Where on Earth am I? Don't Worry, GPS Satellites will Guide you

1. Introduction and Principle of GPS

Makarand Phatak

Since ancient times navigators have been taking the help of celestial objects to find angles between the horizontal and the lines of sight to the celestial objects, in order to determine their position on the earth. Now, they would take the help of couple of man-made objects whizzing around the earth to find their position with very high accuracy, by measuring ranges from these objects. These objects are satellites and this article gives a brief introduction to a satellite based navigation system, called the GPS (Global Positioning System) in two parts. The first part introduces the basic principle of operation of GPS.

Makarand Phatak is working as a researcher at Aerospace Systems Private Limited, Bangalore. He received his PhD from the Indian Institute of Science in 1992. Impressed by the statement "The best way to learn is to teach and the best way to teach is to learn" he seeks joy in teaching and learning within his capacity.

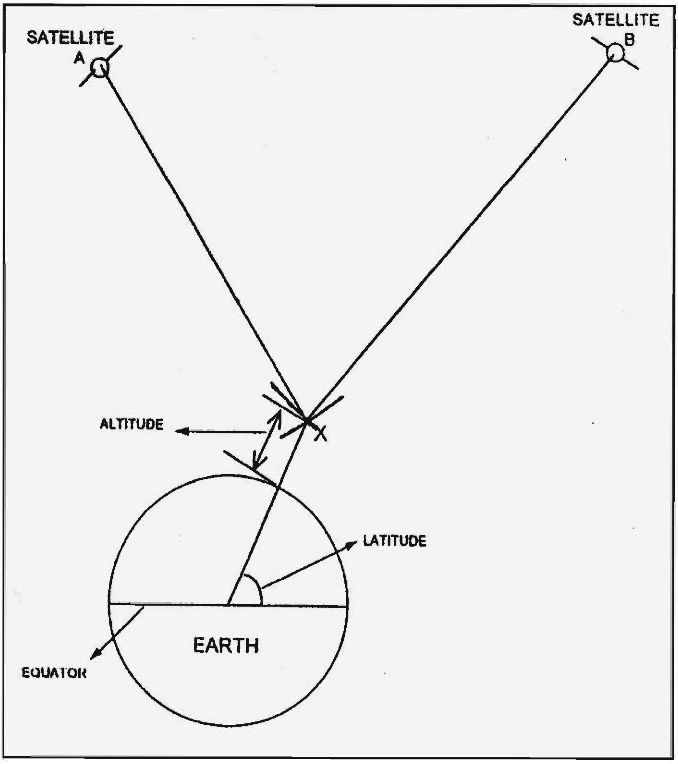
Introduction

Spanish poet Antonio Machado has written,"Traveller, there is no path, paths are made by walking". Yet a traveller, while exploring an unknown path will frequently need an answer to the question: Where on earth am I? Any vehicle, manned or unmanned, moving through air or sea or on desert land where there are no reference marks, frequently needs information on its changing position, that is, latitude, longitude and altitude. Apart from moving vehicles, positions of points which are stationary on the earth are also of interest for survey purposes. In olden times position determination was done with the help of celestial objects. For example, Galileo had proposed to measure longitude by using two time-keepers, one in the sky in the form of regular eclipses of moons of Jupiter, and the other on the earth in the form of a swinging pendulum. In modern times also time-keeping is used along with radio waves, and different radio systems have been created for position determination. For example, ground based Omega and Loran, aircraft mounted with Doppler radar, and space based Transit, GPS and GLONASS. Out of these GPS is most popular and successful. Created by Department of Defense (DoD), USA, this ambitious system uses dedicated satellites to transmit radio signals. Using GPS one can obtain position information anytime, anywhere on the earth; and this information is quite accurate. And it can also provide other useful information such as velocity and the precise time. Russia has also created the satellite based system, GLONASS (Global Navigation Satellite System), which has recently become operational.

Principle of GPS

For the sake of convenience consider the space to be two and not three dimensional. So, ignoring longitude, consider the earth not to be spherical in a three dimensional space but to be a circle in a two dimensional space. Imagine that at the positions of points A and B in space, there are two satellites (see *Figure 1*). The satellites are also confined to move in the two dimensional space. However motion of the satellites need not concern us for the time being. Each satellite has a perfect clock which shows not only the time but also its position, i.e its latitude and altitude. Near the earth, at the point X, there is an observer who is equipped with a perfect clock and two telescopes. This observer is interested in knowing his position and for that he has oriented the two telescopes towards the clocks of the satellites at A and B. It is assumed that all the clocks are

Figure 1. Illustration of positioning.



synchronized, that is at any instant all the clocks show the same time (relativistic effects are ignored, they are considered later). To specify time ignore the number of seconds, minutes, hours of a day, and concentrate only on milliseconds[1]. Suppose that the observer's clock is indicating a time of 100 milliseconds. At the same instant he observes that the clocks of satellites at A and B are showing 35 and 30 milliseconds respectively. This means that when the observer is looking at the clock of the satellite at A (or B), he is really seeing the time shown by the clock 100 – 35 = 65 (or 100 – 30 = 70) milliseconds in the past. This is so because it takes 65 (or 70) milliseconds for the light which has started from satellite at A (or B) to reach X. The observer multiplies the difference between his clock reading and clock reading of satellite at A (or B) by the speed of light (300 Km per millisecond) to find the range AX (or BX) as 19500 Km (or 21000 Km). Through his telescopic observation he also gets the positions of the satellites at A and B. After getting this information he locates points A and B on a graph paper, draws circles of radii equal to ranges AX and BX with A and B as centers respectively and finds the point X as a point of intersection of these two circles such that the altitude of X is less than the altitudes of A and B. The other point of intersection of the two circles has an altitude greater than the altitudes of A and B and this solution is rejected. Once X is obtained, he can then easily get the latitude and altitude of X (see *Figure 1*).

1 millisecond is 1000th part of a second and is the main unit of time used in this article.

Suppose, the clocks of the satellites are accurate atomic clocks but the observer has an inexpensive clock which is not accurate. Further, suppose that his clock is showing 110 milli-seconds instead of the true time 100 milliseconds. Since the clocks of satellites are accurate, when the observer looks at them from his telescopes, he sees that they show the times of 35 milli-seconds and 30 milliseconds respectively as before. Then the observer will calculate the ranges AX and BX respectively to be 22500 and 24000 Kms. These are obviously wrong by 3000 Kms and hence the determination of the position of X will be totally wrong.

There is a simple way to correct this error. Suppose there is one more satellite with an accurate atomic clock at position, say C, and the clock of the satellite at C is showing the time of 40 milliseconds to the observer when he looks using a third telescope. Then the calculated range CX will be (110 – 40) × 300 = 21000 Kms. Now, he draws three circles with centers A, B and C as explained before whose radii are respectively the calculated ranges AX, BX and CX He will find that they do not intersect at one point at all. Then he should go on decreasing or increasing the three radii by the same magnitude till the three circles intersect at one point (in the example considered all the radii would be reduced by a magnitude equivalent to 3000 Kms). When such a point is obtained, that is nothing but the point X. From the scale used on the graph paper he can determine the latitude and the altitude. Also, the magnitude of adjustment required, when divided by the speed of light, gives the offset of the clock of the observer from the reference time. Thus if there is one extra satellite with an accurate atomic clock there is no need for the observer to have an expensive atomic clock. With the observer having a clock with offset, the measurement is affected by the offset and it is called pseudo-range measurement instead of range measurement. Sometimes, the term range will still be used in this article in place of pseudo-range for convenience.

The method of adjusting the radii to find X as described above requires trial and error. A systematic geometric construction with the help of algebra can be used to locate X. See *Box 1* for details.

In practice an observer will not use telescopes to find ranges and will not use graph paper to find the position. Instead he uses a GPS receiver to find the position. A GPS receiver is of the size of a small tape recorder. A patch antenna that can fit into a palm is connected to this receiver. The antenna, instead of receiving light as in the case of a telescope, receives radio

Box 1. Apollonius, Newton and GPS

One of the problems posed by Apollonius of Perga, a Greek mathematician, is about drawing a circle touching two circles and passing through a point. The problem of two dimensional GPS is nothing but the problem of Apollonius which appears to have been solved in 3rd century B.C! (*Figure a*) Unfortunately the original Apollonius construction is lost, but other constructions are available. Notable among them is the hyperbolic one due to Newton who considered the problem of the determination of a fourth point when its two range differences from three points are given.

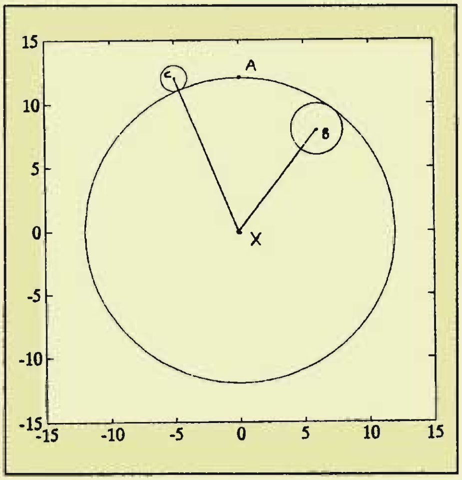
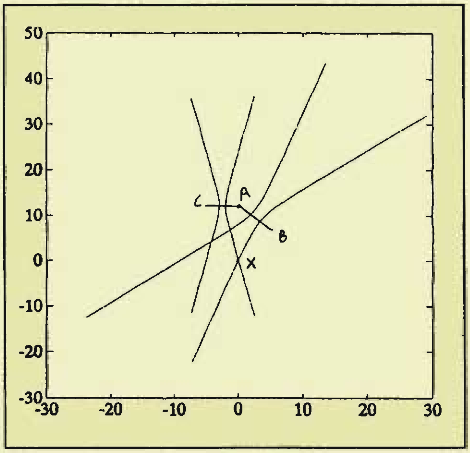


Figure a. 2D-GPS problem as a problem of Apollonius.

Consider the 2-D GPS problem of locating X, given locations of A, B, C and the pseudo-ranges AX+b, BX+b and CX+b, where AX, BX and CX are ranges of X from A, B and C respectively and b is a common unknown offset. Define range difference of X from A and B as AXB = AX – BX and similarly of X from A and C as AXC = AX – CX. Note that by subtracting measurements of pseudo-ranges we get range differences AXB and AXC since the common offset b cancels out ((AX + b) – (BX + b) = AX – BX = AXB and (AX + b) – (CX + b) = AX – CX = AXC). Thus the problem reduces to: Given locations A, B, C and the range differences AXB and AXC, locate X such that AX–BX = AXB and AX – CX = AXC. It can be modeled as a problem of Apollonius as follows. Suppose we draw two circles, one with center B and radius AXB and the other with center C and radius AXC. Then draw a circle such that it is tangential to these two circles and passes through point A. Then the center of the circle is point X. See *Figure b*.

Figure b. Newton's construction for the problem of Apollonius.



Now we consider the solution to the related problem as given by Newton. Geometrically, locus of all points X such that AX – BX = AXB is a hyperbola with foci at A and B. Similarly, locus of all points X such that AX – CX = AXC is another hyperbola with foci at A and C. Then, point X lies on the intersection of the two hyperbolae.

Algebraically, the two hyperbolae are given by the equations

$$b_1^2 = r(a_1 + c_1 \cos \theta)$$

Continued

$b^2_2 = r(a_2 + c_2 \cos(\alpha - \theta))$

where $2a_1 = \rho_2 - \rho_1$ and $2a_2 = \rho_3 - \rho_1$, ρ_1, ρ_2, ρ_3 are the measured pseudo-ranges AX + b, BX + b and CX + b respectively, $2c_1$ and $2c_2$ are the ranges AB and AC respectively, r and θ are the polar coordinates of X with AB as the reference axis and A as the origin, α is the angle BAC and $b_1^2 = c_1^2 - a_1^2$, $b_2^2 = c_2^2 - a_2^2$. See *Figure c*. Eliminate r from the above two equations to obtain the following quadratic equation in $\cos\theta$.

$(J^2 + M^2)\cos^2\theta + 2JK\cos\theta + K^2 - M^2 = 0$, where

$J = b_2^2 c_1 - b_1^2 c_2 \cos\alpha$, $K = b_2^2 a_1 - b_1^2 a_2$, $M = b_1^2 c_2 \sin\alpha$.

Solving the quadratic gives two solutions. Solution corresponding to X near the earth is retained, the other one is discarded. *Figure b* shows three of the four solutions, the fourth one is beyond the article. Note that in the above algebra two solutions are eliminated because the equations of hyperbolae are represented in terms of r and not r^2.

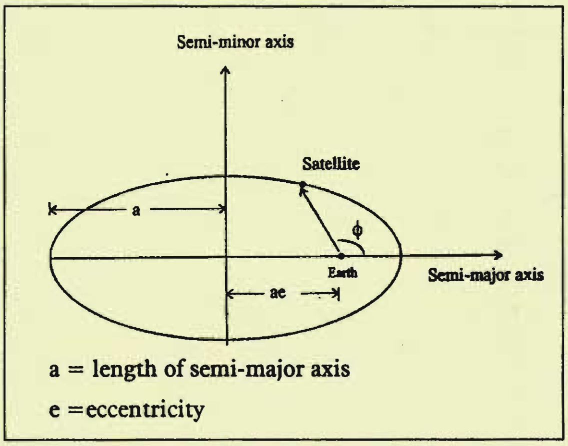
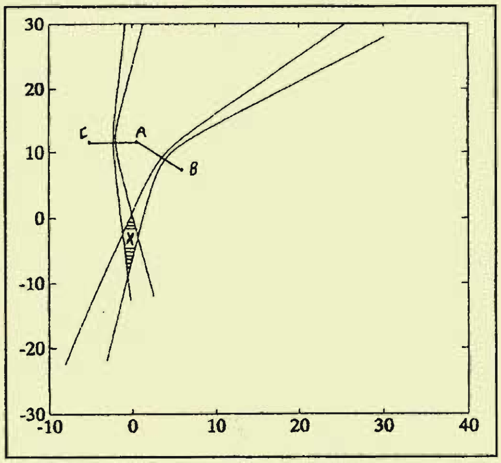


Figure c. Elliptical orbit of a satellite.

Now consider a situation in which each of the measured pseudo-ranges has some uncertainty due to errors. Each of the pseudo-range is expected to be within an interval. Then instead of two hyperbolae, we have four and they intersect to give a diamond of errors, see *Figure d*. The size of the diamond depends on the positions of the satellites and the user and can be described by a parameter called as GDOP (Geometric Dilution of Precision). Precise definition of GDOP is beyond the scope of this article; it is enough to note that more the GDOP more the error in position.

Figure d. Diamond of errors.



The solution presented here offers some geometrical insight. There are several other methods of locating X; some are computationally attractive, some are optimal in the sense that they minimize the effect of random errors in locating the position.

Box 2. Kepler and GPS

Tycho Brahe gave Johannes Kepler the problem of the motion of Mars. Kepler prophesied that he would solve the problem in eight days. Kepler was still on the problem eight years later! But, finally solved the problem and discovered in the process the famous three laws of planetary motion.These laws are used to determine the positions of GPS satellites.

From Kepler's first law we get that the orbit of a satellite is an ellipse with the center of the earth at one of its foci. (See *Figure c*). The equation of the ellipse is $r = a(1-e^2)/1 + e\cos\phi$.

From Kepler's second law we get that the radius vector (vector from the center of the earth to the satellite) sweeps out equal areas in equal intervals of time. From the equation of the ellipse then the following differential equation is obtained.

$$\frac{d\phi}{dt} = \frac{h(1+e\cos\phi)^2}{a^2(1-e^2)^2}, \; h = \text{constant angular momentum} .$$

The constant h can be determined from Kepler's third law which states that the square of the period of revolution is proportional to the cube of the semi-major axis.

$T^2 = 4\pi^2 a^3/\mu$ where μ is Kepler's constant (From his universal theory of gravitation Newton showed that μ is the product of the mass of the earth and the universal gravitational constant G). Finally it can be shown that $h^2 = a(1-e^2)\mu$.

With h known, the differential equation in ϕ gives ϕ dependence on time t starting from the initial condition $\phi(t_0) = \phi_0$. In general the time variation of ϕ is complicated.

signals from GPS satellites. The radio signals contain special codes, called Gold codes, and messages, called navigation messages. Codes contain information about finer divisions of satellite clocks and messages mainly contain information about orbits of satellites which is used to find satellite positions and coarser divisions of satellite clocks. The receiver not only maintains its own time with the help of an inexpensive crystal oscillator clock but also performs three calculations pertaining to 1) finding the position and the clock offset of the receiver. 2) calculating the satellite positions and 3) calculating the ranges of GPS satellites. One method for the first calculation is explained in *Box 1*. For the second calculation Kepler's laws are used to determine satellite

positions at the time of interest from the parameters of elliptical orbit. See *Box 2* for details. It was stated in the beginning that the relativistic effects which affect clocks have been ignored in these calculations. What are these relativistic effects? What does one gain from a knowledge of position on the earth? These questions have been addressed to in the second part of this article.

Suggested Reading

[1] **R J Milliken and C J Zoller. Principle of operation of NAVSTAR and system characteristics.** ***Navigation, Journal of the Institute of Navigation.*** **25. 2, 1978.**

[2] **B W Parkinson, T Stansell, R Beard and K Gromov. A history of satellite navigation.** ***Navigation, Journal of the Institute of Navigation.*** **42. 1, 1995.**

[3] **B W Parkinson and J J Spilker (ed.).** ***Global Positioning System: Theory and Applications.*** **published by American Institute of Aeronautics and Astronautics. Inc. Vol I and II, 1996.**

[4] **E D Kaplan.** ***Understanding GPS: Principles and Applications.*** **Artech House Publishers, Boston. USA, 1996.**

[5] **J Hoshen. The GPS equations and the problem of Apollonius.** ***IEEE Transactions on Aerospace and Electronic Systems.*** **32. 3. July, 1996.**

Address for Correspondence
Makarand Phatak
Aerospace Systems Private Ltd.
Pragathi, 70/1, Miller Road
Bangalore 560 052, India.

Did You Know?

1. There is no exact formula for the perimeter of an ellipse in terms of ordinary functions. This led to the invention of *elliptic functions*, in terms of which this and other previously intractable problems could be solved.

An approximate formula given by Ramanujan in 1914 is $\pi[3(a+b)-\sqrt{(a+3b)(3a+b)}]$.

2. *To draw a hyperbola mechanically.* Fix the ends of two strings at the proposed positions of the foci. Tie the other ends together, arranging the lengths so that their difference is equal to the desired length of the transverse axis. Thread both strings through a small ring and place a pencil in the ring. If the ring and pencil are now moved so that the strings are kept taut the pencil will describe one branch of a hyperbola.

'A Book of Curves' by E H Lockwood
Cambridge University Press, 1961.