1 Introduction

1.1 The origins of surveying

Since the dawn of civilization, man has looked to the heavens with awe searching for portentous signs. Some of these men became experts in deciphering the mystery of the stars and developed rules for governing life based upon their placement. The exact time to plant the crops was one of the events that was foretold by the early priest astronomers who in essence were the world's first surveyors. Today, it is known that the alignment of such structures as the pyramids and Stonehenge was accomplished by celestial observations and that the structures themselves were used to measure the time of celestial events such as the vernal equinox.

Some of the first known surveyors were Egyptian surveyors who used distant control points to replace property corners destroyed by the flooding Nile River. Later, the Greeks and Romans surveyed their settlements. The Dutch surveyor Snell van Royen was the first who measured the interior angles of a series of interconnecting triangles in combination with baselines to determine the coordinates of points long distances apart. Triangulations on a larger scale were conducted by the French surveyors Picard and Cassini to determine a baseline extending from Dunkirk to Collioure. The triangulation technique was subsequently used by surveyors as the main means of determining accurate coordinates over continental distances.

The chain of technical developments from the early astronomical surveyors to the present satellite geodesists reflects man's desire to be able to master time and space and to use science to further his society. The surveyor's role in society has remained unchanged from the earliest days; that is to determine land boundaries, provide maps of his environment, and control the construction of public works.

1.2 Development of global surveying techniques

The use of triangulation (later combined with trilateration and traversing) was limited by the line of sight. Surveyors climbed to mountain tops and developed special survey towers to extend this line of sight usually by small amounts. The series of triangles was generally oriented or fixed by astronomic points where special surveyors had observed selected stars to determine the position of that point on the surface of the earth. Since these oped during World War II to position aircraft. Beginning in the late 1940s, HIRAN arcs of trilateration were measured between North America and Europe in an attempt to determine the difference in their respective datums. A significant technological breakthrough occurred in 1957 after the launch of Sputnik, the first artificial satellite, when scientists around the world experienced that the Doppler shift in the signal broadcast by a satellite could be used as an observable to determine the exact time of closest approach of the satellite. This knowledge, together with the ability to compute satellite ephemerides according to Kepler's laws, led to the present capability of instantaneously determining precise position anywhere in the world.

The immediate predecessor of today's modern positioning system is the Navy Navigation Satellite System (NNSS), also called TRANSIT system. This system was composed of six satellites orbiting at altitudes of about 1 100 km with nearly circular polar orbits. The TRANSIT system was developed by the U.S. military, primarily, to determine the coordinates of vessels and aircraft. Civilian use of this satellite system was eventually authorized, and the system became widely used worldwide both for navigation and surveying.

Some of the early TRANSIT experiments by the former U.S. Defense Mapping Agency (DMA) and the U.S. Coast & Geodetic Survey showed that accuracies of about one meter could be obtained by occupying a point for several days and reducing the observations using the postprocessed precise ephemerides. Groups of Doppler receivers in translocation mode could also be used to determine the relative coordinates of points to submeter accuracy using the broadcast ephemerides. This system employed essentially the same Doppler observable used to track the Sputnik satellite; however, the orbits of the TRANSIT satellites were precisely determined by tracking them at widely spaced fixed sites.

1.3 History of the Global Positioning System

The Global Positioning System (GPS) was developed to replace the TRANSIT system because of two major shortcomings in the earlier system. The large time gaps in coverage were the main problem with TRANSIT. Since nominally a satellite passed overhead every 90 minutes, users had to interpolate their position between "fixes" or passes. The second problem with the TRANSIT system was its relatively low navigation accuracy.

In contrast, GPS answers the questions "What time, what position, and what velocity is it?" quickly, accurately, and inexpensively anywhere on the globe at any time.

astronomic positions could be in error by hundreds of meters, each continent was virtually (positionally) isolated and their interrelationship was imprecisely known.

1.2.1 Optical global triangulation

Some of the first attempts to determine the interrelationship between the continents were made using the occultation of certain stars by the moon. This method was cumbersome at best and was not particularly successful. The launch of the Russian Sputnik satellite in 1957, however, had tremendously advanced the connection of the various geodetic world datums. In the beginning of the era of artificial satellites, an optical method, based (in principle) on the stellar triangulation method developed in Finland as early as 1946, was applied very successfully. The worldwide satellite triangulation program, often called the BC-4 program (after the camera that was used), for the first time determined the interrelationships of the major world datums. This method involved photographing special reflective satellites against a star background with a metric camera that was fitted with a specially manufactured chopping shutter. The image that appeared on the photograph consisted of a series of dots depicting each star's path and a series of dots depicting the satellite's path. The coordinates of selected dots were precisely measured using a photogrammetric comparator, and the associated spatial directions from the observing site to the satellite were then processed using an analytical photogrammetric model. Photographing the same satellite from a neighboring site simultaneously and processing the data in an analogous way yields another set of spatial directions. Each pair of corresponding directions forms a plane containing the observing points and the satellite, and the intersection of at least two planes results in the spatial direction between the observing sites. In the next step, these oriented directions were used to construct a global network with the scale being derived from several terrestrial traverses. An example is the European baseline running from Tromsø in Norway to Catania on Sicily. The main problem in using this optical technique was that clear sky was required simultaneously at a minimum of two observing sites separated by some 4 000 km, and the equipment was massive and expensive. Thus, optical direction measurement was soon supplanted by the electromagnetic ranging technique because of all-weather capability, greater accuracy, and lower cost of the newer technique.

1.2.2 Electromagnetic global trilateration

First attempts to (positionally) connect the continents by electromagnetic techniques was by the use of an electronic ranging system (HIRAN) devel-

1.3.1 Navigating with GPS

One aim of navigation is instantaneous positioning and velocity determination. As stated, one of the main problems with the TRANSIT system was the fact that it was not able to provide continuous positioning.

Satellite constellation
To provide a continuous global positioning capability, a scheme to orbit a sufficient number of satellites to ensure that (at least) four were always electronically visible was developed for GPS. Several schemes were proposed and it was found that 24 evenly spaced satellites placed in circular 12-hour orbits inclined 55° to the equatorial plane would provide the desired coverage for the least expense. In any event, this constellation provides the minimum of four satellites in good geometric position 24 hours per day anywhere on the earth. There are often more than the minimum number of satellites available for use and it is during these periods that surveyors perform special surveys. In fact, assuming a 10° elevation angle, there are brief periods where up to 10 GPS satellites are visible on the earth.

Point positioning
The GPS satellites are configured, primarily, to provide the user with the capability of determining his position, expressed for example by latitude, longitude, and elevation. This is accomplished by the simple resection process using the distances measured to satellites.

Consider the satellites frozen in space at a given instant. The space coordinates $\underline{\varrho}^S$ relative to the center of the earth of each satellite (see also Fig. 1.1) can be computed from the ephemerides broadcast by the satellite by an algorithm presented in Chap. 4. If the ground receiver defined by its geocentric position vector $\underline{\varrho}_R$ employed a clock that was set precisely to GPS system time (Sect. 3.3) the true distance or range ϱ to each satellite could be accurately measured by recording the time required for the (coded) satellite signal to reach the receiver. Each range defines a sphere (more precisely: surface of a sphere) with its center at the satellite. Hence, using this technique, ranges to only three satellites would be needed since the intersection of three spheres yields the three unknowns (e.g., latitude, longitude, and height) which could be determined from the three range equations

$$\varrho = \|\underline{\varrho}^S - \underline{\varrho}_R\| \,. \tag{1.1}$$

GPS receivers apply a slightly different technique. They typically use an inexpensive crystal clock which is set approximately to GPS time. Thus, the clock of the ground receiver is offset from true GPS time, and because

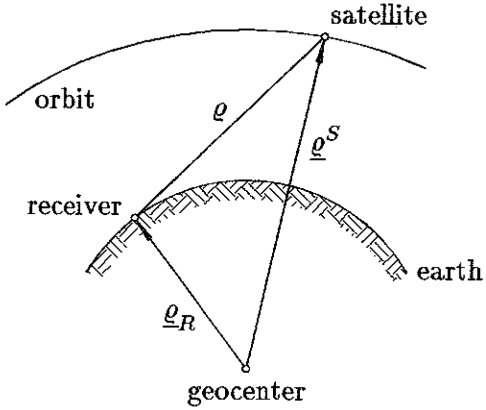

Fig. 1.1. Principle of satellite positioning

of this offset, the distance measured to the satellite differs from the "true" range. These distances are called pseudoranges R since they are the true range plus a range correction $\Delta\varrho$ resulting from the receiver clock error or bias δ. A simple model for the pseudorange is

$$R = \varrho + \Delta\varrho = \varrho + c\,\delta \tag{1.2}$$

with c being the velocity of light.

Four simultaneously measured pseudoranges are needed to solve for the four unknowns; these are three components of position plus the clock bias. Geometrically, the solution is accomplished by a sphere being tangent to the four spheres defined by the pseudoranges. The center of this sphere corresponds to the unknown position and its radius equals the range correction.

It is worth noting that the range error $\Delta\varrho$ could be eliminated in advance by differencing the pseudoranges measured from one site to two satellites or two different positions of one satellite. In the second case, the resulting range difference or delta range corresponds to the observable in the TRANSIT system. In both cases, the delta range defines a hyperboloid with its foci placed at the two satellites or the two different satellite positions for the geometric location of the receiver.

Considering the fundamental equation (1.1), one can conclude that the accuracy of the position determined using a single receiver essentially is affected by the following factors:

- accuracy of each satellite position,
- accuracy of pseudorange measurement,
- geometry.

Systematic errors in the satellite position and eventual satellite clock biases in the pseudoranges can be reduced or eliminated by differencing the pseudoranges measured from two sites to the satellite. This interferometric approach has become fundamental for GPS surveying as demonstrated below. However, no mode of differencing can overcome poor geometry.

A measure of satellite geometry with respect to the observing site is a factor known as Geometric Dilution of Precision (GDOP). In a geometric approach, this factor is inversely proportional to the volume of a body. This body is formed by points obtained from the intersection of a unit sphere with the vectors pointing from the observing site to the satellites. More details and an analytical approach on this subject are provided in Sect. 9.6.

Velocity determination
The determination of the instantaneous velocity of a moving vehicle is another goal of navigation. This can be achieved by using the Doppler principle of radio signals. Because of the relative motion of the GPS satellites with respect to a moving vehicle, the frequency of a signal broadcast by the satellites is shifted when received at the vehicle. This measurable Doppler shift is proportional to the relative radial velocity. Since the radial velocity of the satellites is known, the radial velocity of the moving vehicle can be deduced from the Doppler observable. A minimum of four Doppler observables is required to solve for the three components of the vehicle's velocity vector and, possibly, one frequency bias.

In summary, GPS was designed to solve many of the problems inherent to the TRANSIT system. Above all, GPS provides 24 hours a day instantaneous global navigation. The system as originally designed, however, did not include provision for the accurate surveying that is performed today. This surveying use of GPS resulted from a number of fortuitous developments described below.

1.3.2 Surveying with GPS

From navigation to surveying
As previously described, the use of near-earth satellites for navigation was demonstrated by the TRANSIT system. In 1964, I. Smith filed a patent describing a satellite system that would emit time codes and radio waves that would be received on earth as time delayed transmissions creating hyperbolic lines of position. This concept would become important in the treatment of GPS observables to compute precise vectors. A few years later, another patent was filed by R. Easton further refining the concept of comparing the

phase from two or more satellites.

In 1972, C. Counselman along with colleagues from the Massachusetts Institute of Technology (MIT) reported on the first use of interferometry to track the Apollo 16 Lunar Rover module. The principle they described is in essence the same technique they used later in developing the first geodetic GPS receiver and corresponds to differencing pseudoranges measured from two receivers to one satellite. The present use of the GPS carrier phase to make millimeter vector measurements dates from work by the MIT group using Very Long Baseline Interferometry (VLBI) performed between 1976 and 1978 where they proved that (for short lines) millimeter accuracy was obtainable using the interferometric technique.

In 1978, the Miniature Interferometer Terminals for Earth Surveying (MITES) were proposed. They detail how a satellite system can be used for precise surveying. This concept was further refined to include the Navigation System with Timing and Ranging (NAVSTAR). Also, the codeless technique that later became important in developing high-accuracy dual frequency receivers was described for the first time.

Observation techniques

When referring to high accuracy, GPS surveying implies the precise measurement of the vector between two (or more) GPS receivers.

The observation technique where both receivers involved remain fixed in position is called static surveying. The static method formerly required hours of observation and was the technique that was primarily used for early GPS surveys. A second technique where one receiver remains fixed, while the second receiver moves is called kinematic surveying. In 1986, B. Remondi first demonstrated that subcentimeter vector accuracies could be obtained between a pair of GPS survey instruments with as little as a few seconds of data collection using this method. B. Remondi also first developed another survey technique which is denoted as pseudokinematic technique. In this technique, a pair of receivers occupies a pair of points for two brief periods that are separated in time. This method, also denoted intermittent static, snapshot static, or reoccupation has demonstrated accuracies comparable to the static method.

Data processing is performed in two conceptually different ways. (1) Relative positioning is the technique where simultaneously observed data are processed; strictly speaking, the result is not obtainable instantaneously due to the required simultaneity. (2) The differential positioning technique involves placing one receiver at a fixed site of known position. Comparing computed ranges with measured pseudoranges, the reference site can transmit corrections to a roving receiver to improve its measured pseudoranges.

This technique provides instantaneous positioning and is, thus, mainly applied in kinematic surveys.

Hardware developments
The following sections contain reference to various terms that are more fully described in subsequent chapters. These are the C/A-code (Coarse/Acquisition) and P-code (Precision) which are basically code bits that are modulated on the two carrier signals broadcast by the GPS satellites. Code correlation as well as codeless techniques strip these codes from the carrier so that the phase of the (reconstructed) carrier can be measured. Brand names mentioned in this section are included for historical purposes since they represent the first of a certain class or type of receiver.

An interferometric technology for codeless pseudoranging was developed by P. MacDoran at the Jet Propulsion Laboratory (JPL). This Satellite Emission Range Inferred Earth Surveying (SERIES) technique was later improved for commercial geodetic applications. The culmination of the VLBI interferometric research applied to earth orbiting satellites was the production of a "portable" codeless GPS receiver that could measure short baselines to millimeter accuracy and long baselines to one part per million (ppm). This performance of the receiver, trade-named the Macrometer Interferometric Surveyor™, was demonstrated by the former U.S. Federal Geodetic Control Committee (FGCC).

A parallel development was being carried out by the DMA in cooperation with the U.S. National Geodetic Survey (NGS) and the U.S. Geological Survey (USGS). In 1981, these agencies developed specifications for a portable dual frequency code correlating receiver that could be used for precise surveying and point positioning. Texas Instruments Company was awarded the contract to produce a receiver later trade-named the TI-4100. The NGS geodesists C. Goad and B. Remondi developed software to process its carrier phase data in a manner similar to the method used by the MIT group (i.e., interferometrically).

The physical characteristics of the TI-4100 were significantly different from the Macrometer. The TI-4100 was a dual frequency receiver that used the P-code to track a maximum of four satellites, while the original Macrometer was a rack mounted codeless single frequency receiver that simultaneously tracked up to six satellites. There were also significant logistic differences in performing surveys using these two pioneer instruments. The TI-4100 received the broadcast ephemerides and timing signals from the GPS satellites so units could be operated independently, while the Macrometer required that all units be brought together prior to the survey and after the survey so that the time of the units could be synchronized. Also, the Macrometer required that the ephemerides for each day's tracking be generated at the home office prior to each day's observing session.

The next major development in GPS surveying occurred in 1985 when manufacturers started to produce C/A-code receivers that measured and output the carrier phase. The first of this class of receivers was trade-named the Trimble 4000S. This receiver required the data to be collected on an external (i.e., laptop) computer. The 4000S was the first of the generic C/A-code receivers that eventually were produced by a host of manufacturers. The first Trimble receivers were sold without processing software; however, the company soon retained the services of C. Goad who produced appropriate vector computation software which set the standard for future software developers.

Today's GPS receivers include all features of the early models and additionally have expanded capabilities. By far the major portion of receivers produced today are the C/A-code single frequency type. For precise geodetic work, however, the more expensive dual frequency receivers are the standard.

Software developments

The development of GPS surveying software has largely paralleled the development of hardware.

The NGS has been one of the primary organizations in the world in developing independent GPS processing software. As previously mentioned, C. Goad and B. Remondi pioneered this development.

The NGS first produced processing software that used the Macrometer phase measurements and the precise ephemerides produced by the U.S. Naval Surface Warfare Center (NSWC). Other Macrometer users had to apply the processing software developed by the Macrometer manufacturer which required the use of specially formatted ephemerides produced (and sold) by them. The NGS software was also adapted for the TI-4100 format data and finally for other receivers that were subsequently used.

The original software developed by both the NGS and manufacturers computed individual vectors one at a time. These vectors were then combined in a network or geometric figure, and the coordinates of all points were determined using least squares adjustment programs.

The NGS and the Macrometer manufacturer eventually developed processing software that simultaneously determined all vectors observed during a given period of time (often called session). The second generation multi-baseline software included the ability to determine corrections to the satellite orbits and was often called orbital relaxation software. This technique was pioneered by G. Beutler's group at the Bernese Astronomical Institute. Today, relaxation software is no longer applied since precise orbital information is available to the public.

Today, a huge number of scientific as well as commercial software is available. A review of software features is given in Chap. 11.

Ephemerides service
The first GPS surveys performed in late 1982 using Macrometers depended on orbital data derived from a private tracking network. Later, the broadcast ephemerides were used to supplement this private tracking data. The TI-4100 receiver obtained the ephemerides broadcast by the satellites so that processing programs could use this ephemerides to process vectors. The NSWC originally processed the military ephemerides obtaining "precise" postprocessed ephemerides which were turned over to the NGS for limited distribution to the public.

Today, various organizations around the world provide satellite tracking data from points that are referenced to a global datum. The tracking stations collect code range and phase data for both frequencies for all satellites. The data are processed and archived on a daily basis. Precise ephemerides are now available within a few days to the public upon request.