Surveying 2

By Dr. Khamis Naba Sayl

المحاضرة الثامنة

Global position system (GPS)

Global position system (GPS)

1-INTRODUCTION

During the 1970s, a new and unique approach to surveying, the global positioning system (GPS), emerged. This system, which grew out of the space program, relies upon signals transmitted from satellites for its operation. It has resulted from research and development paid for by the military to produce a system for global navigation and guidance. More recently other countries are developing their own systems. Thus, the entire scope of satellite systems used in positioning is now referred to as global navigation satellite systems (GNSS). Receivers that use GPS satellites and another system such as GLONASS. These systems provide precise timing and positioning information anywhere on the Earth with high reliability and low cost. The systems can be operated day or night, rain or shine, and do not require cleared lines of sight between survey stations. This represents a revolutionary departure from conventional surveying procedures, which rely on observed angles and distances for determining point positions. Since these systems all share similar features, the global positioning system will be discussed in further detail herein.

2- OVERVIEW OF GPS

Precise distances from the satellites to the receivers are determined from timing and signal information, enabling receiver positions to be computed. In satellite surveying, the satellites become the reference or control stations, and the ranges (distances) to these satellites are used to compute the positions of the receiver. Conceptually, this is equivalent to resection in traditional ground surveying work, where distances and/or angles are observed from an unknown ground station to control points of known position.

The global positioning system can be arbitrarily broken into three parts: (a) the space segment, (b) the control segment, and (c) the user segment. The space segment consists nominally of 24 satellites operating in six orbital planes spaced at 60° intervals around the equator. Four additional satellites are held in reserve as spares. The orbital planes are inclined to the equator at 55°.

This configuration provides 24-h satellite coverage between the latitudes of 80°N and 80°S. The satellites travel in near-circular orbits that have a mean altitude of 20,200 km above the Earth and an orbital period of 12 sidereal hours.1 The individual satellites are normally identified by their Pseudo Random Noise (PRN) number, (described below), but can also be identified by their satellite vehicle number (SVN) or orbital position.

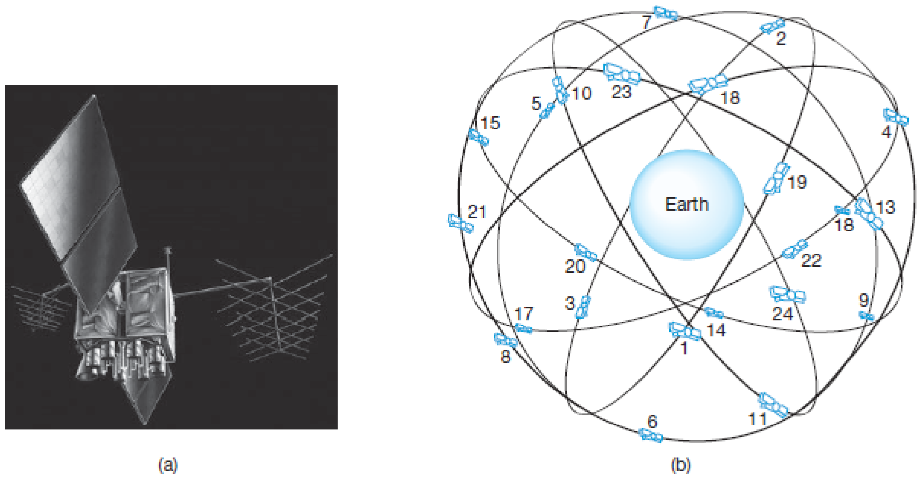

Figure 13.2 (a) A GPS satellite and (b) the GPS constellation.

The control segment consists of monitoring stations which monitor the signals and track the positions of the satellites over time. The initial GPS monitoring stations are at Colorado Springs, and on the islands of Hawaii, Ascension, Diego Garcia, and Kwajalein. The tracking information is relayed to the master control station in the Consolidated Space Operations Center (CSOC) located at Schriever Air Force base in Colorado Springs. The master control station uses this data to make precise, near-future predictions of the satellite orbits, and their clock correction parameters. This information is uploaded to the satellites, and in turn, transmitted by them as part of their broadcast message to be used by receivers to predict satellite positions and their clock biases (systematic errors).

The user segment in GPS consists of two categories of receivers that are classified by their access to two services that the system provides. These services are referred to as the Standard Position Service (SPS) and the Precise Positioning Service (PPS). The SPS is provided on the L1 broadcast frequency and more recently the L2 at no cost to the user. This service was initially intended to provide accuracies of 100 m in horizontal positions, and 156 m in vertical positions at the 95% error level. However, improvements in the system and the processing software have substantially reduced these error estimates. The PPS is broadcast on both the L1 and L2 frequencies, and is only available to receivers having valid cryptographic keys, which are reserved almost entirely for DoD use. This message provides a published accuracy of 18 m in the horizontal, and 28 m in the vertical at the 95% error level.

3-BASIC PRINCIPLE OF POSITION FIXING

Position fixing in three dimensions may involve the measurement of distance (or range) to at least three satellites whose X, Y and Z position is known, in order to define the user's Xp, Yp and Zp position. In its simplest form, the satellite transmits a signal on which the time of its departure (t_D) from the satellite is modulated. The receiver in turn notes the time of arrival (t_A) of this time mark. Then the time which it took the signal to go from satellite to receiver is

$(t_A - t_D) = t$ called the delay time. The measured range R is obtained from

$R_1 = (t_A - t_D)c = \Delta tc$

where c = the velocity of light.

Whilst the above describes the basic principle of range measurement, to achieve it one would require the receiver to have a clock as accurate as the satellite's and perfectly synchronized with it. As this would render the receiver impossibly expensive, a correlation procedure, using the pseudo-random binary codes (P or C/A), usually 'C/A', is adopted. The signal from the satellite arrives at the receiver and triggers the receiver to commence generating its own internal copy of the C/A code. The receiver-generated code is cross-correlated with the satellite code (Figure 9.10). The ground receiver is then able to determine the time delay (t) since it generated the same portion of the code received from the satellite. However, whilst this eliminates the problem of the need for an expensive receiver clock, it does not eliminate the problem of exact synchronization of the two clocks. Thus, the time difference between the two clocks, termed clock bias, results in an incorrect assessment of t. The distances computed are therefore called 'pseudo-ranges'.

The use of four satellites rather than three, however, can eliminate the effect of clock bias. A line in space is defined by its difference in coordinates in an X, Y and Z system:

$$R = (\Delta X^2 + \Delta Y^2 + \Delta Z^2)^{\frac{1}{2}}$$

If the error in R, due to clock bias, is δR and is constant throughout, then:

$$R_1 + \delta R = [(X_1 - X_p)^2 + (Y_1 - Y_p)^2 + (Z_1 - Z_p)^2]^{\frac{1}{2}}$$

$$R_2 + \delta R = [(X_2 - X_p)^2 + (Y_2 - Y_p)^2 + (Z_2 - Z_p)^2]^{\frac{1}{2}}$$

$$R_3 + \delta R = [(X_3 - X_p)^2 + (Y_3 - Y_p)^2 + (Z_3 - Z_p)^2]^{\frac{1}{2}}$$

$$R_4 + \delta R = [(X_4 - X_p)^2 + (Y_4 - Y_p)^2 + (Z_4 - Z_p)^2]^{\frac{1}{2}}$$

where X_n, Y_n, Z_n = the coordinates of satellites 1, 2, 3 and 4 ($n = 1$ to 4)
X_p, Y_p, Z_p = the coordinates required for point P
R_n = the measured ranges to the satellites

Solving the four equations for the four unknowns X_p, Y_p, Z_p and δR also solves for the error due to clock bias.

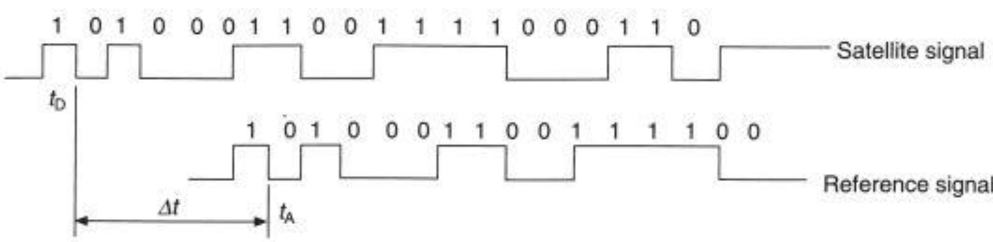

Fig. 9.10 *Correlation of the pseudo-binary codes*

REFERENCE COORDINATE SYSTEMS

In determining the positions of points on Earth from satellite observations, three different reference coordinate systems are important. First of all, satellite positions at the instant they are observed are specified in the “space-related” satellite reference coordinate systems. These are three-dimensional rectangular systems defined by the satellite orbits. Satellite positions are then transformed into a three-dimensional rectangular geocentric coordinate system, which is physically related to the Earth. As a result of satellite positioning observations, the positions of new points on Earth are determined in this coordinate system. Finally, the geocentric coordinates are transformed into the more commonly used and locally oriented geodetic coordinate system. The following subsections describe these three coordinate systems.

The Geodetic Coordinate System

Although the positions of points in a satellite survey are computed in the geocentric coordinate system, in that form they are inconvenient for use by surveyors (geomatics engineers). This is the case for three reasons: (1) with their origin at the Earth’s center, geocentric coordinates are typically extremely large values, (2) with the X-Y plane in the plane of the equator, the axes are unrelated to the conventional directions of north-south or east-west on the surface of the Earth, and (3) geocentric coordinates give no indication about relative elevations between points. For these reasons, the geocentric coordinates are converted to geodetic coordinates of latitude ϕ longitude λ and height (h) so that reported point positions become more meaningful and convenient for users.

Figure 13.6 also illustrates a quadrant of the reference ellipsoid, and shows both the geocentric coordinate system (X,Y,Z), and the geodetic coordinate system (ϕ, λ, h). Conversions from geocentric to geodetic coordinates, and vice versa are readily made. From the figure it can be shown that geocentric coordinates of point P can be computed from its geodetic coordinates using the following equations:

$$X_P = (R_{N_P} + h_P)\cos\phi_P \cos\lambda_P$$
$$Y_P = (R_{N_P} + h_P)\cos\phi_P \sin\lambda_P \qquad (13.1)$$
$$Z_P = [R_{N_P}(1 - e^2) + h_P]\sin\phi_P$$

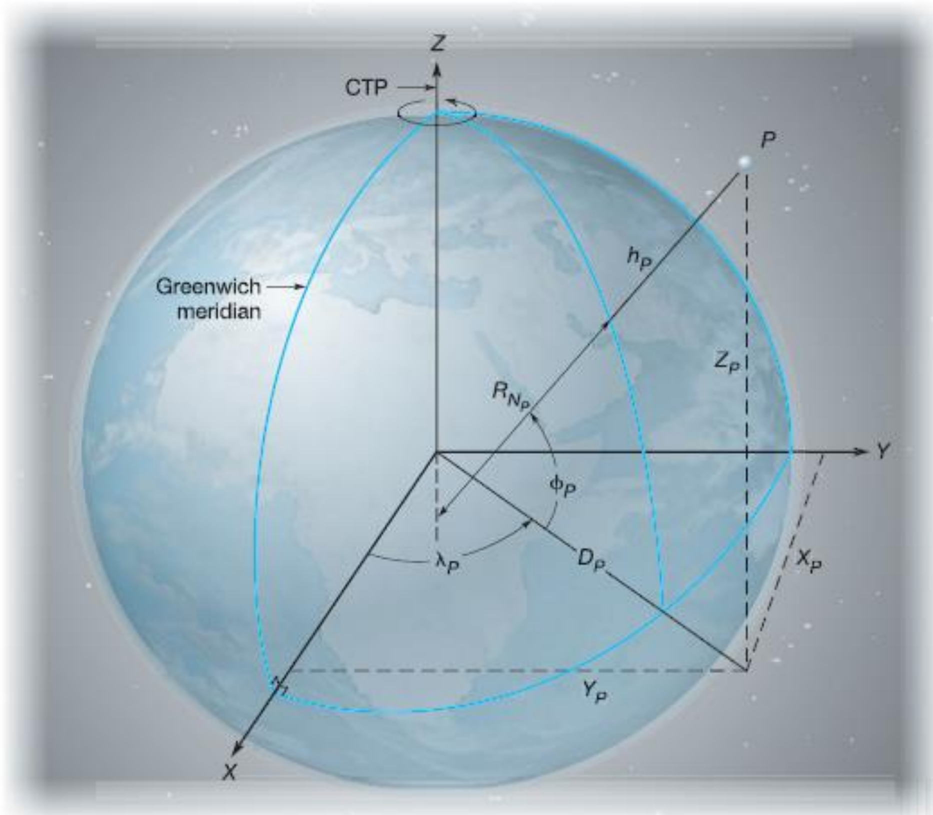

Figure 13.6
The geodetic and geocentric coordinate systems.

where

$$R_{N_P} = \frac{a}{\sqrt{1 - e^2 \sin^2 \phi_P}} \qquad (13.2)$$

In Equations (13.1), X_p , Y_p and Z_p are the geocentric coordinates of any point P, and the term e, which appears in both Equations (13.1) and (13.2), is the eccentricity of the WGS84 reference ellipsoid. Its value is 0.08181919084. In Equation (13.2), R_{NP} is the radius in the prime vertical of the ellipsoid at point P, and a, as noted earlier, is the semimajor axis of the ellipsoid. In Equations (13.1) and (13.2), north latitudes are considered positive and south latitudes negative.

Similarly, east longitudes are considered positive and west longitudes negative. Additionally, the programming for the conversion of geodetic coordinates to geocentric coordinates and vice versa is demonstrated.

Example 13.1

The geodetic latitude, longitude, and height of a point A are 41°15′18.2106″N, 75°00′58.6127″ W, and 312.391 m, respectively. Using WGS84 values, what are the geocentric coordinates of the point?

Solution

Substituting the appropriate values into Equations (13.1) and (13.2) yields

$$R_{N_A} = \frac{6{,}378{,}137}{\sqrt{1 - 0.0066943799 \sin^2(41°15'18.2106'')}} = 6{,}387{,}440.3113 \text{ m}$$

$$X_A = (6{,}387{,}440.3113 + 312.391) \cos 41°15'18.2106'' \cos(-75°00'58.6127'')$$
$$= 1{,}241{,}581.343 \text{ m}$$

$$Y_A = (6{,}387{,}440.3113 + 312.391) \cos 41°15'18.2106'' \sin(-75°00'58.6127'')$$
$$= -4{,}638{,}917.074 \text{ m}$$

$$Z_A = [6{,}387{,}440.3113(1 - 0.00669437999) + 312.391)] \sin(41°15'18.2106'')$$
$$= 4{,}183{,}965.568 \text{ m}$$

Conversion of geocentric coordinates of any point P to its geodetic values is accomplished using the following steps (refer again to Figure 13.6).

Step 1: Compute D_P as

$$D_P = \sqrt{X_P^2 + Y_P^2} \tag{13.3}$$

Step 2: Compute the longitude as[5]

$$\lambda_P = 2\tan^{-1}\left(\frac{D_P - X_P}{Y_P}\right) \tag{13.4}$$

Step 3: Calculate approximate latitude, ϕ_0[6]

$$\phi_0 = \tan^{-1}\left[\frac{Z_P}{D_P(1 - e^2)}\right] \tag{13.5}$$

Step 4: Calculate the approximate radius of the prime vertical, R_N, using ϕ_0 from step 3, and Equation (13.2).

Step 5: Calculate an improved value for the latitude from

$$\phi = \tan^{-1}\left(\frac{Z_P + e^2 R_{N_P}\sin(\phi_0)}{D_P}\right) \tag{13.6}$$

Step 6: Repeat the computations of steps 4 and 5 until the change in ϕ between iterations becomes negligible. This final value, ϕ_P, is the latitude of the station.

Step 7: Use the following formulas to compute the geodetic height of the station. For latitudes less than 45°, use

$$h_P = \frac{D_P}{\cos(\phi_P)} - R_{N_P} \tag{13.7a}$$

For latitudes greater than 45° use the formula

$$h_P = \left[\frac{Z_P}{\sin(\phi_P)}\right] - R_{N_P}(1 - e^2) \tag{13.7b}$$

Example 13.2

What are the geodetic coordinates of a point that has X, Y, Z geocentric coordinates of 1,241,581.343, –4,638,917.074, and 4,183,965.568, respectively? (Note: Units are meters.)

Solution

To visualize the solution, refer to Figure 13.6. Since the X coordinate value is positive, the longitude of the point is between 0° and 90°. Also, since the Y coordinate value is negative, the point is in the western hemisphere. Similarly since the Z coordinate value is positive, the point is in the northern hemisphere.

Substituting the appropriate values into Equations (13.3) through (13.7) yields

Step 1:

$$D = \sqrt{(1{,}241{,}581.343)^2 + (-4{,}638{,}917.074)^2} = 4{,}802{,}194.8993$$

Step 2:

$$\lambda = 2\tan^{-1}\left(\frac{4{,}802{,}194.8993 - 1{,}241{,}581.343}{-4{,}638{,}917.074}\right) = -75°00'58.6127'' \text{ (West)}$$

Step 3:

$$\phi_0 = \tan^{-1}\left[\frac{4{,}183{,}965.568}{4{,}802{,}194.8993(1 - 0.00669437999)}\right] = 41°15'18.2443''$$

Step 4:

$$R_N = \frac{6{,}378{,}137}{\sqrt{1 - 0.00669437999 \sin^2(41°15'18.2443'')}} = 6{,}387{,}440.3148$$

Step 5:

$$\phi_0 = \tan^{-1}\left[\frac{4{,}183{,}965.568 + e^2 6{,}387{,}440.3148 \sin 41°15'18.2443''}{4{,}802{,}194.8993}\right]$$

$$= 41°15'18.2107''$$

Step 6: Repeat steps 4 and 5 until the latitude converges. The values for the next iteration are

$$R_N = 6{,}387{,}440.3113$$
$$\phi_0 = 41°15'18.2106''$$

Repeating with the above values results in the same value for latitude to four decimal places, so the latitude of the station is 41°15′18.2106″ N.

Step 7: Compute the geodetic height using Equation (13.7a) as

$$h = \frac{4{,}802{,}194.8993}{\cos 41°15'18.2106''} - 6{,}387{,}440.3113 = 312.391$$

The geodetic coordinates of the station are latitude = 41°15′18.2106″ N, longitude = 75°00′58.6127″ W, and height = 312.391 m. Note that this example was the reverse computations of Example 13.1, and it reproduced the starting geodetic coordinate values for that example.

It is important to note that geodetic heights obtained with satellite surveys are measured with respect to the ellipsoid. That is, the geodetic height of a point is the vertical distance between the ellipsoid and the point as illustrated in Figure 13.7. As shown, these are not equivalent to elevations (also called orthometric heights) given with respect to the geoid. To convert geodetic heights to elevations, the geoid height (vertical distance between ellipsoid and geoid) must be known. Then elevations can be expressed as:

$$H = h - N \tag{13.8}$$

where H is elevation above the geoid (orthometric height), h the geodetic height (determined from satellite surveys), and N the geoidal height. Figure 13.7 shows the correct relationship of the geoid and the WGS84 ellipsoid in the continental United States. Here the ellipsoid is above the geoid, and geoid height (measured from the ellipsoid) is negative. The geoid height at any point can be estimated with mathematical models developed by combining gravimetric data with distributed networks of points where geoidal height has been observed.

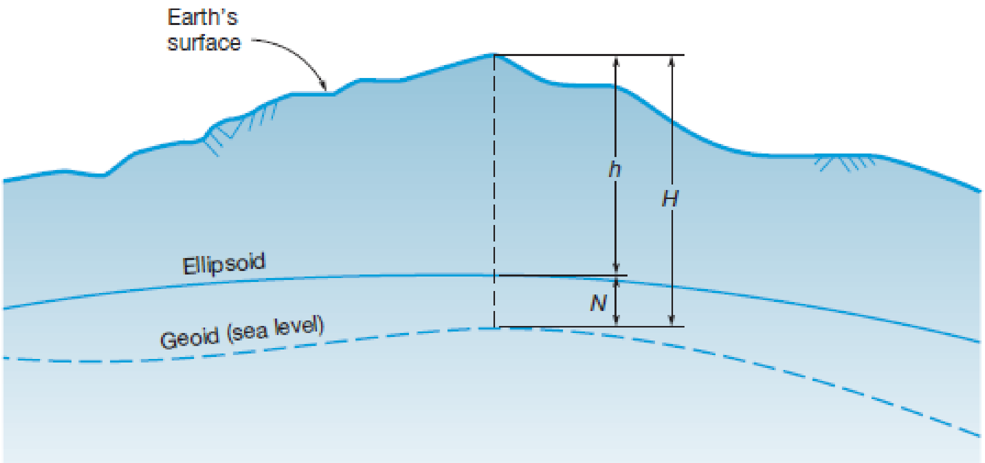

Figure 13.7 Relationships between elevation H, geodetic height h, and geoid undulation N.