LECTURE NOTES ON

SATellite NAVIGATION
Satellite navigation

“Satellite navigation”, or better “satellite-based navigation”, means navigation based on signals transmitted by artificial satellites. Sometimes, this notation is misunderstood, in the sense that “satellite navigation” is interpreted as navigation of the satellites themselves, i.e. the determination and the correction of their orbits.

Historical overview

The beginning of satellite navigation can be dated back to the launch of the Russian satellite Sputnik in 1957. With this satellite flying in space, many studies were developed on the determination of its orbit by exploiting the Doppler effect. In principle, with the opposite reasoning, the estimated Sputnik orbit could be used for terrestrial positioning and navigation: if a single ground station is sufficient to determine a satellite orbit by Doppler shift data, then an unknown receiver position on the Earth can be derived from the same type of Doppler measurements, assuming that the satellite orbit is known. Let us recall here that the Doppler effect produces a frequency variation (frequency shift) of the signal as perceived by a receiver in relative motion with respect to the emitter. In particular, defining as \( v \) their relative velocity (along the line of sight, i.e. in the direction between emitter and receiver), the Doppler effect to a first approximation can be expressed as

\[
\Delta f = f_r - f_e = -\frac{f_e}{c} v
\]

where \( f_e \) and \( f_r \) are the frequencies of the emitted and received signal, respectively. In other words, the Doppler effect provides a measure of the velocity of the moving emitter (or of the moving receiver) and, after integration in time, the distance between emitter and receiver.

On the basis of this experience, in the Sixties the United States Navy developed a system called Transit. In 1964 this system became operational for military use; in 1967 it became operational for civilians as well. It was dismissed in 1996. The subsequent GPS system is based on the same principles of the Transit system. Transit satellites flew over circular and polar orbits at an altitude of 1075 km (LEO = Low Earth Orbit); typically seven satellites were active and each of them transmitted signals on two different carriers (150 and 400 MHz). Every time a satellite was in view of a ground station, this station received the two signals emitted by the satellite, computed the satellite orbit by exploiting the Doppler effect and transmitted to the satellite the predicted orbit for the next 12 hours. Knowing the satellite orbit, any receiver could compute its position when the satellite was in view, with an accuracy of the order of tens of meters in navigation mode.
It has to be recalled that in 1974 the Soviet Union launched to space the Tsikada system, which was based on the same principles of Transit. Nominally, the Tsikada system is still operational. However the real revolution in the satellite navigation happened with the second generation of satellite systems, launched around the Eighties, namely the American system GPS (Global Positioning System) and the Russian counterpart GLONASS (GLObal Navigation Satellite System). In the next future, two new systems will be operational, namely the European system called Galileo and the Chinese system called Compass.

Global Positioning System (GPS)

The GPS positioning system is based on the reception of radio-signals emitted by a constellation of artificial satellites orbiting the Earth. The system is designed to allow, in any time and place, the tri-dimensional positioning of objects (including moving objects).

The system architecture consists of 3 parts:
- Space segment, i.e. the satellite constellation;
- Control segment, i.e. the set of ground stations monitoring and managing the system;
- User segment, i.e. the civil and military users that make use of suitable receivers with antenna.
**Space segment**

The nominal GPS constellation consists of 24 operational satellites, which have nearly circular orbits at an altitude of about 20200 km above the Earth (MEO = Mean Earth Orbit). The orbital period is 12 hours. Each satellite has an average lifetime of 7.5 years. The 24 satellites are deployed on 6 evenly-spaced orbital planes, with an inclination of 55° with respect to the equatorial plane. Nominally, there are 4 satellites per orbital plane.

![GPS constellation](image)

**GPS constellation**

A global coverage with at least four satellites is always guaranteed. It is possible to observe up to eight satellites with a minimum elevation of 15° above the horizon, up to 10 satellites with an elevation higher than 10° and up to 12 satellites with an elevation higher than 5° (the elevation is defined as the angle between the satellite-receiver direction and the plane tangential to the ellipsoid at the receiver location, see Figure below).

![Satellite elevation η](image)

**Satellite elevation η**

The satellites, with a mass of the order of 800 kg and equipped with solar panels having a surface of about 7.2 m², has been launched in different periods and belong to different blocks:
- Block I: experimental and launched from 1978 to 1985 (all of them are no longer used);
- Block II, SVN 13-21: launched from 1989 to 1990;
- Block IIA, SVN 22-40: launched from 1990 to 1997;
- Block IIR, SVN 41-62: launched from 1997 up to today (last launch 2007 October 17);
- Block IIF: under development.

The Block I and Block IIA satellites carry four oscillators, two cesium atomic clocks and two rubidium atomic clocks; the Block IIR satellites carry three rubidium atomic clocks.
**Control segment**

It consists of ground control stations. There are five stations, placed in equatorial areas (Hawaii and Colorado Springs (USA), Ascension in the South Atlantic Ocean, Diego Garcia in the Indian Ocean and Kwajalein in the North Pacific Ocean), with the task of “tracking” and correcting satellite orbits, monitoring and synchronizing the satellite clocks, uploading the satellite ephemerides (i.e. the satellite positions) predicted for the subsequent 12/24 hours, which will be in turn sent from the satellites to the users. The station located at Colorado Spring represents the master station, collecting data from the other control stations and carrying out the main computations.

![Diego Garcia station](image)

**User segment**

It consists of users having receivers with antenna. These instruments are passive, in the sense that they do not emit any signal but they are just able to receive data. The position that will be determined is the position of the phase centre of the antenna, which can be kept fixed on a tripod-benchmark or can be in motion. The receiver can be indifferently on ground, on an airplane, on a space-shuttle (at an altitude of 150 km above the Earth) and so on.
The precision of the system is guaranteed by the fact that the signal emitted from the satellites are controlled by atomic clocks, which stably “oscillate” at the fundamental frequency of 10.23 MHz (corresponding to a wavelength of the order of 30 m). Starting from this frequency, it is possible to generate the three main parts of the GPS data message:

- **Carriers**: 2 sinusoidal waves called L1 and L2;
- **Codes**: 2 binary sequences called C/A (Coarse Acquisition Code) and P (Precise Code);
- **Navigation message**: code D.

### Carriers
They are generated by multiplying the fundamental frequency by 154 (carrier L1) and by 120 (carrier L2), thus obtaining

- **Carrier L1**: \( f_1 = 1575.42 \text{ MHz} \), \( \lambda_1 \approx 19 \text{ cm} \)
- **Carrier L2**: \( f_2 = 1227.60 \text{ MHz} \), \( \lambda_2 \approx 24 \text{ cm} \)

The use of these dual frequencies is essential to eliminate the different sources of error, such as for example the delay due to the ionospheric refraction.

### Codes
They are sequences of binary values (+1 or −1), generated by an algorithm that is periodically repeated in time (for this reason these sequences are called “pseudo-random”). The codes are used to modulate the amplitude of the carriers. Defining the code period \( T \) as the time needed for the transmission of the whole sequence (before repeating it), two codes are available:

- **Code C/A**: \( f_1 = 0.1 \times 10.23 \text{ MHz} \), \( \lambda_1 \approx 300 \text{ m} \), 1023 bits (called chips) \( \rightarrow T = 1 \text{ ms} \)
- **Code P**: \( f_2 = 10.23 \text{ MHz} \), \( \lambda_2 \approx 30 \text{ m} \), \( T = 37 \text{ weeks} \)
The C/A code is available for civilian use and is presently modulated upon the L1 carrier only; it is used to identify the satellite (each sequence is different from satellite to satellite). On the other hand, the P code is common to all satellites, is modulated upon both carriers and guarantees higher accuracy; the GPS design includes the possibility for the system manager (Department of Defence) to make visible or hidden the P code in such a way that it can be used by all the user or by authorized user only (typically military users). This procedure is called Anti-Spoofing (A/S) and consists in encrypting the P code to the Y code (encrypted).

![Code modulation of the carrier](image)

**Navigation message**

It is just a binary code, which is modulated on the carriers and which consists of 25 blocks of 30 seconds each. It contains information on the satellite health status, on the satellite clock and on the satellite ephemerides, i.e. the parameters for the computation of the satellite position.
It has to be noted that the Block IIF satellites (modernized GPS) will make available a third frequency (denoted as L5 and obtained by multiplying the fundamental frequency by 115) for the civilian use. The name L5 has been chosen because currently the satellites transmit some additional signals called L3 and L4 for a military use only. A clear separation between codes for civilian users and codes for military users is foreseen too.

Basically, there are two different types of GPS observations:
- Pseudo-range observations (pseudo-distance);
- Phase observations.

**Pseudo-range observations**
The idea is to measure the signal “flight time”, i.e. the time interval between the emission of the signal by the satellite and the reception of the same signal by the receiver. The measure is performed by a correlation procedure between two signals. The receiver, after identifying the satellite (by means of the C/A code), can generate a replica of the code sequence and compare it with the signal received from the satellite. The phase shift between the two signals is a measure of the flight time. This procedure can be indifferently applied to the C/A code or to the P code.

![Measurement principle of code observations](image)

**Measurement principle of code observations**

Defining \( P^r_s(t) \) as the measured distance between the satellite \( s \) and the receiver \( r \) (pseudo-range observations), in principle the following relation holds:

\[
P^r_s(t) = c \Delta t^r_s(t)
\]

where \( c \) is the velocity of the light in the vacuum and \( \Delta t^r_s(t) \) is the measured flight time. Different time scales have to be considered, namely:

1. reference time scale, which is also called GPS time;
2. satellite time scale, which is not perfectly synchronized with the GPS time and therefore it is characterized by an offset \( dt^s(t) \);
3. receiver time scale, with an offset \( dt_r(t) \).

Note that these offsets (also called clock delays) change in time, for instance due to drift effects in the satellite and receiver clocks.
Therefore, the measured flight time will be equal to

\[
\Delta t^r_s(t) = [t_r + dt_r(t)] - [t^s + dt^s(t)] = \Delta t + dt_r(t) - dt^s(t)
\]
where \( t_r \) and \( t_s \) are, respectively, the emission and reception time of the signal with respect to the reference time scale (GPS time). \( \Delta t \) is the actual flight time since it refers to the same time scale (GPS time) and it is of the order of 66 ms, taking into account the altitude of the satellite orbit. Therefore the pseudo-range observation equation, disregarding the atmospheric delays that will be discussed afterwards, can be written as

\[
P_r'(t) = \rho_r' + c\left[dt_r(t) - dt_s'(t)\right]
\]

The atomic clocks of the satellites can be considered as synchronized to each other and the corresponding clock delay can be modelled by a second order polynomial. This introduces errors from tens of centimetres to some meters. The same reasoning cannot be repeat for the receiver clock delays since the receivers generally have quartz oscillators that are characterized by a lower stability. Modelling the receiver clock delays would lead to positioning errors of the order of some hundreds of kilometres, which is of course unacceptable. For this reason the clock delay of the receiver is considered as an unknown together with the satellite-receiver distance \( \rho_r' \). This distance can be written as

\[
\rho_r' = \sqrt{(X_r - X')^2 + (Y_r - Y')^2 + (Z_r - Z')^2}
\]

where \((X_r, Y_r, Z_r)\) are the unknown coordinates of the receiver, while \((X', Y', Z')\) are the satellite coordinates derived from the satellite orbits. These orbits can be known with different precision according to the type of ephemerides used, namely:

1. ephemerides predicted by the control segment and transmitted by the satellites to the receiver (therefore in real time). These ephemerides have a precision of some meters.
2. “rapid” ephemerides, computed a-posteriori (within half day) by the International GNSS Service (IGS). These ephemerides have a precision of about 10 centimetres.
3. “precise” ephemerides, computed a-posteriori (within 12 days) again by IGS with a precision of about 5 centimetres.

Defining \( n_s \) as the number of visible satellites and \( n_t \) as the number of observed epochs, we have:

1. in the static positioning (i.e. with the receiver still in place), \( n_s \cdot n_t \) observations and \( 3 + n_t \) unknowns (i.e. three coordinates plus the receiver clock delay for every epoch), therefore

\[
n_s n_t \geq 3 + n_t, \quad n_t \geq \frac{3}{n_s - 1}
\]

meaning that only 2 satellites for a minimum period of 3 epochs are sufficient to obtain a solution. This solution is however bad-conditioned and it is generally preferred to use at least 4 satellites in order to guarantee a solution even with a single observation epoch.

2. in the kinematic positioning, the number of unknowns becomes \( 4n_t \) because the coordinates of the receiver change in time, therefore

\[
n_s n_t \geq 4n_t, \quad n_s \geq 4
\]

meaning that at least 4 satellites are always required to track the trajectory of the moving vehicle.

**Phase observations**

The satellite-receiver range can be also determined by means of phase measurements on the L1 and L2 carriers. Let us imagine to follow the satellite along its orbit, considering an initial epoch \( t_0 \) and a generic epoch \( t \).
At the epoch $t_0$ the range $R$ observed by the receiver can be written as

$$R_\circ^r(t_0) = \lambda \phi_\circ^r(t_0) + \lambda N_\circ^r(t_0)$$

where $\phi_\circ^r(t_0)$ is the observed phase shift between the signal received from the satellite and the replica internally generated by the receiver, $N_\circ^r(t_0)$ is the integer number of cycles performed by the carrier wave to cover the satellite-receiver path and, finally, $\lambda$ is the carrier wavelength. As a matter of fact, the receiver measures $\phi_\circ^r(t_0)$ only, while the term $N_\circ^r(t_0)$ represents a new unknown of the problem, which is called “initial phase ambiguity”.

At the epoch $t$ the satellite is located at a different position along the orbit and the new measurement of the satellite-receiver range can be again expressed as

$$R_\circ^r(t) = \lambda \phi_\circ^r(t) + \lambda N_\circ^r(t)$$

where

$$N_\circ^r(t) = N_\circ^r(t_0) + C(t_0, t)$$

This time the receiver is able to count the number of cycles (positive or negative) $C(t_0, t)$ performed by the carrier from the epoch $t_0$ of the first phase-lock to the epoch $t$. The only additional unknown is therefore the initial phase ambiguity at the epoch $t_0$. Naturally it is implicitly assumed that the receiver maintains the signal lock with the satellite during all the observation period. The loss of this lock, for example due to an obstacle along the optical path between satellite and receiver or to circuitry problems of the receiver, is called cycle-slip and causes the loss of the count of the integer number of cycles and therefore the introduction of a new unknown phase ambiguity.

Now, taking also into account the clock delays

$$R_\circ^r(t) = \rho_\circ^r + c[dt_r(t) - dt^*_r(t)]$$

and expressing the phase equation in terms of the real observable $\phi_\circ^r(t)$ (in cycles), it holds that

$$\phi_\circ^r(t) = \frac{\rho_\circ^r}{\lambda} - N_\circ^r(t) + f[dt_r(t) - dt^*_r(t)]$$

with the frequency $f = c/\lambda$. 

Measurement principle of phase observations
As regards the basic satellite configuration that is necessary to get a solution, we have:

1. in the static positioning, \( n_s \) observations and \( 3n_t + n_s \) unknowns (i.e. three coordinates, one clock delay for every epoch and one initial phase ambiguity for each satellite), therefore

\[
n_s n_t \geq 3 + n_t + n_s \\
\]

meaning that only 2 satellites for a minimum period of 5 epochs are sufficient to obtain a solution. As in the case of code observations, this solution is bad-conditioned and in general it is preferred to consider at least 4 satellites, thus obtaining a solution with a minimum of three observation epochs.

2. in the kinematic positioning the number of unknowns becomes \( 4n_t + n_s \) because the coordinates of the observed point change in time, therefore

\[
n_s n_t \geq 4n_t + n_s \\
\]

meaning that at least 5 satellites are required to get a solution with 5 observation epochs. If 6 satellites are available, 3 epochs are needed. If 8 satellite are available, only 2 epochs are sufficient. However, the kinematic solution with a single observation is not possible due to the initial phase ambiguities. In this case it is necessary to define a “initialization” procedure, for instance keeping the receiver still until the initial phase ambiguities are estimated. Special techniques allowing for the determination of the initial phase ambiguities even when the receiver is moving have been developed and are known as On The Fly (OTF) techniques.

The precision of GPS measurements also depends on the atmospheric effects (or delay), which are connected to the fact that the signal emitted by the satellite crosses the atmosphere before arriving at the receiver. The propagation velocity of the signal changes according to the physical state of the atmosphere. The resulting delay has to be taken into account because the corresponding positioning error is generally higher than 5 meters and can be even of the order of 200 meters. Basically, it is possible to detect three types of atmospheric delay:

- optical path curvature
- tropospheric refraction
- ionospheric refraction

**Optical path curvature**

When crossing a medium, any electromagnetic signal follows the optical path that implies the minimum travelling-time (Fermat’s principle). This path does not necessarily coincide with the geometrical distance. It can be modelled according to the following empirical model

\[
R(\eta) = \frac{1.92}{\eta^2 + 0.6}
\]

where \( R \) is the curvature effect on the path length in meters, while \( \eta \) is the elevation angle in degrees. This curvature effect will be neglected later on.
Optical path curvature of the electromagnetic signal

Tropospheric refraction
The troposphere represents the “low” part of the atmosphere, between the Earth surface and an altitude of about 40 km. It can be divided into two parts:
- “hydrostatic” part: between the Earth surface and an altitude of about 11 km
- “dry” part: in the altitude interval between 11 km and 40 km.

The tropospheric refraction always produces a delay in the signal transmission, so that the measure of the satellite-receiver range is systematically longer. The tropospheric delay is independent from the signal frequency and so it is the same for the L1 and L2 carriers. It depends on atmospheric parameters (such as pressure, water vapour, temperature) and on the zenith angle \( z \) of the satellite. The tropospheric delay significantly increases for a zenith angle above 75° (corresponding to an elevation angle of 15°). Among the various models used to describe the tropospheric delay, we mention here the Saastamoinen model and the Hopfield model. Some typical values of the tropospheric delay are reported in the following table, as a function of the satellite elevation.

<table>
<thead>
<tr>
<th>Elevation (°)</th>
<th>Low (m)</th>
<th>Medium (m)</th>
<th>High (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>8.9</td>
<td>9.2</td>
<td>10.4</td>
</tr>
<tr>
<td>30</td>
<td>4.6</td>
<td>4.8</td>
<td>5.4</td>
</tr>
<tr>
<td>45</td>
<td>3.3</td>
<td>3.4</td>
<td>3.8</td>
</tr>
<tr>
<td>90</td>
<td>2.3</td>
<td>2.4</td>
<td>2.7</td>
</tr>
</tbody>
</table>

Low: dry and cold weather. Medium: standard weather. High: humid and warm weather

Ionospheric refraction
The ionosphere represents the “high” part of the atmosphere, from an altitude of 40-50 km to an altitude of about 1000 km. The ionospheric delay (which is positive for code observations and negative for phase observations) depends on the signal frequency and therefore it is different for the L1 and L2 carriers. This delay is mainly related to the total electron content (TEC), which in turn depends on the intensity of the solar activity and on the intensity of the solar radiation incident on the atmosphere (day-time and night-time period). Among the various models used to describe the ionospheric delay, we mention here the Klobuchar model. Some typical values of the ionospheric delay are reported in the following table, as a function of the satellite elevation.

<table>
<thead>
<tr>
<th>Elevation (°)</th>
<th>Low (m)</th>
<th>Medium (m)</th>
<th>High (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>3.9</td>
<td>18</td>
<td>180</td>
</tr>
<tr>
<td>30</td>
<td>2.0</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>45</td>
<td>1.4</td>
<td>7</td>
<td>70</td>
</tr>
<tr>
<td>90</td>
<td>1.0</td>
<td>5</td>
<td>50</td>
</tr>
</tbody>
</table>

Low: night-time period. Medium: day-time period, normal solar activity. High: day-time period, high solar activity.
Defining $J$ as the ionospheric delay and $T$ as the tropospheric delay, the observation equation for pseudo-range code observations can be written as

$$P_r^s (t) = \rho_r^s + c [dt_r (t) - dt^s (t)] + J_r^s (t) + T_r^s (t)$$

while, the observation equation for phase observations is

$$\phi_r^s (t) = \frac{\rho_r^s}{\lambda} - N_r^s (t) + f [dt_r (t) - dt^s (t)] - \frac{J_r^s (t)}{\lambda} + \frac{T_r^s (t)}{\lambda}$$

The relation between the two different ionospheric effects on the L1 and L2 carriers is

$$J_{r,L2} = \frac{\lambda_{L2}^2}{\lambda_{L1}^2} J_{r,L1}$$

In addition to the atmospheric delays there are other sources of error in the GPS system, such as the ones related to relativistic effects, to variations of the antenna phase centre, to the so-called multi-path (which happens when part of the signal arrives at the receiver antenna not directly from the satellite, but reflected by other surfaces), to the receiver electronics, etc.

Then there are the random errors that, roughly, are of the order of 3 m for the C/A code, 30 m for the P code and 2 mm for the carriers.

Let us consider now three different techniques for positioning/navigation:

- Absolute positioning (point positioning)
- Relative positioning
- Differential GPS positioning (DGPS or in general DGNSS)

**Point positioning**

The absolute positioning of a receiver means the estimation of the receiver coordinates and of its clock delay (typically using code observations). It can be performed in static mode, i.e. standing on the same position for a certain time period, or in kinematic mode when the receiver is in motion. In the latter case it is necessary to estimate the absolute position of the receiver epoch by epoch, with at least 4 visible satellites.

The estimates of the unknown is based on a least-squares (LS) adjustment, after linearizing the observation equations (starting from an approximate value of the coordinates).

Since we are interested to the navigation solution, let us set up the LS problem for a single epoch, assuming that approximate values of the receiver coordinates ($\tilde{X}_r, \tilde{Y}_r, \tilde{Z}_r$) are available; moreover, ionospheric and tropospheric delays are assumed to be known, as well as the satellite position and the satellite clock delay (by using predicted or precise ephemerides and the information transmitted in the navigation message).

Taking into account the measurement noise $\nu_p$, it holds that

$$P_r^s (t) = \rho_r^s (t) + c [dt_r (t) - dt^s (t)] + J_r^s (t) + T_r^s (t) + \nu_p =$$

$$= \sqrt{(X_r - X^s)^2 + (Y_r - Y^s)^2 + (Z_r - Z^s)^2 + c dt_r (t) + [- c dt^s (t) + J_r^s (t) + T_r^s (t)]} + \nu_p$$

13
By linearizing $\rho^s_r$ with respect to the receiver coordinates, we get

$$\rho^s_r \approx \tilde{\rho}_r^s + \frac{\partial \rho^s_r}{\partial X_r} (X_r - \tilde{X}_r) + \frac{\partial \rho^s_r}{\partial Y_r}(Y_r - \tilde{Y}_r) + \frac{\partial \rho^s_r}{\partial Z_r}(Z_r - \tilde{Z}_r) =$$

$$= \tilde{\rho}_r^s + \frac{(\tilde{X}_r - X^s)}{\sqrt{(\tilde{X}_r - X^s)^2 + (\tilde{Y}_r - Y^s)^2 + (\tilde{Z}_r - Z^s)^2}} (X_r - \tilde{X}_r) +$$

$$+ \frac{(\tilde{Y}_r - Y^s)}{\sqrt{(\tilde{X}_r - X^s)^2 + (\tilde{Y}_r - Y^s)^2 + (\tilde{Z}_r - Z^s)^2}} (Y_r - \tilde{Y}_r) +$$

$$+ \frac{(\tilde{Z}_r - Z^s)}{\sqrt{(\tilde{X}_r - X^s)^2 + (\tilde{Y}_r - Y^s)^2 + (\tilde{Z}_r - Z^s)^2}} (Z_r - \tilde{Z}_r) =$$

$$= \tilde{\rho}_r^s + \frac{(\tilde{X}_r - X^s)}{\tilde{\rho}_r^s} (X_r - \tilde{X}_r) + \frac{(\tilde{Y}_r - Y^s)}{\tilde{\rho}_r^s} (Y_r - \tilde{Y}_r) + \frac{(\tilde{Z}_r - Z^s)}{\tilde{\rho}_r^s} (Z_r - \tilde{Z}_r)$$

In this way the linearized pseudo-range equation becomes

$$P^s_r = \frac{(\tilde{X}_r - X^s)}{\tilde{\rho}_r^s} (X_r - \tilde{X}_r) + \frac{(\tilde{Y}_r - Y^s)}{\tilde{\rho}_r^s} (Y_r - \tilde{Y}_r) + \frac{(\tilde{Z}_r - Z^s)}{\tilde{\rho}_r^s} (Z_r - \tilde{Z}_r) +$$

$$+ cdt(t) \left[ \tilde{\rho}_r^s - cdt^s + J_r^s + T_r^s \right] + \nu_p =$$

$$= \frac{(\tilde{X}_r - X^s)}{\tilde{\rho}_r^s} \partial X_r + \frac{(\tilde{Y}_r - Y^s)}{\tilde{\rho}_r^s} \partial Y_r + \frac{(\tilde{Z}_r - Z^s)}{\tilde{\rho}_r^s} \partial Z_r + cdt + b_r^s + \nu_p$$

where

$$\partial X_r = X_r - \tilde{X}_r; \quad \partial Y_r = Y_r - \tilde{Y}_r; \quad \partial Z_r = Z_r - \tilde{Z}_r$$

$$b_r^s = \tilde{\rho}_r^s - cdt^s + J_r^s + T_r^s$$

The above equation can be written for all the $n$ visible satellites at the considered epoch $t$. The resulting LS model is then

$$Y_0 = Ax + b + \nu$$

where

$$x = \begin{bmatrix} \delta X_r \\ \delta Y_r \\ \delta Z_r \\ cdt_r \end{bmatrix}$$
where the noise vector $\nu$ is characterized by the following covariance matrix

$$C_\nu = \sigma_0^2 Q = \sigma_0^2 I$$

meaning that observations from different satellites are considered uncorrelated and of the same precision. The LS estimate of the unknown vector $\tilde{x}$ is given by

$$\begin{bmatrix} \delta\tilde{X}_r \\ \delta\tilde{Y}_r \\ \delta\tilde{Z}_r \\ c\delta t_r \end{bmatrix} = \left( A^\top A \right)^{-1} A^\top \left( Y_0 - b \right)$$

and, consequently, the estimated coordinates of the receiver are

$$\begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} = \begin{bmatrix} \tilde{X}_r \\ \tilde{Y}_r \\ \tilde{Z}_r \end{bmatrix} + \begin{bmatrix} \delta\tilde{X}_r \\ \delta\tilde{Y}_r \\ \delta\tilde{Z}_r \end{bmatrix}$$

There are some rough indexes that allow to evaluate the quality of the survey, both in static and kinematic mode. These indexes, which are not always implemented in low-level GPS receivers, are:
- the number of visible satellites (the more satellites, the better accuracy);
- the signal-to-noise ratio S/N, which is crucial to maintain the signal lock with the satellite;
- the positioning dilution of precision (PDOP); this index is given by

$$\text{PDOP} = \sqrt{\text{Tr} N^{-1}}$$

where $N^{-1} = (A^\top A)^{-1}$ is the inverse of the normal matrix of the LS problem; the PDOP index, multiplied by $\sigma_0$ represents the overall precision of the pure geometrical determination of the point. In addition to the number of tracked satellites, the matrix $N$ also depends on their position with respect to the receiver (sky plot), i.e. on the spatial distribution of the satellites in the sky. Typically the receiver discards observations with a PDOP value above a certain threshold.
Besides the PDOP, other similar indexes are sometimes taken into account, such as the dilution of precision in the height coordinates or in the horizontal coordinates (VDOP, HDOP); note that, for intrinsic reasons, the VDOP value is normally equal to 1.5 or 2 times the HDOP value.

![Bad satellite geometry: high PDOP](image1)

![Good satellite geometry: low PDOP](image2)

The point positioning technique, in navigation mode, leads to errors in the trajectory determination of the order of 1 – 3 meters (not worse than 10 meters). A significant improvement can be obtained by using a relative positioning/navigation.

**Relative positioning**

Let us consider two receivers that are collecting observations from the same satellite $k$ at the same epoch $t$. The two receivers are called master $m$ (typically still) and rover $r$ (typically in motion). Let us first consider code observations. Given the observation equations for the two receivers

\[
P_m^k = \rho_m^k + c[dt_m - dt_r] + J_m^k + T_m^k + v_m^k
\]

\[
P_r^k = \rho_r^k + c[dt_r - dt_r] + J_r^k + T_r^k + v_r^k
\]
the so-called “single differences” are the differences of the codes observed by the two receivers

\[ P_{rm}^k = P_r^k - P_m^k = (\rho_r^k - \rho_m^k) + c(dt_r - dt_m) + (J_r^k - J_m^k) + (T_r^k - T_m^k) + (v_r^k - v_m^k) \]

where the dependence on the satellite clock delays is vanished. Note also that the ionospheric and tropospheric terms are so spatially correlated that their differences can be neglected if master and rover are less than 10 km from each other.

Let us consider now two satellites \( k \) and \( h \), tracked by the receivers \( r \) and \( m \) at the same epoch \( t \). The so-called “double differences” are obtained by computing a further differentiation, namely

\[ P_{rm}^{kh} = (P_r^k - P_m^k - P_r^h - P_m^h) = \left[ (\rho_r^k - \rho_m^k) - (\rho_r^h - \rho_m^h) \right] + \left[ (J_r^k - J_m^k) - (J_r^h - J_m^h) \right] + \left[ (T_r^k - T_m^k) - (T_r^h - T_m^h) \right] + \left[ (v_r^k - v_m^k) - (v_r^h - v_m^h) \right] = P_{rm}^{rh} + J_{rm}^{kh} + T_{rm}^{kh} + V_{rm}^{kh} \]

where also the dependence on the receiver clock delays is disappeared. In the general case of more than 2 satellites, one of them is chosen as reference (pivot) and the double differences are computed with respect to this chosen satellite to avoid linearly dependent combination.

Since the master receiver is kept still while the rover is moving from point to point, it is natural to assume that the master coordinates are known with a higher precision and the linearization is made with respect to the coordinates of the rover receiver.

The linearization of the double differences is equivalent to linearizing the pseudo-range observations and then computing the double differences of the linearized quantities, i.e.

\[ P_{rm}^{kh} = (P_r^k - P_m^k - P_r^h - P_m^h) \approx \left[ \bar{\rho}_r^k + \frac{(\bar{X}_r - X_r^k)}{\bar{\rho}_r^k} \right] \delta X_r + \left[ \bar{\rho}_r^h + \frac{(\bar{X}_r - X_r^h)}{\bar{\rho}_r^h} \right] \delta X_r + \left[ \bar{\rho}_r^h + \frac{(\bar{Y}_r - Y_r^h)}{\bar{\rho}_r^h} \right] \delta Y_r + \left[ \bar{\rho}_r^h + \frac{(\bar{Z}_r - Z_r^h)}{\bar{\rho}_r^h} \right] \delta Z_r - \rho_{m}^k \]

\[ = P_{rm}^{rh} + J_{rm}^{kh} + T_{rm}^{kh} + V_{rm}^{kh} \]

namely
$$P_{rm}^{kl} \approx \begin{bmatrix} \frac{\tilde{X}_r - X^k}{\tilde{\rho}_r^k} & \frac{\tilde{X}_r - X^h}{\tilde{\rho}_r^h} \\ \frac{\tilde{Y}_r - Y^k}{\tilde{\rho}_r^k} & \frac{\tilde{Y}_r - Y^h}{\tilde{\rho}_r^h} \\ \frac{\tilde{Z}_r - Z^k}{\tilde{\rho}_r^k} & \frac{\tilde{Z}_r - Z^h}{\tilde{\rho}_r^h} \end{bmatrix} \delta X_r + \begin{bmatrix} \frac{\tilde{Y}_r - Y^k}{\tilde{\rho}_r^k} & \frac{\tilde{Y}_r - Y^h}{\tilde{\rho}_r^h} \\ \frac{\tilde{Z}_r - Z^k}{\tilde{\rho}_r^k} & \frac{\tilde{Z}_r - Z^h}{\tilde{\rho}_r^h} \end{bmatrix} \delta Y_r + \begin{bmatrix} \frac{\tilde{Z}_r - Z^k}{\tilde{\rho}_r^k} & \frac{\tilde{Z}_r - Z^h}{\tilde{\rho}_r^h} \end{bmatrix} \delta Z_r + b_{rm}^{kh} + \nu_{rm}^{kh}$$

where

$$\delta X_r = X_r - \tilde{X}_r; \quad \delta Y_r = Y_r - \tilde{Y}_r; \quad \delta Z_r = Z_r - \tilde{Z}_r$$

$$b_{rm}^{kh} = (\tilde{\rho}_r^k - \rho_{rm}^k) - (\tilde{\rho}_r^h - \rho_{rm}^h) + J_{rm}^{kh} + T_{rm}^{kh}$$

This equation can be written for each of the \(n-1\) double differences (where \(n\) is the number of visible satellites) computed at the epoch \(t\). The resulting LS model is

$$Y_0 = Ax + b + \nu$$

where

$$A = \begin{bmatrix} \frac{\tilde{X}_r - X^k}{\tilde{\rho}_r^k} & \frac{\tilde{X}_r - X^h}{\tilde{\rho}_r^h} & \cdots & \frac{\tilde{X}_r - X^{n-1}}{\tilde{\rho}_r^{n-1}} \\ \frac{\tilde{Y}_r - Y^k}{\tilde{\rho}_r^k} & \frac{\tilde{Y}_r - Y^h}{\tilde{\rho}_r^h} & \cdots & \frac{\tilde{Y}_r - Y^{n-1}}{\tilde{\rho}_r^{n-1}} \\ \frac{\tilde{Z}_r - Z^k}{\tilde{\rho}_r^k} & \frac{\tilde{Z}_r - Z^h}{\tilde{\rho}_r^h} & \cdots & \frac{\tilde{Z}_r - Z^{n-1}}{\tilde{\rho}_r^{n-1}} \end{bmatrix}$$

$$Y_0 = \begin{bmatrix} P_{rm}^{kh} \\ \vdots \end{bmatrix} \quad b = \begin{bmatrix} \vdots \end{bmatrix}$$

Note that the noise \(\nu_{rm}^{kh}\) of the double-difference observations is correlated. For example let us consider three satellites \(k, h, l\), where \(k\) is the pivot; the two possible double differences will be characterized by the following noise

$$\nu_{rm}^{kh} = (\nu_r^k - \nu_m^k) - (\nu_r^h - \nu_m^h)$$

$$\nu_{rm}^{kl} = (\nu_r^k - \nu_m^k) - (\nu_r^l - \nu_m^l)$$

Assuming that the original phase observations have a noise variance \(\sigma_0^2\) and that they will be uncorrelated to each other, that is

$$E(\nu_r^k, \nu_m^h) = \sigma_0^2 \delta_{kh} \delta_{lm}$$

then, it holds
In general the cofactor matrix $Q$ can be written as

$$Q = \begin{bmatrix} 4 & 2 & 2 & 2 \\ 2 & 4 & 2 & 2 \\ 2 & 2 & \ddots & \vdots \\ 2 & 2 & \cdots & 4 \end{bmatrix}$$

so that the corresponding noise covariance matrix is

$$C_{\nu\nu} = \sigma_0^2 Q$$

A least-squares solution based on double differences of the code observations allows to estimate the baseline between master and rover positions, and therefore to reconstruct the rover motion, with an error of the order of decimetres.

Let us move now to the phase observations:

$$\phi^k_m(t) = \frac{\rho^k_m}{\lambda} - N^k_m(t) + f[dt_m - dt^k_m(t)] - \frac{J^k_m}{\lambda} + \frac{T^k_m}{\lambda} + \nu^k_m$$

$$\phi^k_r(t) = \frac{\rho^k_r}{\lambda} - N^k_r(t) + f[dt_r(t) - dt^k_r(t)] - \frac{J^k_r}{\lambda} + \frac{T^k_r}{\lambda} + \nu^k_r$$

The single differences are given by:

$$\phi^k_{rm} = \phi^k_r - \phi^k_m = \left(\frac{\rho^k_r}{\lambda} - \frac{\rho^k_m}{\lambda}\right) - \left(N^k_r - N^k_m\right) + c(dt_r - dt_m) - \left(\frac{J^k_r}{\lambda} - \frac{J^k_m}{\lambda}\right) + \left(\frac{T^k_r}{\lambda} - \frac{T^k_m}{\lambda}\right) + (\nu^k_r - \nu^k_m)$$

while the double differences are:

$$\phi^{kh}_{rm} = \left(\phi^k_r - \phi^k_m\right) - \left(\phi^h_r - \phi^h_m\right) = \left[\left(\frac{\rho^k_r}{\lambda} - \frac{\rho^k_m}{\lambda}\right) - \left(\frac{\rho^h_r}{\lambda} - \frac{\rho^h_m}{\lambda}\right)\right] - \left[\left(N^k_r - N^k_m\right) - \left(N^h_r - N^h_m\right)\right] +$$

$$- \left[\left(\frac{J^k_r}{\lambda} - \frac{J^k_m}{\lambda}\right) - \left(\frac{J^h_r}{\lambda} - \frac{J^h_m}{\lambda}\right)\right] + \left[\left(\frac{T^k_r}{\lambda} - \frac{T^k_m}{\lambda}\right) - \left(\frac{T^h_r}{\lambda} - \frac{T^h_m}{\lambda}\right)\right] + \left[\left(\nu^k_r - \nu^k_m\right) - \left(\nu^h_r - \nu^h_m\right)\right] =$$

$$= \frac{\rho^{kh}_{rm}}{\lambda} - N^{kh}_{rm} + \frac{J^{kh}_{rm}}{\lambda} + \nu^{kh}_{rm}$$

Since the phase observations are affected by a noise at least one order of magnitude smaller than the code pseudo-range, it is expected that the estimate of the motion by using phase double differences is significantly better. Nevertheless, the phase double-difference observation equations contain a further unknown (the initial phase ambiguity) that is an integer number by definition. As already
stated before, the navigation solution, epoch by epoch, is not possible by using phase observations without a preliminary estimate of the initial phase ambiguity. A first solution to the problem is to use a “float” estimate (not integer), i.e. the following average in time:

\[
\hat{N}_{rm}^{kh} = \frac{1}{t_n-t_0} \sum_{t_{m_0}}^{t_n} \left[ \frac{P_{rm}^{kh}(t)}{\lambda} - \phi_{rm}^{kh}(t) \right]
\]

assuming that no cycle-slips occur in the considered time interval. Note that the used code-phase combination does not depend on the satellite-receiver distance (geometry free), while the dependence on the ionospheric delay \( J_{rm}^{kh} \) can be neglected if master and rover are not too far from each other. However, the standard deviation of this combination is quite high, approximately equal to code error expressed in cycles, namely

\[
\sigma\left( \frac{P}{\lambda} - \phi \right) \approx 3 \text{ cycles}
\]

Therefore the use of a single epoch to estimate \( N \) not only would lead to a non-integer number, but in a reasonable confidence interval, say \( 4\sigma \), many integer numbers could be taken as estimates of \( N \). On the contrary, by using the time average proposed before the corresponding error standard deviation is

\[
\sigma(\hat{N}) = \frac{\sigma\left( \frac{P}{\lambda} - \phi \right)}{\sqrt{t_n-t_0}}
\]

Assuming a sampling rate of one second and considering about 4 minutes (240 seconds) of observations without cycle-slips, we get

\[
\sigma(\hat{N}) \approx \frac{3}{15} = 0.2 \text{ cycles}
\]

so that only an integer number is included in the \( 4\sigma \) interval. This number is the estimate of \( N \).

When \( N \) is fixed, it is possible to linearize the phase observation equations, similarly to what has been done for the code observations. Then it is possible to estimate the rover coordinates by LS with a precision of the order of centimetres.

Note that to avoid cycle-slips for instance due to an obstacle between the satellite and the receiver, in a kinematic survey it is generally preferred keeping the rover still for about ten minutes before starting the survey. In this way the probability that a cycle-slip occurs is reduced and so it is possible to fix the initial phase ambiguities. Sometimes, in order to increase the accuracy of the estimated trajectory, a “stop-and-go” approach is used, which consists in periodically stopping the rover, thus exploiting the higher number of available observations for the stopping point.

Finally it has to be noted that the maximum distance of 10 km between master and rover, below which the atmospheric errors can be disregarded, can be significantly increased by using a network of permanent GPS stations.

**Differential GPS**

Similarly to relative positioning, differential GPS is based on the use of (at least) two receivers: a “master” that is located at a known position and a “rover” that is in motion. However, differently
from relative positioning, in the case of differential GPS the observations of the two receivers are not directly differentiated, but the known position of the master is used to compute corrections to be applied to the rover measurements.

The differential GPS technique was developed as a countermeasure to the so-called Selective Availability (SA). Originally, the GPS system was designed to offer civil users an accuracy in navigation of the order of 400 meters. Early tests on Block I satellites revealed the possibility to achieve an accuracy an order of magnitude better than expected. For this reason, U.S. Department of Defence introduced a degradation of the transmitted signal (called Selective Availability) on the Block II satellites starting from March 1990. This degradation was kept until May 2000 and since then it is not applied any longer.

Selective Availability consisted of two components: the former, called δ-process, affected the satellite clock stability, while the latter, called ε-process, corrupted the ephemerides transmitted by the satellites, namely the information on the satellite orbits. The resulting disturbance was of the order of 50-150 meters in terms of positioning on ground. The use of differential GPS allows to obviate the drawback of Selective Availability, obtaining an accuracy (at least in the horizontal positioning) of the order of 1-10 meters.

Although the Selective Availability is not active any more, differential positioning is still used with the aim of obtaining quite high accuracy (of the order of decimetres) even with low-level (cheap) receivers.

Differential GPS consists in applying corrections to the rover data by exploiting the known position of the master receiver. These corrections can be referred to either positions or measurements.

In the first approach, the master position is first computed by using the satellite observations and then compared to the nominal position (which is assumed to be known). On the basis of this comparison, the correction (translation) to be applied to the position estimated by the rover is computed. This concept is quite simple but requires that both master and rover use the same satellite configuration. This could be difficult, especially in case of large distance between master and rover and in case of local obstructions for the rover.

In the second approach (the one based on measurements), the satellite-master range observations are compared with the “expected” values derived from the master known coordinates and from the ephemerides transmitted by the satellites. This comparison gives rise to individual corrections for each satellite. These corrections are directly applied to the rover measurements and then the rover position is computed from the corrected measurements. By using this approach it is not necessary that master and rover see the same satellite configuration, because each correction is related to the single satellite. On the other hand, the computation of an individual correction for the single satellite requires a higher computational burden for the master station.

In general the corrections are transmitted from the master to the rover via a suitable telemetric link (for example via internet or via mobile phone network).

Let us concentrate now on corrections to measurements (the second approach discussed before) in the case of code observations. Naturally the differential GPS can be also applied to phase observations (in this case the method is called “precise DGPS”), but it finds applications only when a very high accuracy is required (for example in the airplane control). For the majority of common navigation applications, the use of code observations (or, in case, code observations smoothed by phase observations) is well sufficient.

The code pseudo-range of the master \( m \) can be written as

\[
P^s_m(t) = \rho^s_m(t) + \epsilon \left[ dt_m(t) - dt^s(t) \right] + J^s_m(t) + T^s_m(t) + v^s_m(t)
\]

where \( \rho^s_m(t) \) is the satellite-receiver distance with respect to the ephemerides transmitted by the satellite (possible orbital errors are included into the term \( v^s_m(t) \)). Anyway, note that the orbital
errors and the satellite clock delays have been significantly reduced after the dismissal of the Selective Availability, so that nowadays the main source of error is the ionospheric refraction.

The first step to derive the differential corrections consists in comparing the observed pseudo-range with the satellite-receiver distance derived from the knowledge of the master coordinates and from the ephemerides transmitted by the satellite. In other words, at the epoch \( t \), it holds

\[
PRC_m^s(\bar{t}) = \rho_m^s(\bar{t}) - P_m^s(\bar{t}) = -c\left[dt_m(\bar{t}) - dt^s(\bar{t})\right] - J_m^s(\bar{t}) - T_m^s(\bar{t}) - v_m^s(\bar{t})
\]

where the acronym PRC stays for Pseudo-Range Correction. Since the computation and the transmission of the PRC require some time, the differential corrections arrive at the rover with a slight delay. The effect of this latency can be reduced by extrapolating the PRC value at the measurement epoch \( t \) of the rover. In order to obtain a linear extrapolation (of first order) of the PRC, it is necessary to estimate the rate of change of the PRC, the so-called Range Rate Correction (RRC), which is computed by means of numerical differentiation:

\[
RRC_m^s(t) = \frac{dPRC_m^s(t)}{dt}
\]

The extrapolated value of PRC at the epoch \( t \) is therefore given by

\[
PRC_m^s(t) = PRC_m^s(\bar{t}) + RRC_m^s(\bar{t})(t - \bar{t})
\]

where \((t - \bar{t})\) is the latency of the differential corrections. Then the rover receiver applies these corrections to its pseudo-range measurements, namely

\[
P_r^s(t)_{corr} = P_r^s(t) + PRC_m^s(t)
\]

Assuming that the extrapolated correction at the epoch \( t \) has an analogous expression of the computed correction at the epoch \( \bar{t} \), it holds:

\[
P_r^s(t)_{corr} = \rho_r^s(t) + c\left[dt_r(t) - dt^s(t)\right] + J_r^s(t) + T_r^s(t) + v_r^s(t) - c\left[dt_m(t) - dt^s(t)\right] - J_m^s(t) - T_m^s(t) - v_m^s(t) = \\
= \rho_r^s(t) + c dt_{rm}(t) + J_{rm}^s(t) + T_{rm}^s(t) + v_{rm}^s(t)
\]

It is important that both the master and the rover use the same transmitted ephemerides, otherwise the common error terms do not cancel out. Note also that the atmospheric delays between satellite and master and the atmospheric delays between satellite and rover are strongly correlated for baselines up to hundreds of kilometres, therefore their effects are significantly reduced by differentiation. All in all, the previous relation can be simplified as

\[
P_r^s(t)_{corr} = \rho_r^s(t) + c dt_{rm}(t) + v_{rm}^s(t)
\]

where \( dt_{rm}(t) \) is the combination of the master and rover clock delays. The obtained equation has the same structure of the equation used in the point-positioning, with the advantage that the atmospheric bias are reduced due to the differentiation. Again, the combination of the master and rover clock delays has to be considered as an additional unknown in the least-squares solution. This means that a simultaneous observation of at least four common satellites at the master and rover locations is required.
It is important to clarify the difference between differential and relative positioning. In the first case corrections extrapolated at the epoch $t$ are applied to the rover measurements with the aim of computing the rover absolute position; in the second case, simultaneous measurements performed by the master and rover receivers are used with the aim of computing the master-rover baseline vector (namely the difference between the two absolute positions). Traditionally, this latter procedure was accomplished in post-processing, i.e. after the survey. With the introduction of the Real-Time Kinematic (RTK) technique, master measurements (typically phase measurements) are directly transmitted to the rover. Differently from DGPS, no correction or extrapolation is performed, but the difference of the two observations at time $\tilde{t}$ is processed to determine the rover relative position. As a consequence, the trajectory is computed with a certain latency $(t - \tilde{t})$, but with a position accuracy one order of magnitude better than in the case of DGPS. Note that, from the theoretical point of view, differential and relative positioning coincide if the latency becomes zero.

Differential GPS can be based on the concept of single or multiple reference station. The system architecture with a single reference station is the one considered up to now: it is a very simple concept but it has the disadvantage that the positioning accuracy decreases with increasing distance between the two receivers, since the errors are less spatially correlated. In the multi-station approach, many reference receivers (i.e. many masters) are used to generate the corrections. The main advantage of this solution is that the area covered by the service is larger, with a quite homogeneous positioning accuracy. The disadvantages are represented by the increased costs for installation and maintenance and by the additional latency in the computation of differential corrections due to the communication among several masters. Within the multi-station approach, two different techniques are possible for the application of the differential corrections. The first technique (denoted as local-area differential GNSS, LAD GNSS) consists in applying to the rover the corrections of the nearest master station or a weighted average of the corrections of all master stations. Obviously this solution requires a high density of master stations in the area covered by the service and, for this reason, it is mainly applied to local problem. The second technique (denoted as wide-area differential GNSS, WAD GNSS) consists in computing a regular grid of ionospheric corrections over the area of interest. The specific corrections for each line-of-sight between a satellite and the rover location are derived by a proper interpolation of the grid values. Double-frequency receivers are required for the master stations, but their density can be lower than in the case of LAD GNSS.

References


