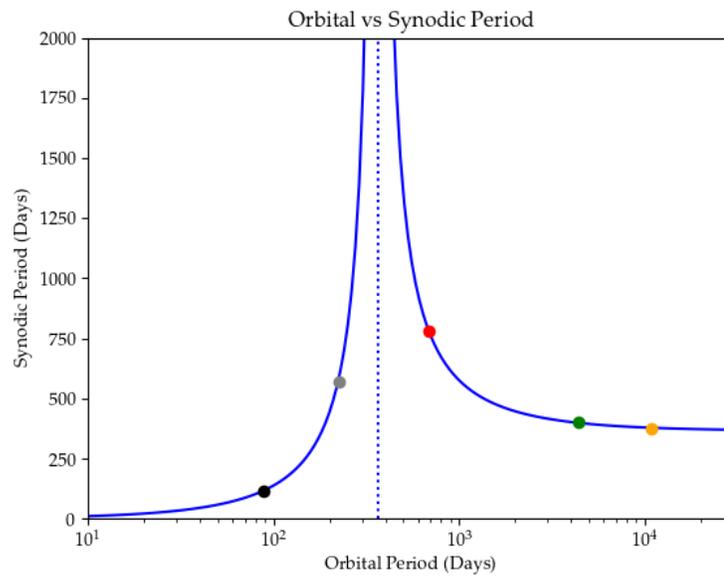


Figure 3.11: Orbital vs synodic period. The dots mark the position of the planets.

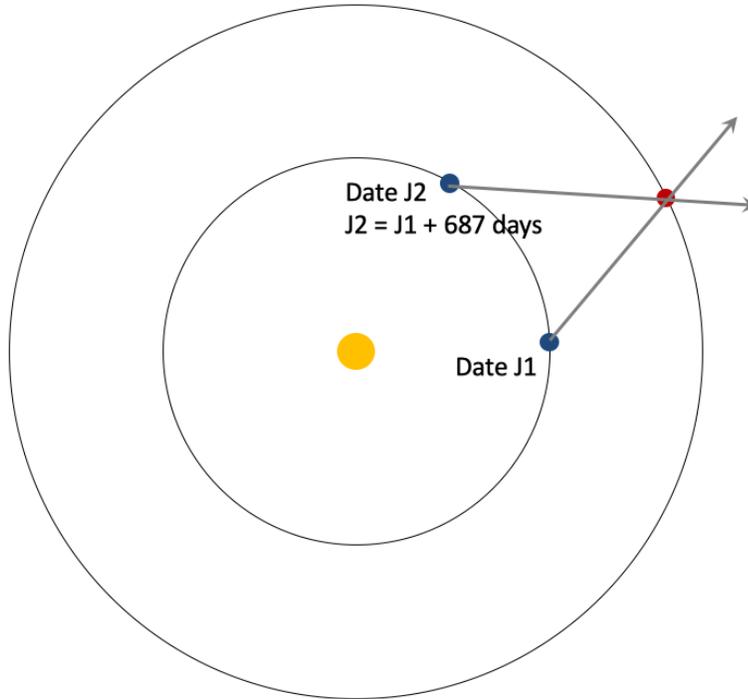


Figure 3.12: Determining Mars position. If two observations are spaced in time by Mars' orbital period of 687 days, Mars will have returned to the same location, but the Earth will be elsewhere, allowing distance determination in AU.

The perpendicular bisector does not strike the line of apsides at the position of the Sun, but at a position about 0.15 AU away from it. Let us measure the distances now in terms of half the length of the line of apsides, and give it the symbol a . The line of apsides measures $2a$. Let the distance from the center of the line of apsides to the Sun be defined as ae , where e is the *eccentricity*. The line of apsides measures $2a = 3AU$. The eccentricity of Mars orbit is thus

$$e = \frac{ae}{2a/2} = \frac{0.15}{1.5} = 0.1 \quad (3.16)$$

3.3.1 Elliptical orbits

Having found the orbit, Kepler had no idea what geometrical shape it corresponded to. Yet, Kepler realized, through geometry, some properties this shape had. Consider Fig. 3.14. The Sun is at point A, and Mars at point P. The distance $AP = r$ is the radius vector from the Sun to the planet.

The orbit is the red curve, which, a priori, we do not know what shape it corresponds to. Some elements are

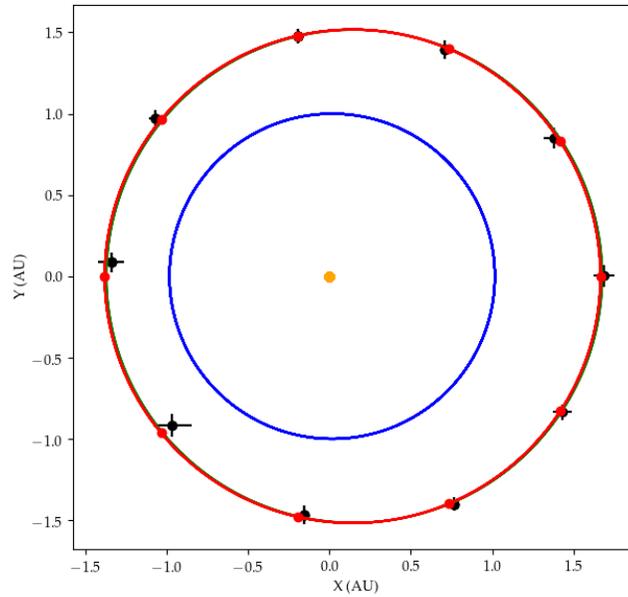


Figure 3.13: In red, Mars' true orbit, with data overplotted.

- The aphelion is point C;
- The line of apsides is bisected at point B, the geometric center of the curve. The distance \overline{BC} is by definition a .
- The distance \overline{AB} is by definition $AB=ae$.
- The perpendicular from P to the line of apsides defines the point H. The coordinates of the P are $x = \overline{BH}$ and $y = \overline{PH}$.

A relation between x and y will tell the shape of the curve.

We prolong the line \overline{PH} until it intersects the circumference at point Q and define the auxiliary angle $E=\widehat{HBQ}$. Given the triangle $\triangle BHQ$, the coordinate x is $BQ \sin\beta$. Given $BQ=a$, we found the first coordinate

$$x = a \cos\beta \quad (3.17)$$

As for y , the triangle $\triangle AHP$ can be used. It is a right triangle where $\overline{AP} = r$ is the hypotenuse; the catheti are $\overline{PH} = y$, and $\overline{AH} = \overline{AB} + \overline{BH} = ae + x$. Thus,

$$y^2 = r^2 - (ae + x)^2 \quad (3.18)$$

Eq. (3.17) gives the value of x , but the value of the radius vector r is so far unknown.

Kepler found r in an ingenious way. The grey circle has center A and radius r . We trace the circle in the hope that a geometric coincidence that helps tell what the length

$$y^2 = a^2 \left[(1 + e \cos \beta)^2 - (e + \cos \beta)^2 \right] \quad (3.20)$$

$$= a^2 (1 + e^2 \cos^2 E - e^2 - \cos^2 E^2) \quad (3.21)$$

$$= a^2 (1 - e^2) (1 - \cos^2 E) \quad (3.22)$$

$$= a^2 (1 - e^2) \sin^2 E \quad (3.23)$$

We can then write $\cos \beta = x/a$, and $\sin \beta = y/(a\sqrt{1-e^2})$ and invoke the trigonometric equality $\sin^2 \theta + \cos^2 \theta = 1$ to find the relationship between the coordinates

$$\frac{y^2}{a^2(1-e^2)} + \frac{x^2}{a^2} = 1 \quad (3.24)$$

This is the equation of an **ellipse**. The semimajor axis is a , and the semiminor axis is $b = a\sqrt{1-e^2}$.

We defined the Sun at the distance ae from the center. It turns out that this is the focal distance, which can be proven by the property of the ellipse that the distance between a point at the ellipse and the two foci is equal to $2a$. Let F be the point where the minor axis intersects the ellipse. A point at the minor axis is equidistant from the two foci, so the distance between F and the focus must be a .

Consider the triangle $\triangle AFB$. The hypotenuse is \overline{AF} , the cathetus \overline{AB} is according to our definition equal to ae , and we found the cathetus \overline{BF} to be $b = a\sqrt{1-e^2}$. Thus,

$$AF^2 = ae^2 + a^2(1-e^2) = a^2 \quad (3.25)$$

We find that $\overline{AF} = a$, so A, the Sun, is a focus.

A planet travels in an ellipse with the Sun at one focus.

This result is known as Kepler's 1st law of planetary motion.

3.3.2 Tycho's accuracy

Before we advance to Kepler's other laws, let us understand why no one before Kepler found out about elliptic orbits. It has to do with the fact that even an ellipse of eccentricity 0.1 is not too far from a circle.

In Fig. 3.15 we show in red the orbit of Mars, an ellipse of eccentricity 0.1, and in orange a circle whose radius is equal to the semimajor axis of the orbit of Mars. The orange circle is centered at the Sun.

It is obvious that the Sun is not at the center of the orbit (the black dot), but the equant and the eccentre, which offset an orbit from an otherwise central body, have been an element of planetary orbital models since Ptolemy. In blue, we plot a *circle* centered at the black dot. Its radius is identical to the orange circle. It barely differs from Mars' ellipse. The ratio b/a between the semiminor and the semimajor axes is $b/a = \sqrt{1-e^2}$, which for $e = 0.1$ yields $b/a \approx 0.995$. Less than 1%.

The difference in the position of Mars in an ellipse (point P) and in an off-centered circle (point Q) is the length \overline{PQ} . Given $\overline{QH} = a \sin \beta$ and $\overline{PH} = a\sqrt{1-e^2} \sin \beta$, the length \overline{PQ} is

$$\overline{PQ} = \overline{QH} - \overline{PH} = a \sin \beta (1 - \sqrt{1-e^2}) \quad (3.26)$$

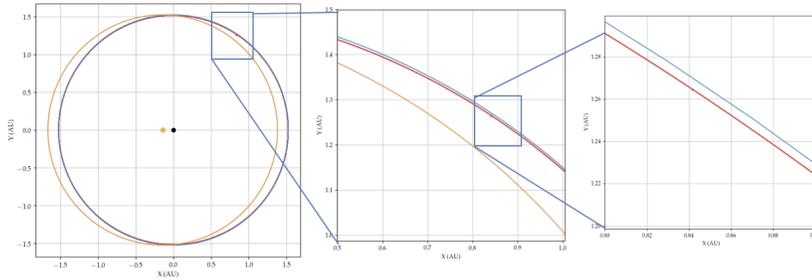


Figure 3.15: Although it is clear that Mars orbit (red curve) is not centered at the Sun (orange dot, with associated Sun-centered orange circle), Mars's orbit does not deviate too much from an off-centered circle (blue line). With our instruments, of 1° precision, we did not detect the eccentricity, and an off-centered circle would have been an acceptable model. To the right, the panels zoom in the difference between the red and blue lines, showing a ~ 0.0075 AU difference, as expected (see Sect 3.3.2). Seen from 1.1 AU away, the difference between the ellipse and the off-centered circle amount to about 20 arc minutes. Kepler used Tycho's data, that had precision of 3 arcminutes in general.

For small eccentricities, we can Taylor-expand the square root to first order, and find $\overline{PQ} \approx ae^2/2 \sin \beta$. Let us assume that $\sin \beta = 1$, then $\overline{PQ} \approx ae^2/2$. For Mars, $a_\oplus = 1.5$ and $e = 0.1$, then $\overline{PQ} \approx 0.0075$ AU.

Seen from Earth at the distance of quadrature $d = \sqrt{a_\oplus^2 - a_\oplus^2} \approx 1.1$ AU, this length corresponds to an arc $\theta \sim 0.075/1.1 \sim 20$ arc minutes.

Given our 1 degree precision, we cannot detect this difference.

Kepler was able to conclude that Mars's orbit was an ellipse because his data had enough precision. He was using the data from Tycho Brahe, which generally had 3 arcminute precision.

3.4 Orbital Elements

For twodimensional elliptical orbits, the geometry of the orbit is specified by the semi-major axis a and the eccentricity e . The orientation of the orbit is given by the longitude of the perihelion $\bar{\omega}$. The position of the planet is given by the true anomaly f . These four quantities would specify the orbit if the plane is given. For an orbit at the ecliptic plane, these suffice. If the orbit is inclined, two extra parameters are needed to specify the orbit. These are the inclination i and the longitude of the ascending node Ω .

These six quantities are referred to as the *orbital elements*. For an inclined orbit, instead of longitude of perihelion, which is measured from the vernal point, one uses the argument of perihelion, which is measured from the ascending node. The orbital elements are

- semimajor axis a
- eccentricity e

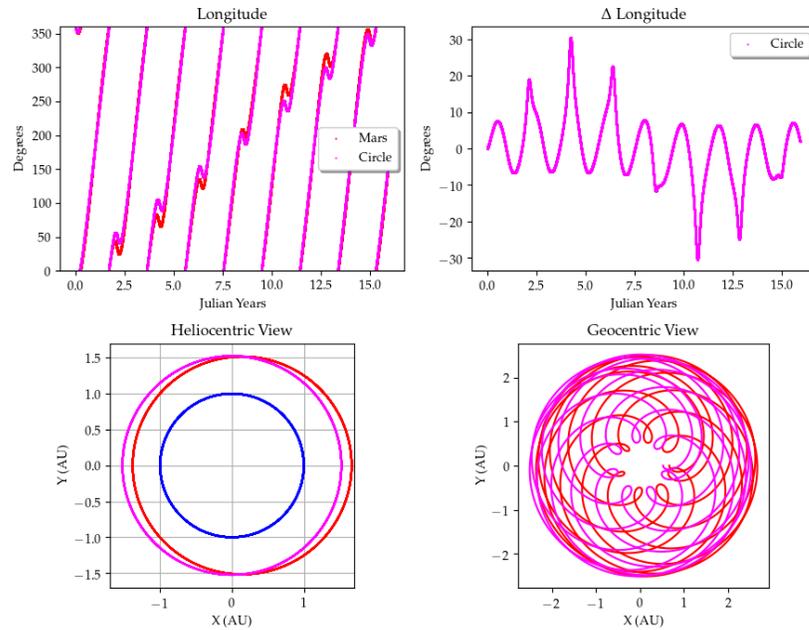


Figure 3.16: Mars orbit versus circular Sun-centred orbit. Upper left plot: Ecliptic longitude vs Time, Mars (red) vs circular orbit (magenta). The longitudes generally match, except at retrogradations. Upper right plot: Longitude residual. The error amounts to as much as 30 degrees. The model is not acceptable. Lower left plot: Heliocentric view of the orbit. Red is mars, blue is Earth, magenta the model. Lower right plot: Geocentric view of the orbit.

- inclination i , measured from the ascending node.
- longitude of the ascending node Ω , measured from the vernal point.
- argument of the perihelion ω , measured from the ascending node.
- true anomaly f , measured from perihelion.

The longitude of perihelion is

$$\bar{\omega} = \Omega + \omega \quad (3.27)$$

which is strange at first because it is a combination of an angle in the ecliptic (Ω) and one in the orbital plane (ω). Yet, as seen, this is useful because it is defined even in ecliptic orbits, where Ω and ω are not defined. The longitude of the planet is $\lambda = \bar{\omega} + f$.

3.5 Kepler's second law

Finding the shape of the orbit is not solving the whole problem of planetary motion. A practical question remains: how to find the planet in the orbit? Ancient wisdom

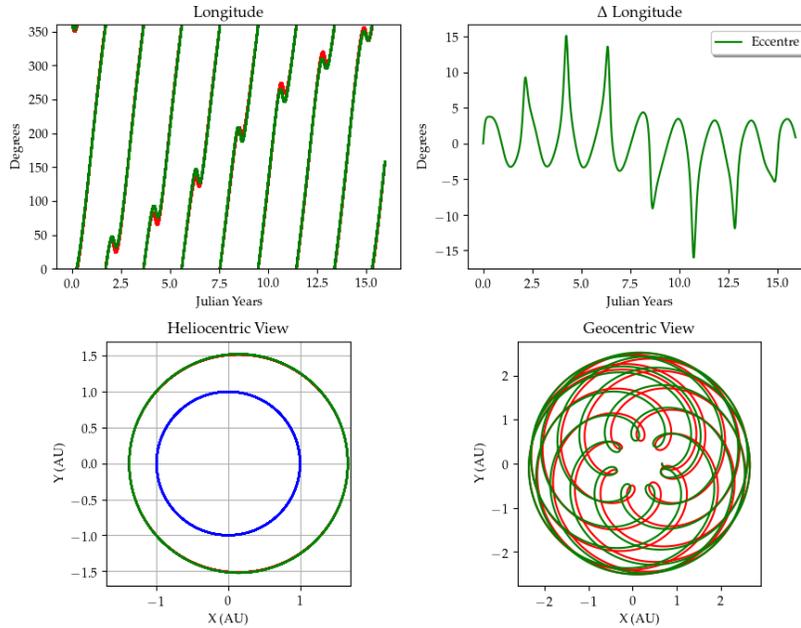


Figure 3.17: Mars orbit versus off-centered circular orbit (the eccentre), keeping uniform motion. Upper left plot: Ecliptic longitude vs Time, Mars (red) vs circular off-centered orbit (green). The longitudes generally match, except at retrogradations. Upper right plot: Longitude residual. The error is better than the Sun-centered model, but still amounts to as much as 15 degrees. The model is not acceptable. Lower left plot: Heliocentric view of the orbit. Red is mars, blue is Earth, green the model. Lower right plot: Geocentric view of the orbit.

insisted in uniform circular motion, because it was their way to understand periodicity. Regularity was found in circles, according to Copernicus “the only figure that can bring back the past”. Although they could see that the Sun (or the Earth for that matter) was not the center of the orbit, the shape of the ellipse was out of the reach of their observational accuracy. Non-uniform motion along the orbit was also evident, and a solution was sought by Ptolemy, the equant, which is crucial to Kepler’s 2nd law.

3.5.1 The Equant model

Let us assume that both Earth and Mars go on circular orbits, centered at the Sun, in uniform motion. At any instant of time, the position of Mars is given by

$$x(t) = a \cos M(t) \quad (3.28)$$

$$y(t) = a \sin M(t) \quad (3.29)$$

where $M(t) = \Omega t$, where t is time, and $\Omega = 2\pi/T$ where T is the period of Mars. Here, M , the mean anomaly, is equal to both the eccentric and true anomalies. Notice

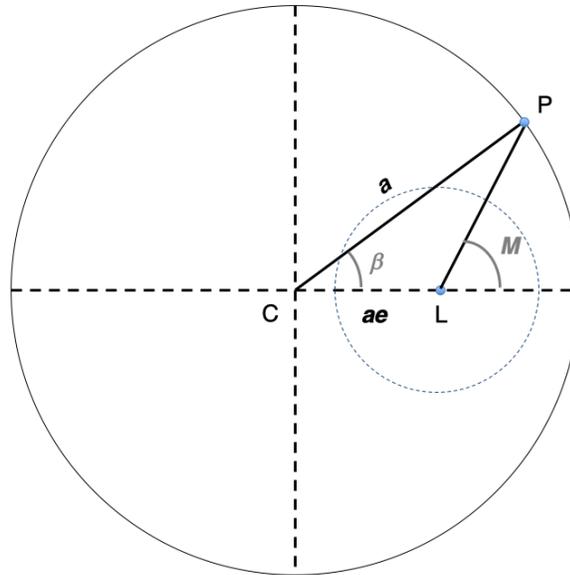


Figure 3.18: Elements of the equant model.

that everything but time in the definition of mean anomaly is constant. Mean anomaly equals time. Mean anomaly is time.

The four panels in Fig. 3.16 show (1) upper left plot: the longitude vs time for this model (circular and sun-centred orbit; magenta) and the actual Mars, in red; (2) upper right plot: deviation between model and actual Mars; (3) lower left plot: The heliocentric view of the orbit; (4) lower right plot: The geocentric view of the orbit.

The model does not reproduce either the shape of the orbit or the longitudes. The predicted positions of retrogradations, specifically, are off by as much as 30 degrees from the actual positions of Mars. The model has to be discarded.

The next model is the eccentre. This model merely shifts the position of the center of the orbit away from the Sun by an amount ae , keeping the uniform motion. At any given instant in time, the position of Mars is now given by

$$x(t) = a \cos E(t) + ae \quad (3.30)$$

$$y(t) = a \sin E(t) \quad (3.31)$$

where ae is the amount we shift the center away from the Sun. The angle E , the eccentric anomaly, is $E(t) = \Omega t$ and equal to the mean anomaly M .

This model is shown in green in Fig. 3.17. It reproduces the orbit, but does not reproduce the velocity of Mars. It predicts oppositions now half the way to the true position, still up to 15 degrees off. Again, this model cannot be right.

To fit the velocity of Mars, Ptolemy added a third device, the equant point, defined as a point about which the angular velocity of a body on its orbit is constant. This point is point L in Fig. 3.18. About L, the planet, located in P, goes around in uniform

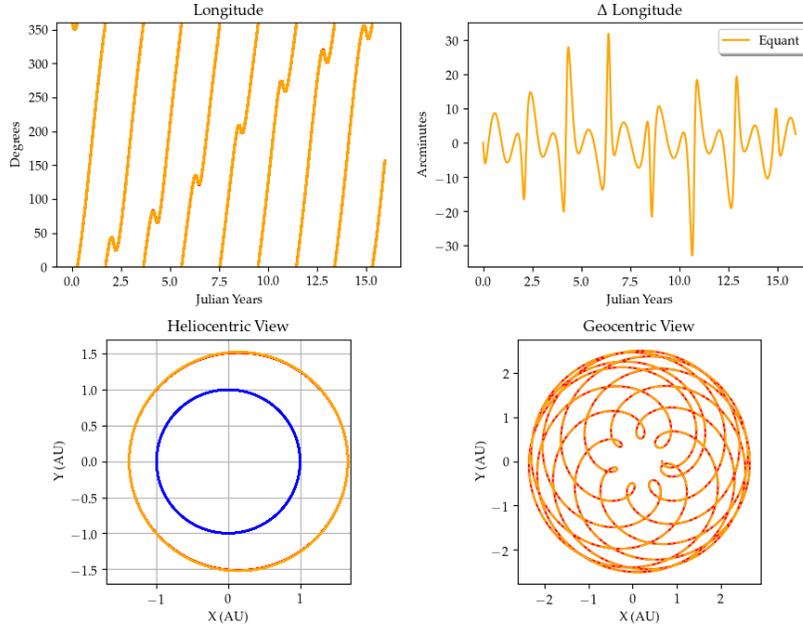


Figure 3.19: Mars orbit versus off-centered circular orbit, with uniform motion about the equant. Upper left plot: Ecliptic longitude vs Time, Mars (red) vs equant model (orange). Upper right plot: Longitude residual. The error is at most 30 arc minutes. Ptolemy's accuracy was 1 degree. The model is acceptable. Lower left plot: Heliocentric view of the orbit. Red is mars, blue is Earth, orange the equant model. Lower right plot: Geocentric view of the orbit. The equant model reproduces location and time of retrogradations.

motion, being described by the angle $M = \Omega t$. The angle β seen from the center of the orbit is related to M by noticing that the triangle ΔPCE has angles $P\hat{E}C = 180^\circ - M$ and $E\hat{P}C = M - \beta$. Applying the law of sines,

$$\frac{\sin(M - \beta)}{ae} = \frac{\sin(180^\circ - M)}{a} \quad (3.32)$$

but $\sin(180^\circ - M) = \sin M$, so

$$\sin(M - \beta) = e \sin M \quad (3.33)$$

Thus, solving for β

$$\beta = M - \sin^{-1}(e \sin M) \quad (3.34)$$

At any time, the position of Mars is given by

$$x(t) = a \cos \beta(t) + a\varepsilon \quad (3.35)$$

$$y(t) = a \sin \beta(t) \quad (3.36)$$

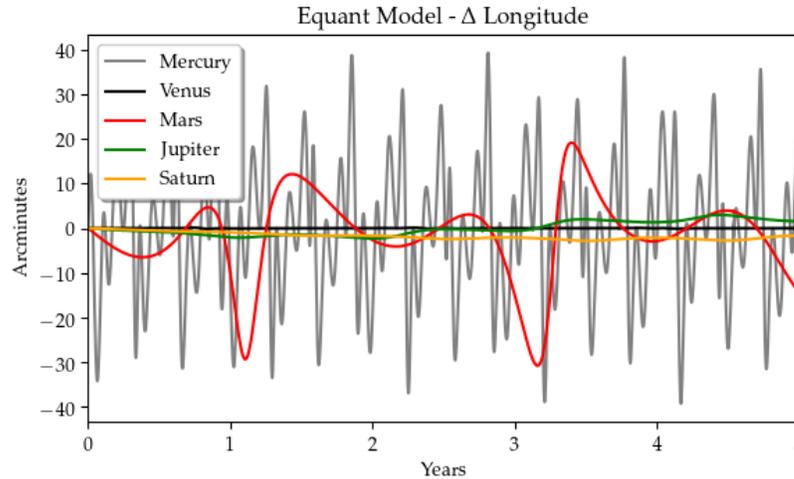


Figure 3.20: Residuals of the equant model for each planet. The residuals reflect orbital eccentricity. Even for Mercury, the most eccentric planet ($e = 0.2$), the equant agrees with the observations down to 40 arcminutes.

The result is shown in Fig. 3.19. The agreement is very satisfactory, to within half a degree. Considering that Ptolemy did not have accuracy under a degree, the equant gives excellent agreement to the observations. Fig. 3.20 extends the equant model to other planets. Each of them has their own equant – which simply reflects the eccentricity of the orbit. Even in the case of Mercury, the planet of highest eccentricity, the agreement with the observations is satisfactory to the degree. This figure also shows why Mars was the subject of Kepler's analysis. Of the outer planets, it is the one whose deviation more blatantly disagreed with the prevailing model. Venus, Jupiter, and Saturn deviate by less than 2 arcminutes, within the accuracy of Tycho's data. Mercury, never too far from the Sun, is simply too difficult to observe.

Notice that there is no reason why the equant has to be equidistant from the center as the central mass. That is, a priori, e and ε need not be the same. The model agrees with the observations to 1 degree accuracy for e to within 5% of ε .

Notice though, that at higher eccentricities the equant model will again start to deviate significantly from the observations. The last figure shows a hypothetical planet of eccentricity $e = 0.45$. The equant model is off by more than 10 degrees. Ptolemy's equant (Eq. 3.34) is a 1st order expansion in terms of eccentricity. For $e = 0.1$ it is satisfactory down to 1 degree accuracy. For $e = 0.45$ one would need higher order corrections.

Ptolemy's solution had a very practical function. Given a point in the line of apsides upon which the planet sweeps equal angles in equal times, the orbit can be parametrized, as given by Eq. (3.34)-Eq. (3.36).

Kepler had two problems. First, Tycho's observations, accurate to 2 arcmins, did not allow for the 30 min error given by the equant model.

Also, his ellipses, with the planet speeding up at perihelion and slowing down in

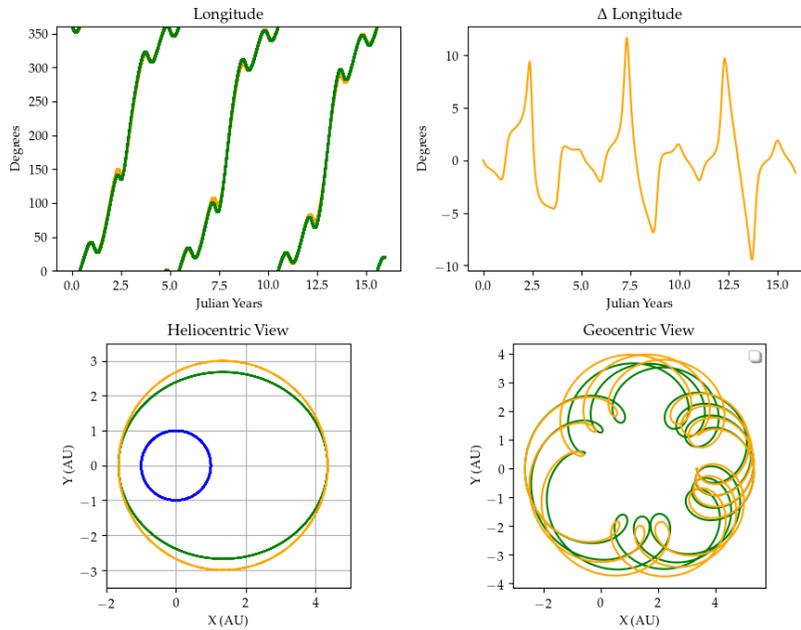


Figure 3.21: A hypothetical planet (green line) with orbital eccentricity 0.45, The equant model (orange) does not reproduce it well anymore. It is essentially a Taylor expansion to 1st order in eccentricity.

aphelion begged the question: what is the equivalent to the equant? What is the point about which a planet sweeps equal angles in equal times? Where is the point along the line of apsides that we can say that angle equals time? As it turns out, Kepler's quest to answer this question culminated with his revolutionary 2nd law, that demolished the idea of uniform motion. The answer is: *there is no equant*.

3.5.2 The impossibility of uniform motion

In trying to find out the location of Mars' equant, Kepler again made use of Tycho's observations. He took four observations of Mars, spaced in time t_1 , t_2 , t_3 , and t_4 . From the point of view of the equant, these observations had to correspond to mean anomalies M_1 , M_2 , M_3 , and M_4 . The equant would be the one point from which Mars was seen at these angles (Fig. 3.22, left panel).

These four observations Kepler used were observations in opposition. This means that the Sun, Earth, and Mars are aligned and the direction we see Mars is the same direction the Sun sees Mars. Lines connecting the four opposition observations intersect at the Sun (Fig. 3.22, right panel).

The challenge now is to match these geometrical constructs to have the Sun, the center of Mars' orbit, and the equant lie on the line of apsides. Kepler's solution is shown in Fig. 3.23. The Sun is the red dot, the equant the blue dot, and the center of the orbit the green dot.

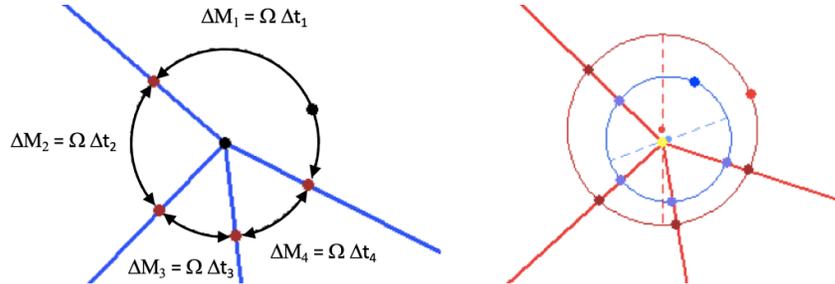


Figure 3.22: The images above show four oppositions of Mars. In the right plot, the blue circle is the orbit of the Earth, the red circle the orbit of Mars. When in opposition, the Sun, the Earth and Mars are aligned, and we see Mars in the same direction as the Sun sees it. Essentially, we change the reference frame to heliocentric. Where the 4 directions intersect is the location of the Sun. The bisection of the line of apsides gives the orbit center. This way, the Sun and the center of the orbit are located.

The left plot shows how to locate the equant. Time the observations. From the reference frame of the equant these time intervals leads to proportional angles. The mean motion is $\Omega = 2\pi/T$ where T is the orbital period. Adjusting the angles to the positions of Mars would reveal the position of the equant.

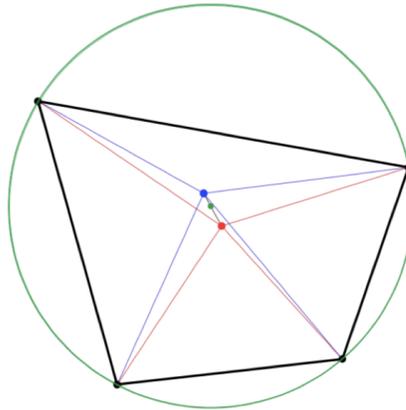


Figure 3.23: Kepler's solution for locating the equant. The Sun is at $ae_1 = 0.11$ and the equant at $ae_2 = 0.07$. This model reproduces Mars' observations to 2 arcminutes. Yet, Tycho's observations did not allow for eccentricities outside $e = 0.07 - 0.09$. Placing $e_1 = e_2$ as in Ptolemy's model, with $e = 0.09$, did not improve the fit, it made it worse in fact, with 8 arcmin discrepancy. This analysis shows the equant cannot exist.

For a circle with a radius of 1, the eccentricity of the Sun (green to red) is $e_1 = 0.11$, and that of the equant (green to blue) is $e_2 = 0.07$. This is Kepler's "best fit" to Tycho's opposition data; with this model he could fit Mars' position to within the observational errors of 2 arcmins.

However, Kepler had already measured the eccentricity of Mars. The whole wealth of Tycho's data did not support eccentricities out of the range 0.80-0.99. The eccentricity of 0.11 could not be correct.

Next Kepler tried to fit the data by using the Ptolemaic construct of having the center of the orbit bisect the Sun-equant line. That is, have $e_1 = e_2$. Since $e_1 + e_2 = 0.18$, then $e_1 = e_2 = 0.09$, which is within the range of eccentricities allowed by Tycho's data.

Kepler uses the updated model and checks against the data. The agreement got significantly worse: 8 arcminutes. In his words:

If I had believed that we could ignore these eight minutes, I would have patched up my hypothesis accordingly. But since it was not permissible to ignore them, those eight minutes point the road to a complete reformation of astronomy.

Kepler had to go back to question his assumptions. But the assumptions were minimal. They amounted to

1. Mars orbits the Sun;
2. The equant exists;
3. Tycho's observations are reliable.

(1) and (3) were beyond doubt correct. The conclusion was astonishing. The equant, a staple of astronomy for 1500 years, cannot exist.

Kepler started this analysis by asking the question: where is the equant? And the answer was: there is no equant.

Kepler ruling out the existence of the equant is truly revolutionary. There is no point about which we can say the planet sweeps equal angles at equal times. Uniform motion does not exist. Time is not given by angle.

3.5.3 Time is equal area

Kepler had disproved fifteen centuries of "equal angles at equal times". That still leaves the problem of how to find the true anomaly as a function of time.

In looking for something that could be a measurement of time, Kepler stumbled, on a hunch, on what this something was.

Because planets are slower when far from the Sun and faster when close, Kepler reasoned that the velocity was inversely proportional to the distance, $v \propto 1/r$. If that is the case, you can multiply both sides by time and write the equality

$$vrt \propto t \tag{3.37}$$

The quantity in the left hand side, whatever it is, is linearly proportional to time. It is the something sought in equal "something" at equal times. But what is its interpretation?

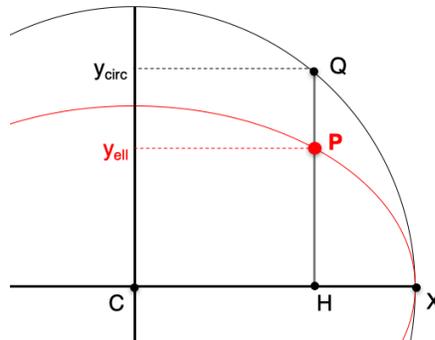


Figure 3.24: The ratio $PH/QH = y_{\text{ell}}/y_{\text{circ}}$ is equal to b/a .

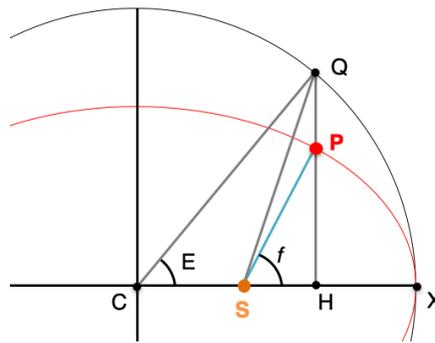


Figure 3.25: The true anomaly f is the angle, with vertex at the Sun, from perihelion to the planet. The eccentric anomaly E is the angle with vertex at the center of the orbit, from perihelion to the planet.

The product vt is the length of the arc swung by the planet. If the time is infinitesimal, $t \rightarrow dt$, the arc is infinitesimal, $dl = vdt$, and then $rdl/2$ is the *area*.

As a planet orbits the Sun, the area it sweeps is proportional to time. Mean anomaly is not given by an angle. Mean anomaly is given by an area.

3.6 Kepler Equation

The mean anomaly is proportional to the area. The question then is, how to compute the area? Kepler did not know calculus, so he could not calculate the area by summing the infinite distances. But after he discovered that the orbit was an ellipse he used the geometry of the ellipse to find out the area.

Consider Fig. 3.24. The main insight is that $\overline{PH}/\overline{QH} = b/a$, which is seen because $\overline{CQ} = a$, and thus $\overline{QH} = a \sin \beta$, and we have already proven that $\overline{PH} = a \sqrt{1 - e^2} \sin \beta = b \sin \beta$.

The area A_{Circle} of the circle is 4 times the area of the quadrant. The area of the quadrant can be found by integrating the vertical distances y_{circ} from $x = 0$ to $x = a$.

$$A_{\text{Circle}} = 4 \int_0^a y_{\text{circ}} dx \quad (3.38)$$

Given $y_{\text{circ}} = a \sin E$ and $x = a \cos E$, then $A_{\text{Circle}} = \pi a^2$, as expected. The area of the ellipse is

$$A_{\text{Ellipse}} = 4 \int_0^a y_{\text{ell}} dx \quad (3.39)$$

Because we can write $y_{\text{ell}} = b/a y_{\text{circ}}$, then

$$A_{\text{Ellipse}} = 4 \int_0^a \frac{b}{a} y_{\text{circ}} dx \quad (3.40)$$

$$= \frac{b}{a} A_{\text{Circle}} = \pi ab \quad (3.41)$$

Given that the planet sweeps equal areas at equal times, a relation between mean anomaly (time) and area can be established. A planet sweeping equal areas at equal times will, within a time t , sweep an area

$$A_{\text{sector}} = \pi ab \frac{t}{T} \quad (3.42)$$

where T is the orbital period. The question is, what is the area of the sector? Consider Fig. 3.25. The area swept from perihelion (X) to point P is the area of the sector A_{SPX} . We can relate it to the area of the sector A_{SQX} by again using the same relation $y_{\text{ell}} = b/a y_{\text{circ}}$

$$A_{\text{SPX}} = \frac{b}{a} A_{\text{SQX}} \quad (3.43)$$

The sector SQX can be broken down as the circular sector CQX, minus the triangle ΔCSQ

$$A_{\text{SQX}} = A_{\text{CQX}} - A_{\text{CSQ}} \quad (3.44)$$

The circular sector CQX comprises an angle E of the full 2π circle, so its area is

$$A_{\text{CQX}} = \pi a^2 \frac{E}{2\pi} \quad (3.45)$$

As for the triangle ΔCSQ , its base is $\overline{\text{CS}} = ae$, and height $\overline{\text{QH}} = a \sin E$

$$A_{\text{CSQ}} = \frac{(ae)(a \sin E)}{2} \quad (3.46)$$

We thus find the area of the elliptic sector SPX,

$$A_{\text{SPX}} = \frac{ab}{2} (E - e \sin E) \quad (3.47)$$

But because area equal time, $A_{\text{SPX}} = \pi ab t/T$. Equating both,

$$\pi abt/T = \frac{ab}{2}(E - e \sin E) \quad (3.48)$$

$$2\pi t/T = (E - e \sin E) \quad (3.49)$$

Given $\Omega = 2\pi/T$, the left hand side is $M = \Omega t$, the mean anomaly. Thus,

$$\boxed{M = E - e \sin E} \quad (3.50)$$

This result is known as **Kepler's equation**. It is a direct consequence of Kepler's 2nd law, and can also be seen as Kepler's 2nd law itself. The left hand side is time. The right hand side is area.

Together with the 1st law, this equation gives the position of the planet in the orbit knowing the time. The radius vector in terms of the eccentric anomaly is

$$r = a(1 - e \cos E) \quad (3.51)$$

or in terms of the true anomaly,

$$r = \frac{a(1 - e^2)}{1 + e \cos f}. \quad (3.52)$$

where the true anomaly is

$$\cos f = \frac{\cos E - e}{1 - e \cos E}. \quad (3.53)$$

3.7 Equation of the Center

By Kepler equation, it is immediate to obtain M if E is known. In practice, it is usually the opposite problem the interesting, obtaining E , or f through M . This is a much more complicated problem, because Kepler's equation is transcendental and therefore does not have a closed-form solution. Series expansions or numerical root-finding is needed to find approximate solutions. In practice, we want to know $f - M$, the difference between the true and mean anomaly. This difference is known as *equation of the center*.

We write first $E - M = e \sin E$. Because $\sin E$ is an odd function, we expand it in Fourier sine series

$$f(e) = E - M = \sum_{s=1}^{\infty} b_s(e) \sin(sM) \quad (3.54)$$

Where the coefficients are

$$b_s(e) = \frac{2}{\pi} \int_0^{\pi} f(e) \sin(sM) dM \quad (3.55)$$

$$= \frac{2}{\pi} \int_0^{\pi} e \sin E \sin(sM) dM \quad (3.56)$$

We use integration by parts to find

$$b_s(e) = -\frac{2}{s\pi} e \sin E \cos sM \Big|_0^{\pi} + \frac{2}{s\pi} \int_0^{\pi} \cos sM d(e \sin E) \quad (3.57)$$

The 1st term integrates to zero. For the second, we write $d(e \sin E) = d(E - M)$, and then

$$b_s(e) = -\frac{2}{s\pi} \int_0^\pi \cos sM dM + \frac{2}{s\pi} \int_0^\pi \cos(sM) dE \quad (3.58)$$

The first integral integrates to zero. For the second one we again make use of Kepler's equation,

$$b_s(e) = \frac{2}{s\pi} \int_0^\pi \cos(sE - se \sin E) dE \quad (3.59)$$

which reveals a Bessel function of 1st kind,

$$J_n(x) = \frac{1}{\pi} \int_0^\pi \cos(n\tau - x \sin \tau) d\tau \quad (3.60)$$

where $x = se$, $n = s$, and $\tau = E$. By comparison, the coefficients are

$$b_s(e) = \frac{2}{s} J_s(se) \quad (3.61)$$

And the series solution to Kepler's equation

$$E = M + e \sin E = M + 2 \sum_{s=1}^{\infty} \frac{1}{s} J_s(se) \sin(sM) \quad (3.62)$$

Let us find the solution to second order. The first Bessel functions are

$$J_1(x) = \frac{x}{2} - \frac{x^3}{16} + \dots \quad (3.63)$$

$$J_2(x) = \frac{x^2}{8} + \dots \quad (3.64)$$

$$(3.65)$$

We thus have

$$E = M + e \sin M + \frac{e^2}{2} \sin 2M + \dots \quad (3.66)$$

3.7.0.1 From eccentric to true anomaly

Knowing the series equation for the eccentric anomaly, we can find the series for the true anomaly. Starting from the equation of the orbit

$$r = \frac{a(1 - e^2)}{1 + e \cos f} \quad (3.67)$$

and taking the derivative

$$\dot{r} = \frac{a(1 - e^2)}{(1 + e \cos f)^2} e \sin f \dot{f} = \frac{re \sin f}{1 + e \cos f} \dot{f} \quad (3.68)$$

At a distance r and true anomaly f , the infinitesimal area is $r^2 df/2$. This area corresponds to an infinitesimal time dt that it takes to traverse it, thus sweeping a fraction dt/T of the total area

$$\frac{r^2 df}{2} = \pi ab \frac{dt}{T} \quad (3.69)$$

Dividing both sides by dt ,

$$r^2 \dot{f} = na^2 \sqrt{1-e^2} \quad (3.70)$$

where $n = 2\pi/T$ is the mean motion. We divide both sides by r^2 , with $r = a(1-e \cos E)$,

$$\dot{f} = \frac{n \sqrt{1-e^2}}{(1-e \cos E)^2} \quad (3.71)$$

and using $dM = ndt$,

$$df = \frac{\sqrt{1-e^2}}{(1-e \cos E)^2} dM \quad (3.72)$$

Considering Kepler's equation

$$dM = dE - e \cos E dE = dE(1 - e \cos E) \quad (3.73)$$

leading to

$$\frac{dE}{dM} = \frac{1}{(1 - e \cos E)} \quad (3.74)$$

Substituting Eq. (3.74) into Eq. (3.75),

$$df = \sqrt{1-e^2} \left(\frac{dE}{dM} \right)^2 dM \quad (3.75)$$

Which can now be integrated using the series solution. To first order in e ,

$$E \approx M + e \sin M \quad (3.76)$$

then

$$\left(\frac{dE}{dM} \right)^2 \approx (1 + e \cos M)^2 \approx 1 + 2e \cos M \quad (3.77)$$

where we dropped the second order term in eccentricity. Plugging Eq. (3.77) into Eq. (3.75)

$$df = (1 + 2e \cos M) dM \quad (3.78)$$

$$\boxed{f - M = 2e \sin M} \quad (3.79)$$

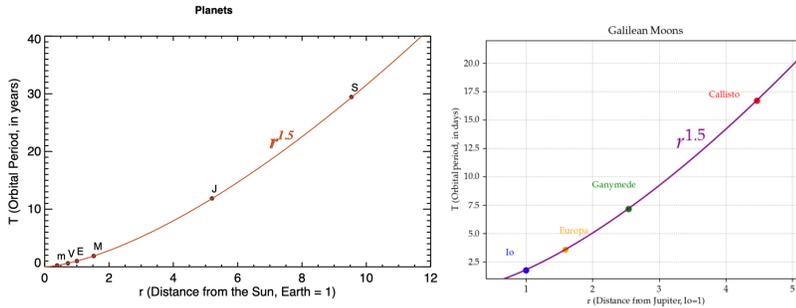


Figure 3.26: Kepler’s 3rd law. A body’s orbital period T is proportional to the power of 1.5 of the orbital distance a . Squaring both sides leads to $T^2 \propto a^3$ as more commonly known. The law is universal, applying to the planets (left) and satellites (right).

3.8 Kepler’s 3rd law

Kepler’s 1st and 2nd law, presented in his book *Astronomia Nova* of 1609, deal with 1-body motion. One planet orbiting a central object. The 3rd one, published ten years later, in 1619, in his book *Harmonices Mundi*, is separate and deals with a system of planets.

Having derived the distances and orbital periods to the planets in the heliocentric model, Kepler noticed that they obeyed a regular relation (Fig. 3.26). A linear law undershoots the relation, the parabola overshoots it. Something in the middle between 1 and 2 should fit. Kepler tried a power law of index 1.5, and it fit the data perfectly. The period is proportional to the power of 1.5 of the orbital distance

$$T \propto a^{1.5} \tag{3.80}$$

Square both sides and one finds the more well known version of the law: the square of the period is proportional to the cube of the semimajor axis.

$$T^2 \propto a^3 \tag{3.81}$$

Kepler’s three laws of motion are thus summarized

$r = \frac{a(1 - e^2)}{1 + e \cos f}$	(3.82)
$M = E - e \sin E$	(3.83)
$T^2 \propto a^3$	(3.84)

Problems

1. An asteroid is found between Mars and Jupiter.
 - (a) Its synodic period is 1.27 years. What is its orbital period around the Sun, in years?

- (b) If the time between quadrature and opposition is 89 days, what is its distance to the Sun, in AU?
- (c) A near Earth asteroid currently between Venus and Mercury is found, also with a synodic period of 1.27 yr.
- What is its orbital period around the Sun?
 - Its maximum elongation is 42° . What is its distance from the Sun in AU?
2. (a) The distances of Io, Europa, Ganymede, and Callisto to Jupiter are 422 000 km, 671 000 km, 1 070 000 km, and 1 883 000 km, respectively. Consider the Earth and Jupiter orbit the Sun at circular orbits at 1 and 5.2 AU ($1 \text{ AU} = 1.49 \times 10^8 \text{ km}$). Show that the maximum angular distance between Io and Jupiter, seen from Earth, is 2 arcminutes, and that the maximum angular distance between Callisto and Jupiter, seen from Earth, is 10 arcminutes.
- (b) The moons of Mars, Phobos and Deimos, have orbits with apoapses 9517 km and 23 470 km. Mars perihelion is at 1.38 AU. Show that Phobos and Deimos widest angular separations to Mars seen from Earth are about half and 2 arcminutes, respectively.
- (c) Phobos and Deimos were discovered only in 1877, more than 250 years after Galileo's sighting of the moons of Jupiter. Yet, seen from Earth, the angular distance between Io and Jupiter and between Deimos and Mars is roughly the same. Galileo could easily spot Io, but not Deimos. Can you tell why?
3. The Hubble Space Telescope (HST) is on a circular, low-Earth orbit, at an elevation $h = 600 \text{ km}$ above the Earth's surface.
- What is its orbital period? For comparison, the Moon is at 384 000 km, and orbits the Earth in 27 days.
 - For an observer who sees HST pass through the zenith, how long is HST above the horizon during each orbit?
4. Communications and weather satellites are often placed in geosynchronous orbits. A geosynchronous orbit is an orbit about the Earth with orbital period P exactly equal to one sidereal day.
- What is the semimajor axis of a geosynchronous orbit?
 - What is the orbital velocity of a satellite on a circular geosynchronous orbit?
5. NASA sends most of its spacecraft to Mars using a trajectory called a Hohmann Transfer, which launches the spacecraft into an ellipse (called the Hohmann Ellipse) with perihelion at the orbit of Earth and aphelion at the orbit of Mars, as drawn below.
- Calculate a_{Hohmann} , the semi-major axis of the Hohmann ellipse, and the time it will take to travel from Earth to Mars following this trajectory. Show your work.

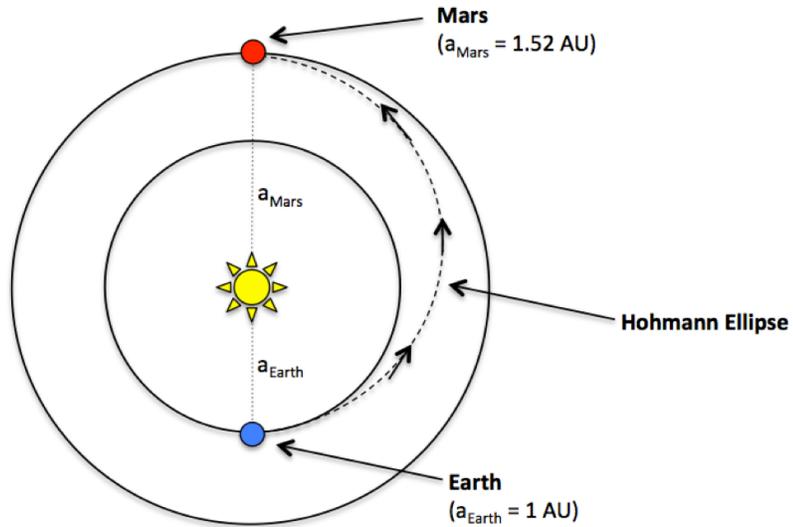


Figure 3.27: .

- (b) The arrow pointing to Mars in the diagram above shows its location when the spacecraft arrives. On the diagram, indicate approximately where Mars will be in its orbit when the spacecraft launches from Earth. Show your work or explain your reasoning.

