Modern Digital Communications:
A Hands-On Approach

GPS: Satellite Position In Various Coordinate Systems

Dr. Nicolae Chiurtu

- Course material of Prof. Bixio Rimoldi -

Last revision: Nov. 14, 2021

Goal

The goal of this part is to learn how to describe the position of a satellite with respect to an earth-centered earth-fixed reference system. This is a system that rotates with the earth.

We start by reviewing Kepler's laws of planetary motion and a few useful facts about ellipses.

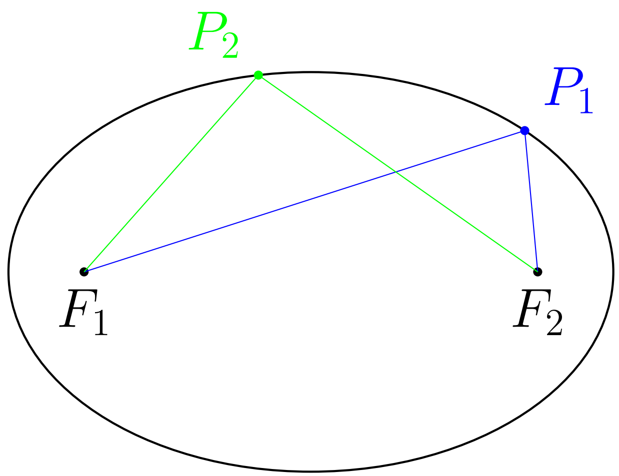
Background Review 1: Kepler's Laws

Kepler's Laws of planetary motion (written for planets of the sun but of course they apply to any planetary motion):

1. (Law of Orbits) All planets move in elliptical orbits, with the sun at one focus.
2. (Law of Areas) A line that connects a planet to the sun sweeps out equal areas in equal times.
3. (Law of Periods) The square of the period of any planet is proportional to the cube of the semimajor axis of its orbit. (The proportionality constant does not depend on the semiminor axis.)

Background Review 2: Useful Facts about Ellipses

An ellipse is the set of points in the plane the sum of whose distances from two fixed points F_1 and F_2 is a constant (see figure). The two fixed points are called the foci (plural of focus).

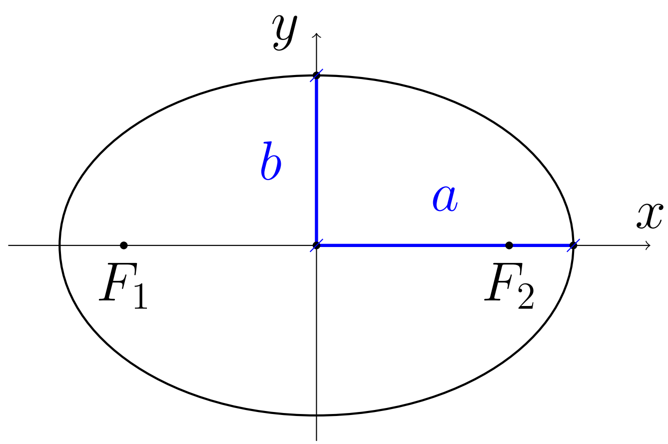


To get the simplest equation of an ellipse we place the foci on the x-axis at equal distance from the origin. Then an ellipse is the set of points (x_e, y_e)

that fulfill

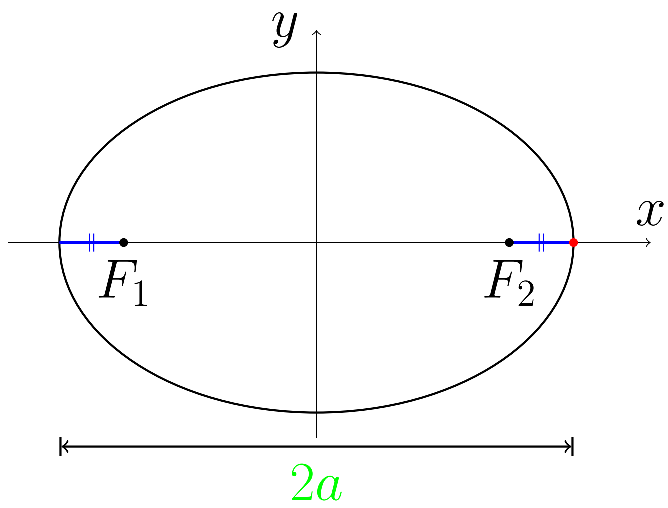
$$\frac{x_e^2}{a^2} + \frac{y_e^2}{b^2} = 1$$

where a and b are the semimajor and semiminor axes, respectively (see figure).



We leave it up to you to verify that the sum of the distances from the foci to a point (x_e, y_e) that fulfills the above equation is indeed constant.

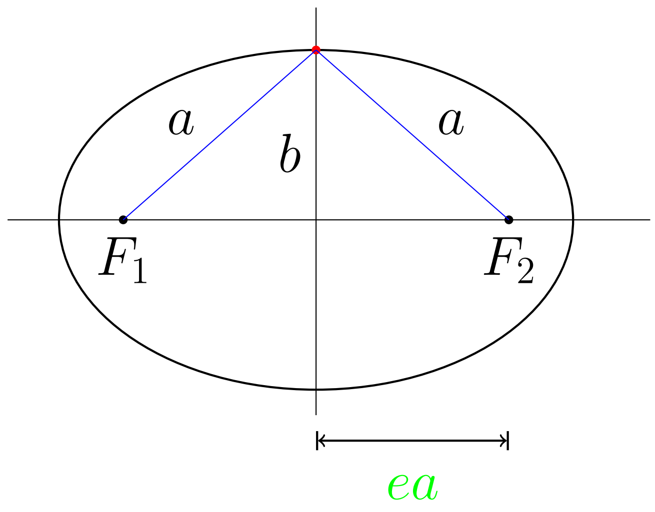
By considering the point with coordinates $(a, 0)$, see next figure, and using the fact that the two blue segments have identical length, we see that the constant equals $2a$.



By considering the point on the ellipse with coordinates $(0, b)$ (see figure), using Pythagoras' theorem we see that the distance from the origin to the foci is ea where

$$e = \sqrt{1 - (b/a)^2}$$

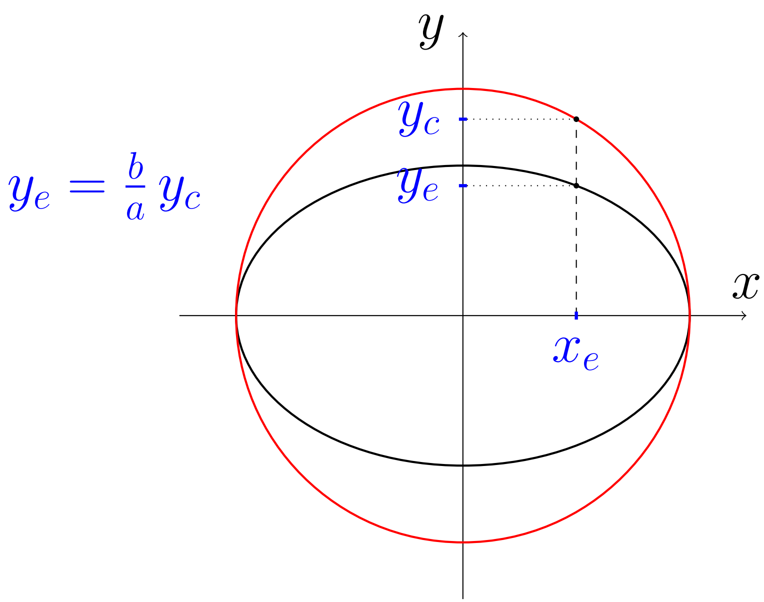
is called the *eccentricity*. Clearly $e \in [0, 1]$ and $e = 0$ when the ellipse is a circle.



The ellipse equation $\frac{x_e^2}{a^2} + \frac{y_e^2}{b^2} = 1$ may be obtained from the equation of the circle $\frac{x_c^2}{a^2} + \frac{y_c^2}{a^2} = 1$ and the transformation

$$x_e = x_c \text{ and } y_e = y_c \frac{b}{a}.$$

This means that the ellipse with semimajor axis a and semiminor axis b may be obtained by "squashing" the circle of radius a along the ordinate so as to reduce its vertical dimension by the factor $\frac{b}{a}$. (See the figure below.)



An immediate consequence is that the area of the ellipse is $A = ab\pi$, i.e., $\frac{b}{a}$

times the area of a circle of radius a. (If this is not clear to you, slice up the circle and the ellipse into narrow vertical stripes of equal width: the stripe inside the ellipse is shorter by a factor $\frac{b}{a}$ compared to the corresponding stripe inside the circle.)

END OF THE BACKGROUND REVIEW

The Need For Several Coordinate Systems

Even for a single satellite it is convenient to introduce several coordinate systems. It is rather evident that we need at least two:

- One that rotates with the earth. It is called the Earth Centered Earth Fixed (ECEF) coordinate system. A stationary object on the surface of the earth has fixed ECEF coordinates.
- One that makes it easy to describe the satellite's position. This is a coordinate system with origin at the earth's center, fixed orientation in space, and x and y axes spanning the orbital plane.

We need to know how to go from one system to the other. This is best done via a sequence of intermediate coordinate systems so that each system is related to the next by a rotation around one of the three main axes.

Convention

All our coordinate systems are related to some rotating object (the earth around the sun, the earth around itself, the satellite around the earth).

We use the following convention: x and y will always span the plane that contains the orbit of the rotating object (or of the equator in case of the earth rotating around itself), with y obtained from rotating x by 90° in the direction of rotation.

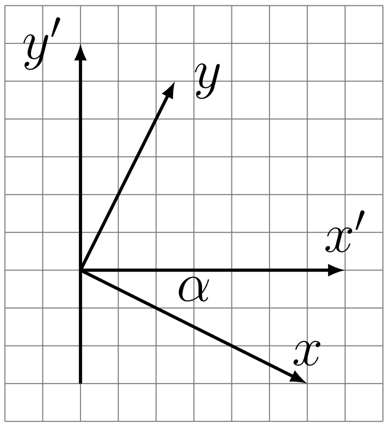
To complete the system into a 3-dimensional coordinate system we pick z according to the *right-hand rule*. (This means that z corresponds to the direction of the cross product $e_x \times e_y$, where e_x and e_y are (unit) vectors pointing in the x and y direction, respectively). Following our convention, z is then lined up with the axis of rotation of the reference object.

To know if an angle is positive or negative we use the right-hand rule: when the thumb points in the direction of the positive z-axis then the other fingers point in the direction of positive angles.

Change of Coordinate System: Rotations

We need to know how to relate coordinate systems that are rotated with respect to one another.

Consider a coordinate system x', y', z' (z' pointing towards you, not shown) obtained by rotating x, y, z by α around z, as shown in the following figure.



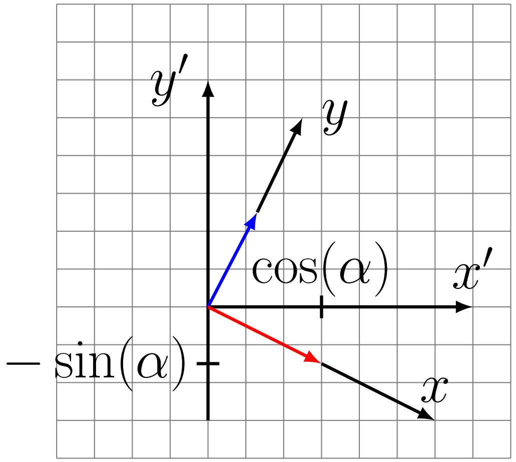
Consistently with our convention, the angle α by which we have to rotate the x, y system to obtain the x', y' system is a positive one.

Let $R_3(\alpha)$ be the matrix defined by

$$p' = R_3(\alpha)p$$

where p and p' are the coordinates of a point with respect to the system x, y, z and x', y', z', respectively.

Here is how we determine $R_3(\alpha)$. Let e_x be the vector with coordinates $(1, 0, 0)^T$ in x, y, z.



The first column of $R_3(\alpha)$ contains the coordinates of e_x in x', y', z'. From the picture we see that they are $(\cos(\alpha), -\sin(\alpha), 0)^T$.

Reasoning similarly with the vector e_y of coordinates $(0,1,0)^T$ in x,y,z we obtain

$$R_3(\alpha) = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Similarly, a rotation by α around x is described by

$$R_1(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & \sin(\alpha) \\ 0 & -\sin(\alpha) & \cos(\alpha) \end{pmatrix},$$

and, a rotation by α around y by

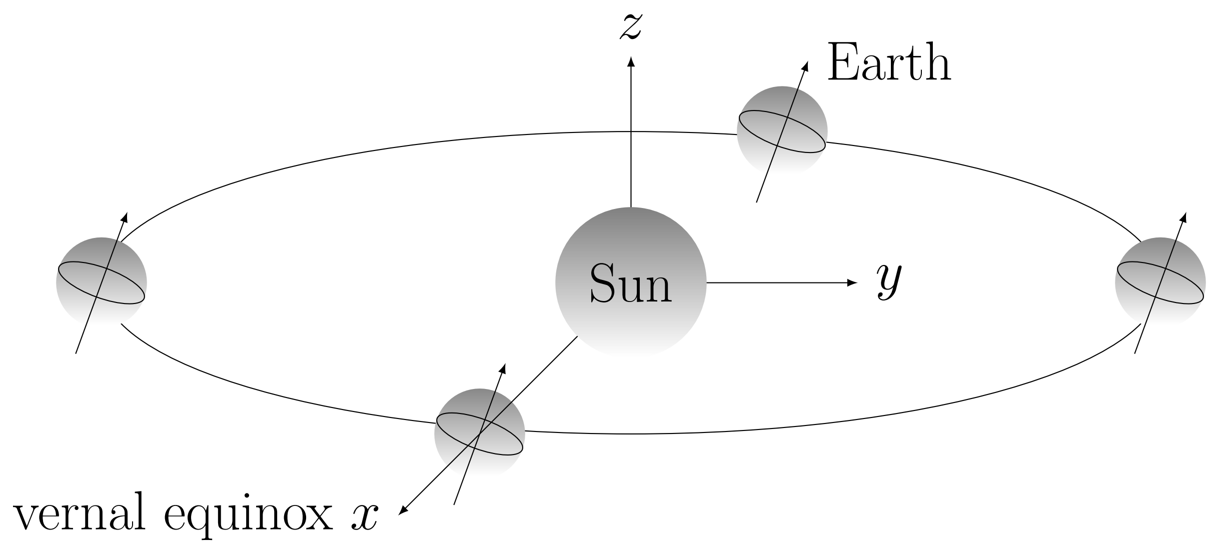
$$R_2(\alpha) = \begin{pmatrix} \cos(\alpha) & 0 & -\sin(\alpha) \\ 0 & 1 & 0 \\ \sin(\alpha) & 0 & \cos(\alpha) \end{pmatrix}.$$

Absolute Reference System

How should we choose our first coordinate system? Is it possible to uniquely describe a coordinate system without another reference system?

Yes it is. A natural choice is to use a Cartesian coordinate system with origin at the center of the sun, with its x, y axes that span the orbital plane of the earth, and x that goes through the vernal equinox (position of the earth during the first day of Fall[1]). We call this the absolute reference system. (See the figure.)

[1] One of two instants each year when the sun is on the equatorial plane.



Notice that on the vernal equinox x goes through the equator and through the earth's center.

It turns out that we will not need the above system. Nevertheless it is comforting to know that there is a natural definition of a "first system" that does not rely on another coordinate system.

The direction of the vernal equinox will play an important role.

Earth Centered Inertial (ECI) System

Next we define the Earth Centered Inertial (ECI) coordinate system and denote it by x_I, y_I, z_I. It has its origin at the center of the earth, the x_I axis points in the direction of the vernal equinox, and the x_I and y_I axis span the equatorial plane. The convention mentioned earlier completely specifies y_I and z_I.

Notice that x_I goes through the equator. Also note that z_I is aligned with the earth's rotation axis.

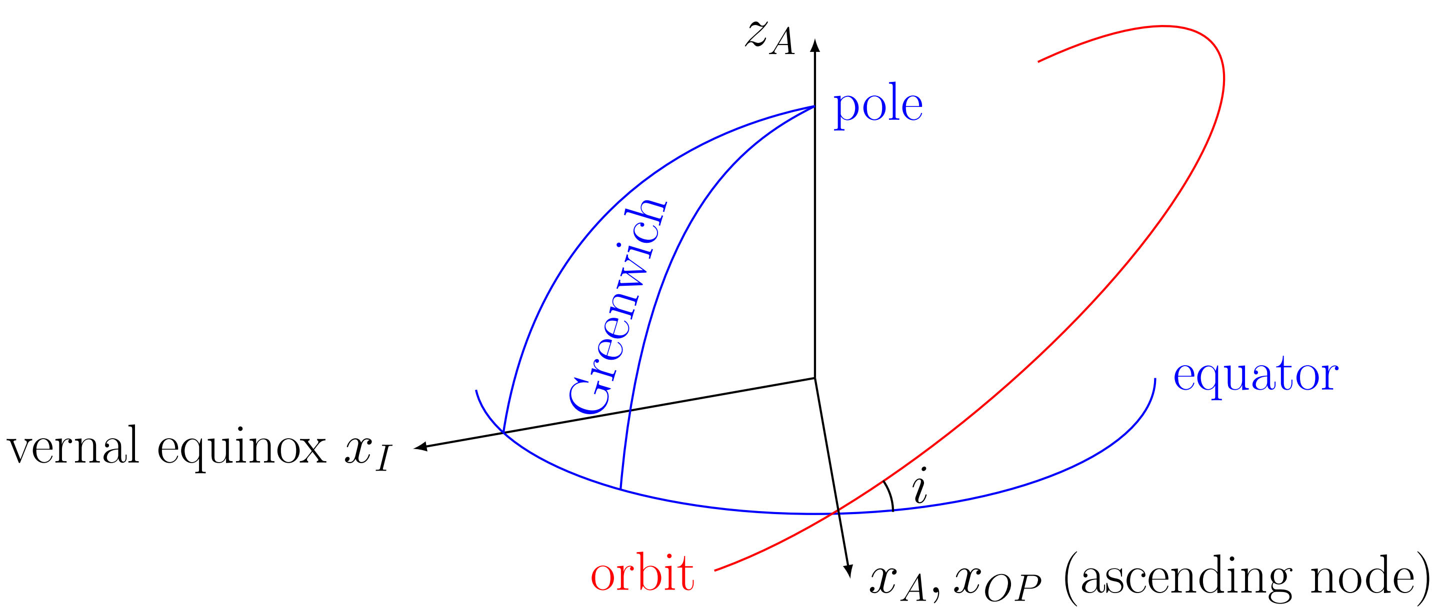
From The ECI To The Satellite's Orbital Plane

The orbit of each satellite is on a plane that goes through the center of the earth. The *orbital plane* is fixed with respect to our ECI.

Since the orbital plane goes through the center of the earth, it crosses the equator at two points. The one that corresponds to the satellite passing from the Southern to the Northern Hemisphere is called the *ascending node.*

Let x_A, y_A, z_A be a coordinate system with origin at the center of the earth, x_A that goes through the ascending node, and x_A, y_A that span the equatorial plane according to the stated convention.

We endow the orbital plane with a coordinate system x_{OP}, y_{OP}, z_{OP} with origin at the center of the earth and x_{OP} that goes through the ascending node. (The direction of y_{OP} and z_{OP} are then determined by our convention.)

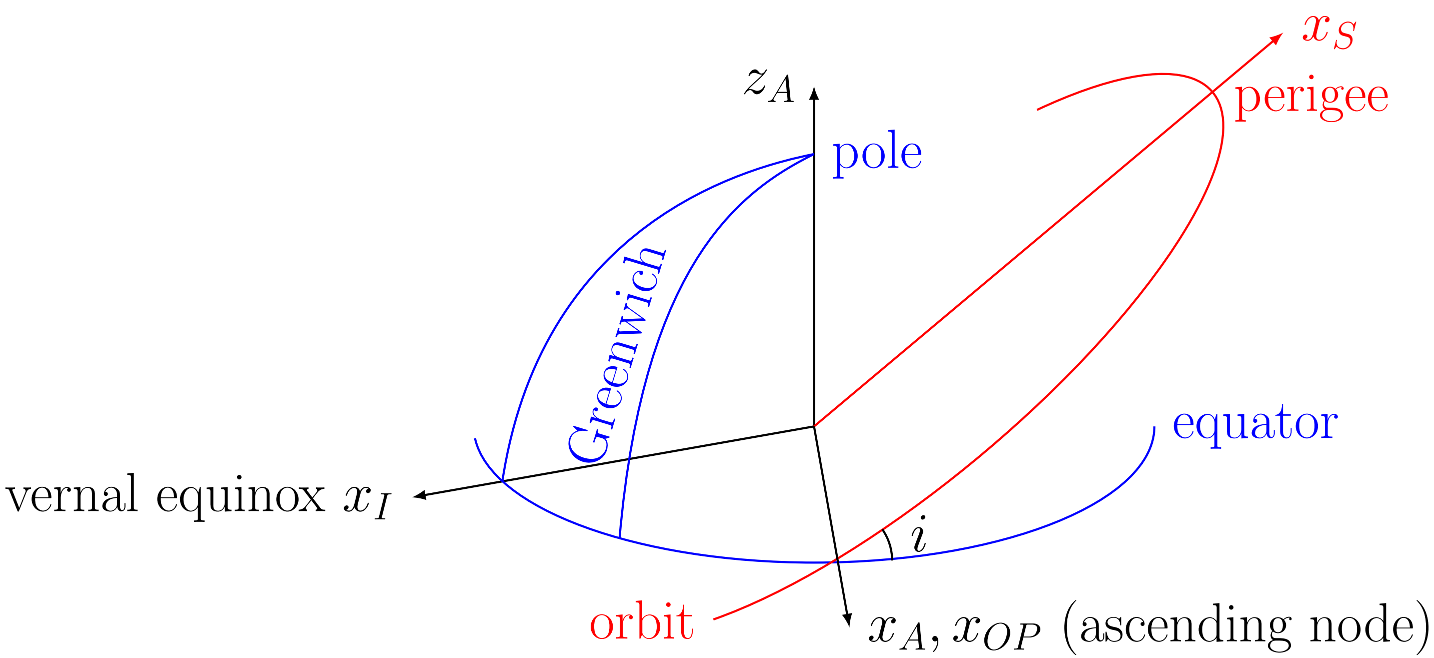


The orientation of the orbital plane may be described by the angle Ω between x_{OP} and x_I and by the *inclination angle* i which is the angle by which we have to rotate the equatorial plane around x_{OP} to obtain the orbital plane.

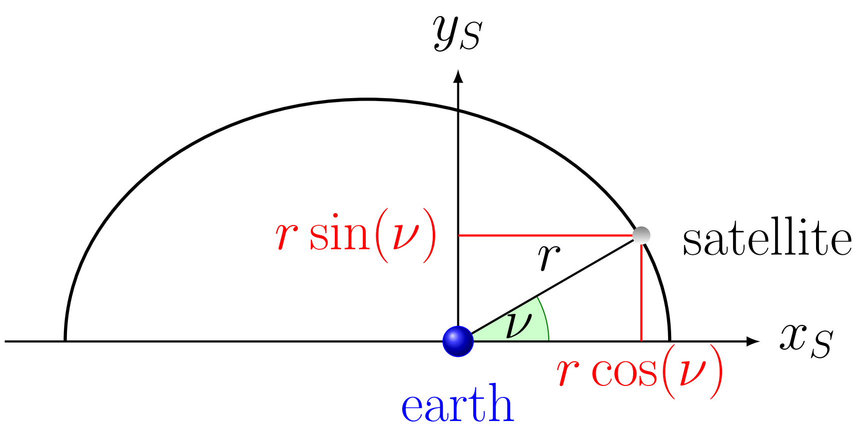
Satellite Within Orbital Plane

A satellite's orbit is an ellipse on the orbital plane, with the earth in one of its two focal points.

It is convenient to introduce one more coordinate system x_S, y_S, z_S such that x_S, y_S span the orbital plane and x_S points in the direction of the perigee (shortest distance from the earth).



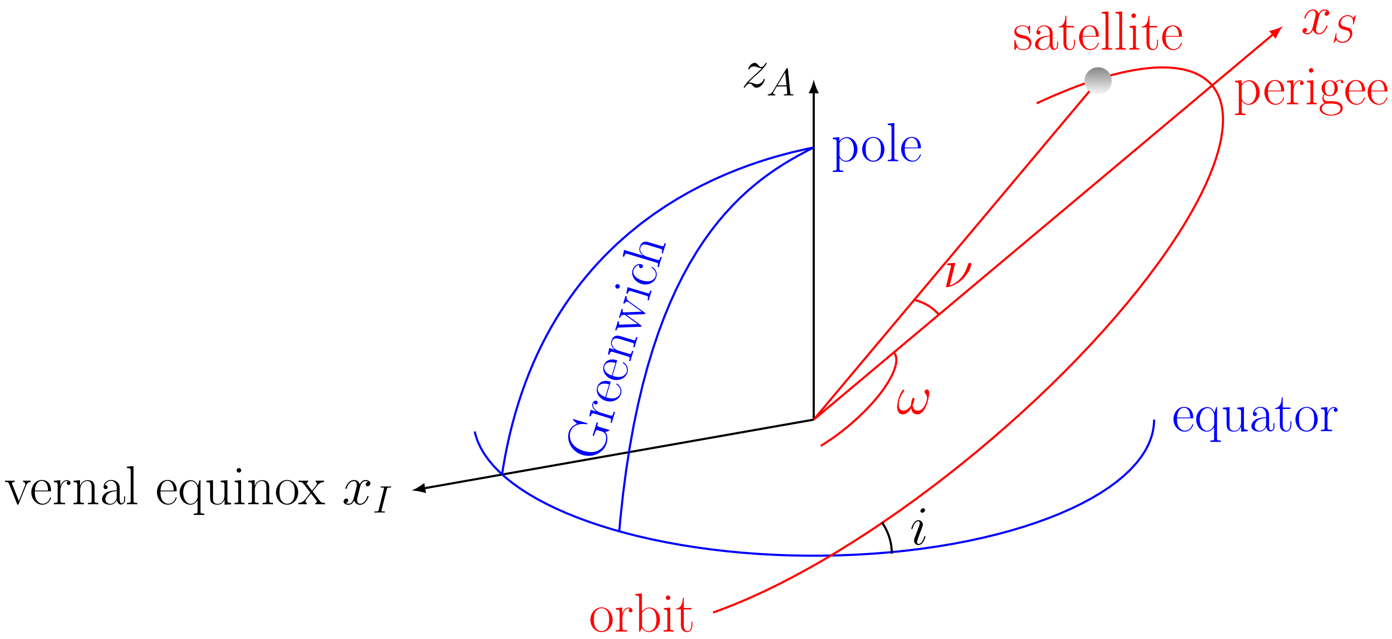
The satellite position may be described by the angle ν from x_S and by the distance $r(\nu)$ from the earth.[2]



It is then straightforward to convert from polar coordinates ν, r to Cartesian coordinates x_S, y_S, z_S (with $z_S = 0$).

[2] The ν in these notes is the angle that describes the position of a satellite in its orbit and has nothing to do with the Doppler shift considered in previous notes and also denoted by ν. The angle ν is used only for intermediate calculations. Hence there is no risk for confusion in the actual implementation.

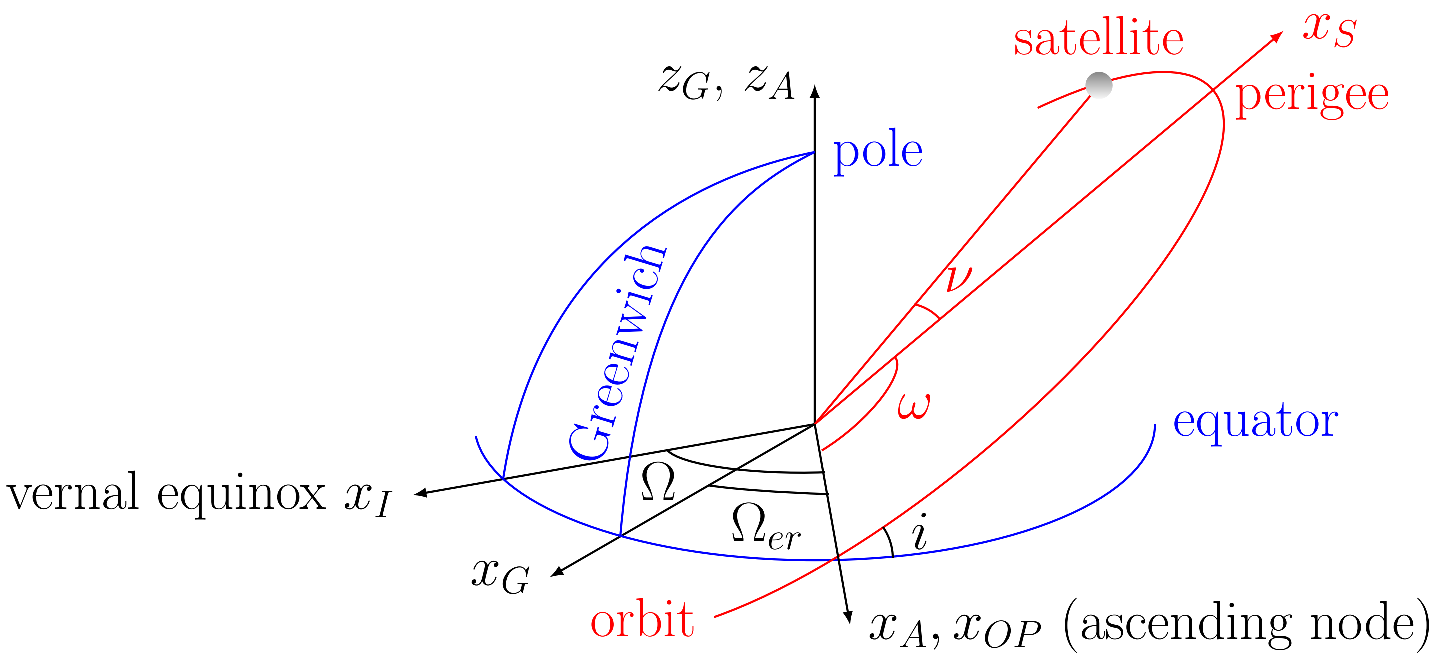
Let ω be the angle from x_{OP} to x_S, so that x_{OP}, y_{OP}, z_{OP} is obtained by rotating x_S, y_S, z_S by $-\omega$ around z_S.



Earth-Centered Earth Fixed Reference System

Eventually we want to have the satellite position in an earth-centered earth-fixed (ECEF) coordinate system.

This is a system x_G, y_G, z_G that has x_G, y_G on the equator plane and rotates with the earth. So a point on the earth has fixed x_G, y_G, z_G coordinates. We let x_G go through the Greenwich meridian.



The x_A, y_A, z_A coordinates are obtained from x_G, y_G, z_G by rotating around z_A by $\Omega_{er} = \Omega_0 - \dot{\Omega}_e t$ (see the figure), where $\dot{\Omega}_e$ is the angular velocity of the earth, t is the GPS time, and Ω_0 is a parameter sent by the satellite. (Because it is the angle from the G to the A system, Ω_{er} decreases as a function of time.[3])

[3] To us it makes more sense to describe the angle from A to G, which increases with time, but the GPS designers have thought otherwise.

Putting The Transformations Together

Let p_i, $i = 1, \ldots, 4$ be the position of a satellite in the coordinate system x_i, y_i, z_i, $i \in \{S, G, I, OP\}$.

$$\begin{aligned}
p_S &= (r\cos(\nu), r\sin(\nu), 0)^T \\
p_{OP} &= R_3(-\omega) p_S = \big(r\cos(\nu+\omega), r\sin(\nu+\omega), 0\big)^T \\
p_A &= R_1(-i) p_{OP} \\
p_G &= R_3(-\Omega_{er}) p_A
\end{aligned}$$

p_G is the sought ECEF coordinate.

The transformation from p_A to p_G depends on time. Hence p_G depends on time even if p_A is a constant (normally it isn't of course).

We let ECEF(ξ) be the ECEF coordinate system valid at GPS time ξ and for an object of interest (satellite or receiver) we let $p(t, \xi)$ be its position at GPS time t with respect to ECEF(ξ).

Unless otherwise specified, when we talk about the ECEF position we mean $p(t, t)$. For an earth-fixed object, this position does not depend on t .

For a fixed Δ, $p(t, t + \Delta)$ does not depend on t for an earth-fixed object.

Notice also that for any t, ξ, and Δ,

$$p(t, \xi + \Delta) = R_3(\dot{\Omega}_e \Delta) p(t, \xi). \tag{1}$$

A Summary and A Refinement

There are several ways to describe the position of a satellite.

Thanks to Newton, if we know the position and velocity of a satellite at a given epoch we can determine the position and the velocity of the satellite at any future (or past) epoch. This is done by solving a differential equation. It takes 6 parameters to specify position and velocity.

Alternatively, we explicitly describe the orbit (the ellipse) and the satellite position within the orbit at some specified epoch. This also requires 6 parameters.

The ones we have described so far are:

- parameters of the orbital plane: *inclination* i with respect to the equatorial plane and *ascending node argument* Ω with respect to the vernal equinox.
- ellipse orientation within orbital plane: the *argument of perigee* ω.
- shape of the ellipse: *semi-major axis* a and *eccentricity* e
- position of the satellite in the orbit at one epoch: this can be done by describing the *true anomaly* ν (the angle between the perigee and the satellite position).

The satellite sends 5 of the above 6 parameters, namely i, Ω, ω, a, e. All these are constant parameters (at least in theory, we'll come back to reality later).

The 6th parameter, namely $\nu(t)$, depends on t. Instead of sending $\nu(t)$ for every value of t (which would require much communication), the satellite sends a constant from which $\nu(t)$ can be derived. The details follow.

The key to understand this are Kepler's laws and the relationship that an ellipse of semimajor axis a and semiminor axis b has with a circle or radius a.

The third law tells us that the period of a satellite on an elliptic orbit of semimajor axis a is the same as that of a satellite on a circular orbit of radius a.

For circular orbits we can compute the velocity by equating the gravitational force $\frac{m_s M_e \mu_e}{a^2}$ (where m_s is the satellite mass, M_e the mass of the earth, and μ_e the gravitational constant) to the centrifugal force $\frac{m_s v^2}{a}$.

By solving for the velocity v we obtain $v = \sqrt{M_e \mu_e / a}$ and then we can determine the period $T = \frac{2\pi a}{v} = 2\pi \sqrt{a^3 / M_e \mu_e}$ as well as the angular velocity n of a satellite on a circular orbit of radius a, namely

$$n = \frac{2\pi}{T} = \sqrt{\frac{M_e \mu_e}{\color{red}{a}^3}}$$

where we are using red for parameters sent by the satellite.

From Kepler's second law we know that the area $A(\nu)$ comprized between the angle ν and the orbit increases linearly with time. Hence it can be written as

$$A(\nu) = \frac{A}{2\pi} M$$

where A is the area of the ellipse and M increases linearly with time (modulo 2π). Specifically, $A = ab\pi$ and M may be written as

$$M = M_0 + n(t - t_{oe}) \pmod{2\pi}.$$

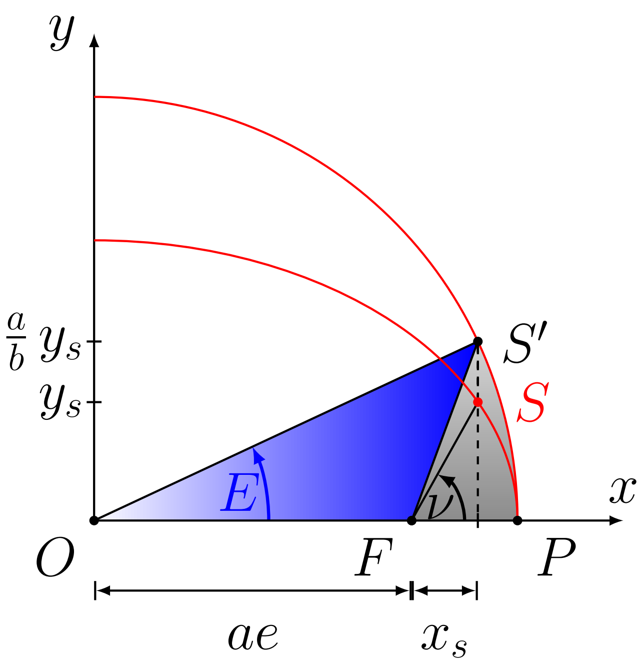
M is called *mean anomaly*[4] and M_0 is the mean anomaly at reference time t_{oe}. For a satellite on a circular orbit, $\nu = M$.

As we see next, M is what allows us to compute the satellite position at all times, based only on M_0 and some constants.

From a, M_0 and t_{oe} we can determine $A(\nu)$ and then the satellite position $p_S = (x_S, y_S, 0)$. The details follow.

[4] Angles are called anomalies in planetary motion's jargon.

We introduce the *eccentric anomaly* E as the angle measured at the center of the orbit between the perigee and the projection of the satellite position on a circle of radius a as shown in the next figure.



To determine E we write the area OPS' two ways, namely as $\frac{a^2\pi}{2\pi}E$ and as $\frac{ae(a\sin(E))}{2} + \frac{a}{b}A(\nu)$, where the first term is the area of the triangle OFS' and the second the area of FPS'.

Equating, using $\frac{a}{b}A(\nu) = \frac{a}{b}\frac{A}{2\pi}M = \frac{a^2M}{2}$, and canceling common terms yields

$$E - e\sin(E) = M.$$

Unfortunately we don't have a closed form solution for E but we can solve iteratively by starting with $E_0 = M$ and following the recursion that computes E_1, E_2, etc. according to $E_{k+1} = M + e\sin(E_k)$ until the change is small (say smaller than 10^{-13}).

Once we have E we immediately obtain the coordinates p_S of the satellite

$$\begin{aligned}
x_S &= a\cos(E) - ae \\
y_S &= (a\sin(E))\frac{b}{a} = b\sin(E) = a\sqrt{1-e^2}\sin(E) \\
z_S &= 0.
\end{aligned}$$

Because of certain correction terms to be introduced shortly, it turns out that we have to compute p_S from r and ν. From x_S, and y_S we obtain

$$
\begin{aligned}
r &= a(1 - e\cos(E)) \\
\nu &= \tan^{-1}\left(\frac{\sqrt{1-e^2}\sin(E)}{\cos(E) - e}\right).
\end{aligned}
$$

Then

$$p_S = (r\cos(\nu), r\sin(\nu), 0)^T.$$