How GPS Devices Work: Data They Receive and What They Do With It

Sailboat Position Tracking

By Zachary Faber Manning, ECE '19

Introduction

The Global Positioning System (GPS) is a system of satellites owned by the United States government. A GPS receiver is the device commonly referred to as a "GPS," which communicates with these satellites in order to determine its location on earth. GPS was originally created for the US military, but since the year 2000, the satellites offer free service to civilian GPS receivers to resolve their position down to an accuracy of approximately three meters.

For many commercial uses such as street directions, conventional GPS provides plenty of accuracy, but another technology called Real-time Kinematic (RTK) GPS can be used for cases where extremely precise position data is important. RTK GPS uses the same satellite signals as conventional GPS, but an RTK GPS receiver also collects information about a physical property of the satellite signals called phase. RTK GPS receivers can use this additional information to improve accuracy over a hundred-fold. However, while conventional GPS only requires communication with a satellite, RTK GPS receivers must also compare its phase measurements with another receiver fixed on the ground.

This report outlines the technical design considerations associated with both conventional and RTK GPS, and presents a collection of existing solutions to these challenges.

Background

The GPS network was first launched in 1973. While the network was always legal for civilian use, the accuracy of civilian GPS was once severely limited for security reasons by a feature called Selective Availability. Selective Availability intentionally degraded the quality of the public GPS messages, while broadcasting a separate, encrypted signal for military applications. This reduced the accuracy of civilian-grade GPS receivers to hundreds of meters. In the year 2000, Selective Availability was discontinued, allowing for a surge in consumer GPS applications. To date there remain additional types of GPS messages which are encrypted for sole use of the United States Military, while the single type of civilian GPS provides a typical precision of about three meters (GPS: The Global Positioning System, Accessed 2018). GPS surveying and other niche GPS applications have provided demand for extremely high precision position data using augmented GPS methods such as RTK GPS.

Overview of GPS

The modern GPS network consists of a ground control segment, as well as over twenty satellites arranged in a mesh of precisely tracked orbits around the Earth. The central technology of GPS is a technique for determining the distance between a satellite and receiver based on the time it takes a message to travel between them (GPS: The Global Positioning System, Accessed 2018). Since radio waves travel at the speed of light, the distance the message travels is simply the time it takes to travel divided by the speed of light. A GPS receiver can determine its position on the ground by using this technique with multiple satellites at once. Since the satellite locations are predictable, the receiver can deduce its own position if it can resolve its distance to enough satellites at once. This process is called triangulation. The measured distance to a single satellite is called a pseudorange, so named because it is a measurement of the true distance, with some associated error. Although only four pseudorange measurements are needed to determine any position on earth, a GPS receiver often connects to additional satellites to provide extra accuracy and to form an estimate of error.

Transmitting Satellite Time

The ultimate goal of a GPS satellite is to transmit the current time extremely precisely. Any connecting GPS receiver can use this time to determine how delayed the transmission is due to its travel time. To illustrate this, imagine your friend sends you a dated letter in the mail. When you receive the message, you can tell how long it was in the mail by comparing the letter's date with the current date. Similarly, a GPS can determine how long a message has been delayed, and use this information to figure out its distance from a satellite.

GPS satellites need to transmit digital data to the GPS receiver to achieve this. However, radio waves are smooth, continuous waves rather than some binary value. A GPS satellite converts its binary data to a continuous wave using a method called Bipolar Phase-Shift Keying (BPSK). BPSK works by transmitting a radio wave (a sine function) that is multiplied by one or negative one, to transmit a one or zero bit, respectively. This transmission happens at a rate of 1.023 million bits per second (1.03Mbps) (GPS and GNSS for Geospatial Professionals, Accessed 2018). On the other end, the GPS receiver can pick up these waves and translate the negative sections of the wave to zeros and the positive sections to ones.

The satellite's transmission consists of a single sequence of 1024 bits, called a gold code, which is repeated continuously. This code is used for precise alignment of the signal, like the second hand of a clock. Each cycle of the gold code is either transmitted normally, or transmitted with ones and zeros swapped, to represent a single zero or one of a message called the NAV Message, at a much slower rate. The NAV Message is used to communicate the current orbits of all active GPS satellites, the current time, and other relevant information about the network (GPS and GNSS for Geospatial Professionals, Accessed 2018).

A GPS receiver can track the relative delay of these codes from multiple different satellites to determine the difference of time taken for the flight of each satellite message.

Since the gold code is repeated every millisecond, there is an ambiguity of how many whole milliseconds of delay exist between the satellite and receiver. To understand this, imagine trying to determine the time of day by looking only at the minute hand of a clock. You can't, because the pattern repeats. However, the receiver can use the slower NAV Message as a coarse estimate of the time, since it only repeats once every 12.5 minutes. Finally, the actual data contained in the NAV Message includes a "week number" and "time of week" which can effectively be used as absolute time (it repeats as well, but only after several decades) [3]. The receiver can combine all three methods to find the absolute time of day that the satellite message was transmitted, accurate down to a few nanoseconds.

Error Reduction through Differencing

The above process cannot actually produce independent measurements of each satellite distance, because while the receiver can measure the time-of-transmission of the satellite signal, it has no internal time reference with which to compare this time measurement. GPS satellites use enormous atomic clocks to keep precise time, but the internal clock of a typical GPS receiver is driven by a quartz oscillator, which loses precision by nanoseconds within seconds of operation. When measuring traveling radio waves, which move at a rate of approximately a third of a meter per nanosecond, this loss of precision quickly makes it impossible to find the distance to a satellite with any accuracy.

Instead, the receiver compares the difference in the times received from several satellites at once (GPS and GNSS for Geospatial Professionals, Accessed 2018). This information is still enough on its own to determine the location of the receiver. The following thought experiment illustrates how this is possible:

Imagine you wake up in a strange house with no idea where you are or what day it is. Just then, a mail truck delivers three letters to the house: one from New York dated May 3rd; one from Michigan dated May 3rd; and one from Florida dated May 1st. Say you know that the mail service takes a day for every 300 miles traveled. You might reason that you're the same distance from New York and Michigan, since those letters had the same date and arrived at the same time. Since the letter from Florida was sent out two days early despite arriving at the same time, you must be another 600 miles farther from Florida than Michigan or New York. Take a look at a map, and try to figure out a location that matches both of these requirements. You should see that you're somewhere around West Virginia.

A GPS receiver can find its position using this information by first taking a guess at its location, and then refining its estimate to more-closely match the relative distances to satellites that it has gathered. The receiver can then estimate its own accuracy by finding how closely it can match a single position to the data (Blewitt, 1997).

For a single GPS receiver, this type of differencing between satellite times is the only way to subtract out errors. However, with multiple receivers, more corrections can be made by cross-referencing their data. This can account for timing errors in the satellites themselves, as well as delays introduced by Earth's atmosphere. Usually, this is done by using a secondary GPS receiver in a known, fixed location. Since these errors are common to both receivers, by measuring the difference between relative signal times to these two receivers, the wandering baseline of the fixed receiver can be subtracted out of the moving receiver. The combination of these two methods is called double differencing, and it is employed in so-called Differential GPS, or DGPS devices. The United States military maintains a network of such stations which provides GPS augmentation based on this principle (Doberstein, 2014).

Increasing GPS Precision with RTK

An often-cited rule of thumb for GPS signal tracking is that a GPS receiver is capable of achieving precision to approximately one percent of the period of the signal it is tracking (GPS and GNSS for Geospatial Professionals, Accessed 2018). For this case, the period of a waveform can be thought of as the time it takes for the smallest recognizable feature of the wave to be transmitted. In these codes, that would be the time to transmit a single bit. The gold codes used in conventional GPS are 1024 bits long, and take one millisecond to repeat. This means a single bit takes 1ms/1024 = 1 microsecond, so its period is 1 microsecond. Using the 1% rule of thumb, we can approximate that a time measurement should have a maximum resolution of roughly (1%)*(1us) = 10 nanoseconds. This corresponds to a time radio waves take to travel about three meters. This is why conventional GPS is limited to a resolution of 3 meters.

RTK GPS gets around this three-meter resolution barrier by also considering the carrier frequency of the satellite signal. The carrier frequency is the sinusoidal wave of the radio transmission itself, used to transmit data. The GPS receiver can obtain another even more accurate measure of the progression of time by tracking the carrier wave as it repeats.

The GPS signal that most receivers use has a carrier frequency of 1575.42MHz (Doberstein, 2014). In one cycle of the carrier frequency, the signal travels a distance of just 19 centimeters. Again using the one percent rule, we can find that a receiver may be able to track this waveform to an accuracy of about 1.9mm. This represents the ideal achievable accuracy of a device capable of tracking carrier phase. The 19 centimeter wavelength of the carrier is significantly shorter than the best precision achievable through conventional GPS, so there is an unresolvable uncertainty in the number of cycles between the receiver and satellite. This is like attempting to determine the time using only the hour hand and second hand of a clock: we know the position very precisely using a method that repeats, like a second hand, while simultaneously we know the position much more coarsely using regular GPS methods, like an hour hand. But since there is a wide gap of accuracy between them, we can't resolve the position. This problem is known as *carrier integer ambiguity*. To obtain the centimeter-level accuracy typical of modern RTK systems, the receiver must find the number of complete wavelengths between the satellite and receiver (GPS and GNSS for Geospatial Professionals, Accessed 2018).

Integer Ambiguity Resolution

Many methods for resolving carrier integer ambiguity rely on many repeated measurements to average-out errors. Better precision can be found by averaging multiple measurements from the same satellite, or by using the natural motion of satellites around the sky to gain additional precision. After carrier integer ambiguity is resolved once, it can be maintained indefinitely without these methods as long as the carrier phase can be continuously measured. This is because once the distance to a satellite is found to begin with, the previous measurement can be used to find each future measurement. In the case where carrier phase is lost and reestablished, however, there is a discontinuity in the measurements where the receiver could have moved by any number of carrier cycles. This is called a cycle slip, and it means that the carrier integer ambiguity must be resolved again (Carter, 1997).

Cycle slips can sometimes be repaired using other sources of motion data, such as accelerometers (Takasu and Yasuda). This type of correction is typically only achievable if the receiver loses data for a very short period of time, so that the error in other measurement types has little time to accumulate.

Conclusion

Most GPS receivers follow a similar structure of distance calculation and position resolution using timing information from satellites. While the details of modern GPS receivers and satellites are incredibly complex, they rely on principles that can be expressed qualitatively. This can help to explain the limitations and behavior of GPS receivers in different environments. These general principles can be great debugging tools for an assortment of modern devices incorporating GPS receivers.

References

1. Blewitt, G. (1997) *Basics of the GPS Technique: Observation Equations* Department of Geomatics, University of Newcastle
2. Carter, C. (1997) *Principles of GPS.* Allen Osborne Associates
3. Doberstein, D. (2014) *Fundamentals of GPS Receivers: A Hardware Approach.* Springer
4. Dixon, T. (1991) *An Introduction to the Global Positioning System and Some Geological Applications, vol. 29. Reviews of Geophysics*
5. *GPS and GNSS for Geospatial Professionals* (Accessed: 18-Dec-2018) https://www.e-education.psu.edu/geog862/home.html.
6. *GPS: The Global Positioning System* (Accessed: 20-Dec-2018) https://www.gps.gov/
7. Rizos, C. (1997*) Principles and Practice of GPS Surveying* Kensington, N.S.W.: School of Geomatic Engineering, University of New South Wales
8. Rubinov, E. *Review of GNSS Formats for Real-Time Positioning* Cooperative Research Centre for Spatial Information, Australia.
9. Takasu, T. and Yasuda, A. *Cycle Slip Detection and Fixing by MEMS-IMU* Tokyo University of Marine Science and Technology.