# How Kepler discovered his laws

A. Eremenko

December 17, 2016

This question is frequently asked on "History of Science and Math", by people who know little about astronomy and its history, so I decided to write this explanation.

The general idea is called successive approximation. The motion of the planets looks very complicated. But it decomposes into simpler motions. For every given time span and accuracy only one or two of these simple motions are relevant, and the other can be neglected. (This is actually the general idea which makes science possible).

1. How it looks from Earth. The most conspicuous motion in the sky is the diurnal motion: the whole sky rotates uniformly about a polar axis with period one day-and-night. Or Earth rotates. What "really" rotates is not important for our purposes. What is important is that this rotation is uniform, and well understood. So it can be subtracted from all other motions and not considered in what follows. Of course, it is also not exactly uniform, but the deviation from the uniformity is so small, that it can be neglected for the questions discussed here. After this rotation is subtracted, we see in the sky the motion of the Sun with period one year, and the motion of a planet, say Mars which does not look exactly periodic but only approximately.

Remark. Non-uniformity of the Earth rotation was essentially known sin antiquity and Middle age. The main point is that such things can be dealt with *separately*, and then corresponding corrections be introduced.

2. The following two facts are very fortunate for astronomers:

a) Each planet moves in one plane, and this plane contains the Sun.

b) The angles between these planes are small (the angle between the plane of Mars orbit and the plane of Earth orbit is 1.85 degrees).

This implies that one can project the Mars orbit onto the plane of the Earth orbit and consider all motions in one plane. The distortion coefficient coming from this projection is roughly speakings $\cos 1.85^o \approx 0.9995$, which is very close to 1. So after the theory is developed for the motion in one plane, a correction can be applied for the inclination of the orbit. All this was well-known to Ptolemy.

The plane which contains the Sun and the Earth orbit (or the Earth and the Sun orbit, before Copernicus) is called the ecliptic. Astronomers use polar coordinates in this plane: distance from the center and longitude.

At the time of Kepler, direct measuring of distances (in miles of feet) was not possible. So the unit of distance had to be related to the dimension of the Earth orbit (it is called the astronomical unit).

3. Observed motion of the Sun. Sun's motion is periodic. It makes one revolution per year (this is actually the definition of the year, technically it is known as the "tropical year"). But the motion is not uniform. However it is well approximated by the sum of two uniform motions on slightly eccentric circles. This was known to Ptolemy, and parameters of these motions were well known. Kepler was using Copernican system in which Sun does not move, and he is talking of the orbit of the Earth. This is just a change in coordinate system, and it has no advantages for the Earth-Sun motion. But it greatly simplifies the motion of the planets.

4. In heliocentric system, Mars rotates around the Sun periodically, with period about 1.88 years. The orbit is not exactly a circle (but close to it) and the motion is not uniform (but close to it). This is already a great simplification in comparison with the visible motion of Mars around the Earth which is complicated: it changes direction, describing loops etc. But in the first approximation this complicated motion can be decomposed into two uniform circular motions, as Ptolemy knew. Passing to the heliocentric coordinates (Copernicus) gives a further simplification. So the problem which Kepler faced was to describe mathematically the small deviation of the Mars heliocentric orbit from the circle, and the small deviation of the motion from the uniform one.

5. What data were available. The only things which could be measured in Kepler's times were angles and time. So all available data were in the form of the values of the angle Sun-Earth-Mars at certain moments of time.

To give the idea of accuracy, the angles could be measured to $1'$, roughly speaking, sometimes to a fraction of $1'$. It is more complicated with time. The most accurate clocks available before the invention of pendulum were water clocks, with accuracy of few minutes per day at the best. But the accuracy of time measurement can be greatly improved when you measure over long periods.

An example. How could the ancients measure the length of the year within minutes? Suppose that you can measure the time of an equinox with accuracy 1 hour (which is possible with quite crude tools). If your observation period is 1 year you have its length with accuracy about 1 hour. But if you can use observations with span 100 years, you know the interval of 100 years with the accuracy 1 hour, that is the length of 1 year with accuracy 1/100 hour.

This demonstrates the principle that using observations over long periods increases accuracy. Kepler did not only use Brahe's observations. He used all available record since antiquity.

6. Now the general principle must be clear. You start with the best available description of the Earth motion and of the Mars motion around the Sun. And try to fit the available data, that is to compute the angles Sun-Earth-Mars for the times of observations, and compare them with measured angles. You find a disagreement. The actual disagreement that Kepler found was of the order of minutes, not degrees. Ptolemy-Copernicus description fits the real motion quite well.

Then you make a hypothesis, how should the orbit and the law of motion be modified to better fit the data. Then you compute again and compare again.

The task is simplified somewhat by the fact that the periods of Earth and Mars are "not commensurable": their ratio $\approx 1.88$ is not a ratio of small integers. This implies that for a fixed position of Earth on its orbit, one has many different positions of Mars, and vice versa.

7. Plato formulated the goal of astronomy as follows: *to represent the visible motion of celestial bodies by a combination of uniform rotations on circles.* This is actually a very reasonable approach since we know that *every* bounded motion can be approximated by linear combinations of motions on circles with constant speeds. This approximation is called the Fourier series, and this is how the motion of the planets is represented even now, for the purposes of prediction and computation. The modern formula would have hundreds

of terms, such result would hardly satisfy Plato.

Already Ptolemy knew that permitting non-uniform circular motion can dramatically simplify the model. He introduced the "equant", a motion on a circle such that the polar angle changes with constant speed *as seen from another point, different from the center*. This auxiliary point is called the equant.

Ptolemy theory described the position of Mars within $8'$. This disagreement was Kepler's concern.

Remark. It is interesting that Ptolemy's theory of non-uniform circular motion (equant) was much criticized by the Arab astronomers and by Copernicus on philosophical grounds. They all accepted the Plato's dogma about *uniform* circular motions. So Copernicus replaced the equant by some more complicated arrangement, and as a result his theory was slightly less accurate than Ptolemy's.

8. Kepler first tried to make more precise the model for the Earth motion (using Ticho's observations of Mars!) With Ptolemy's mathematical tools, excenter and equant, trying to fit parameters with observations as well as possible, he discovered the Law of Areas (now called the Second Kepler's Law). In this intermediate model the planets moved on circles but instead of having constant speed, they obeyed the Law of Areas.

9. Still the model did not fit perfectly the observations of Mars. After *very many* trials Kepler came to the conclusion that the orbit *cannot be a circle*. Then he tried an oval...

And finally he had the great ideas crossed his mind that the best fitting orbit is an ellipse with Sun on one of its foci. This was a truly great discovery, first because he obtained a perfect fit (to the accuracy available from the observations), and second because this was a rejection of a 2000 years old dogma that only circles must be used. Everyone before Kepler used only circles.

It is interesting that Kepler, unlike many other scientists described all his steps, unsuccessful attempts and mistakes in great detail. This makes his work Astronomia Nova a very interesting, but difficult reading.

10. The Third Law was discovered much later, published in his book Harmonia Mundi. Since his youth, Kepler was trying hard to establish some pattern in the periods and distances of planets. Finally he established the simple pattern, just by playing with numbers.

Together the three laws give a complete description of the motion: the first law tells the shape of the orbit, the second one gives velocity as a function of time, up to the constant multiple, the period. And the third law gives the period. So once the shape and size of the ellipse, and position on the ellipse at some moment are known this completely determines the motion.

These laws would be exact if no other bodies were present, only the Sun and one planet. Fortunately (for astronomers and for all of us) the mutual influence between the planets, which in principle could lead to "chaos" is so small, that it can be neglected for the times like hundreds of millions of years. In the case of chaos, the Earth orbit would be non-periodic, the conditions on Earth would change too much and too quickly to support any life.

# References

[1] J. Kepler, Astronomia Nova, English translation: New Astronomy, translated by William H. Donahue, Cambridge, Cambridge Univ. Pr., 1992

[2] N. I Idelson, Essays on the history of planetary theories, in the book: N. I. Idelson, Essays on the history of celestial mechanics, Moscow, Nauka, 1975 (Russian)