THE GLOBAL POSITIONING SYSTEM AND ITS USAGE IN MAKING PATH LENGTH ESTIMATES

MATT OLMSTEAD

Abstract. The Global Positioning System has revolutionized the way positions are found and locations are measured. This technology has impacted many areas from surveying to construction to recreational use. This paper will discuss the Global Positioning System; why it was started, the components of the system, how it is utilized, and how the system is expected to change over the years. In addition, this paper will discuss the optimal points to take measurements to maximize efficiency for several idealized shapes. This will be done by determining how to maintain the path length of the shape with a minimum number of data points.

1. Introduction

There have been many significant technological advances over the last fifty years, one of which is the Global Positioning System (GPS). The Global Positioning System is a satellite navigation system that allows users to accurately determine their location anywhere on Earth. Before this system was invented, the primary method of navigation revolved around the map and compass, but that is not so anymore. Additionally, technology has advanced in the time since GPS was first invented to give more precise measurements. In recent years, GPS has been used for more than simple navigational exercises including applications in geology, agriculture, landscaping, construction, and public transportation and has been used extensively by land surveyors. In this way, GPS is used to measure distances across large areas such as perimeters and path distances. However, this new technology is not perfect; it has drawbacks. This paper will explore some of these drawbacks, with a focus on error estimation.

While it is usually more accurate, or at least more efficient, to measure large distances using GPS, there is still some error involved in the measurement which results from the error of the GPS unit. There are methods to get around this error which will be discussed in Section 2, but sometimes it is not possible to apply these. During these times it will be important to incorporate the error into the measurements and determine the effect of this error (the latter sections of this paper will do just that). Additionally, GPS is not the most efficient method of measuring short distances (on the order of a class room or house) because GPS is not that accurate yet, but is efficient when measuring distances across campus or on larger areas of property. Information regarding the Global Positioning System and its applications in this paper is primarily from three books; Introduction to GPS: The Global Positioning System by Ahmed El-Rabbany [1], Global Positioning System: Theory and Practice by Bernard Hofmann-Wellenhof and Herbert Lichtenegger [2], and GPS for Land Surveyors by Jan Van-Sickle [3].

This paper is divided into several sections. The first section contains the discussion of GPS and how it works. This section begins with a discussion on the history of the Global Positioning System, then explains how measurements are taken using GPS and how GPS is applied to real situations, and ends with a short discussion on other satellite navigation systems. The following sections of the paper look at how to estimate error in path measurements and how to best approximate path lengths using a finite number of data points. The second section is broken into parts, with the first part introducing the bar-bell concept of error approximation, the next part using the bar-bell concept to measure a triangular path while assuming error on the measurements, and the last part of this section estimating the perimeter of a semicircle using these bar-bells. The third section discusses how to estimate the path length of other shapes by taking into account curvature, path length travelled, and angle travelled through. The final section of this paper discusses unsolved problems and future directions to take while looking at this problem.

2. History

The Global Positioning System is the successor to the TRANSIT system. TRANSIT was developed by John Hopkins Applied Physics Laboratory in 1959 and declared operational in 1962 when five satellites were in orbit. TRANSIT uses six satellites and is operated by the U.S. military to give coordinates to vessels and vehicles. However, in 1967 civilian use became allowed, and TRANSIT is still in use today by small vessels and aircrafts. The problems with TRANSIT are due to the small number of satellites which causes time gaps in measurements as the satellites pass overhead at most every 90 minutes. TRANSIT also has a low navigational accuracy. In 1973 the US Air Force set forth to develop the Global Positioning System to improve these deficiencies. While military in use, President Reagan and Congress directed the Department of Defense to promote civilian use of GPS and offer it free of charge after a Soviet shooting of a Korean Airliner in 1983. In 1993, the Secretary of Defense declared Initial Operational Capability of GPS, which means that it is no longer a developmental system.[2]

3. GPS: Configuration

The Global Positioning System is set up to ensure continuous worldwide coverage which it manages to do by having the satellites arranged in six orbital planes with at least four satellites in each plane. Three satellites are needed in each plane in order to attain worldwide coverage, but it is even better with four or more satellites in each orbital plane.[1]

There are three segments to the GPS: the space segment, control segment, and user segment. The space segment consists of the 24 satellite constellation, with each of the satellites transmitting a signal down to earth. Each signal contains five components, two carrier frequencies, two digital codes, and a navigation message. The codes and navigation are sent as binary biphase modulation added to the carriers. The signal is used to determine the distance from the receiver to the satellite, and the navigation message includes the location of the satellite as a function of time. The control segment is made up of a worldwide network of tracking stations with the master control station in Colorado Springs, Colorado. The control segment is responsible for tracking the satellites to predict future satellite locations and also to makes sure the satellite is functioning correctly. The user segment includes all of the users of GPS, both civilian and military, who are able to use receivers in order to find their position on earth.[1]

A GPS satellite (Figure 1) weighs about one ton and measures about 27 feet in length when the solar panels are extended. The satellite is able to generate about 700 watts of power and moves at about 8700 miles per hour. The satellites each have three-dimensional stabilization to make sure that their solar panels are pointed toward the sun and that their 12 helical antennae are pointed toward the earth. There are some times when the satellites must pass through the shadow of the earth and there are on-board batteries to provide power during this time. The satellites have thermostatically controlled heaters and reflective insulation so they can stay at the appropriate temperature to ensure that the oscillators (clock) operate correctly. The GPS satellites are checked out at a facility at Cape Canaveral, FL prior to launch to make sure they are functioning properly.[3]

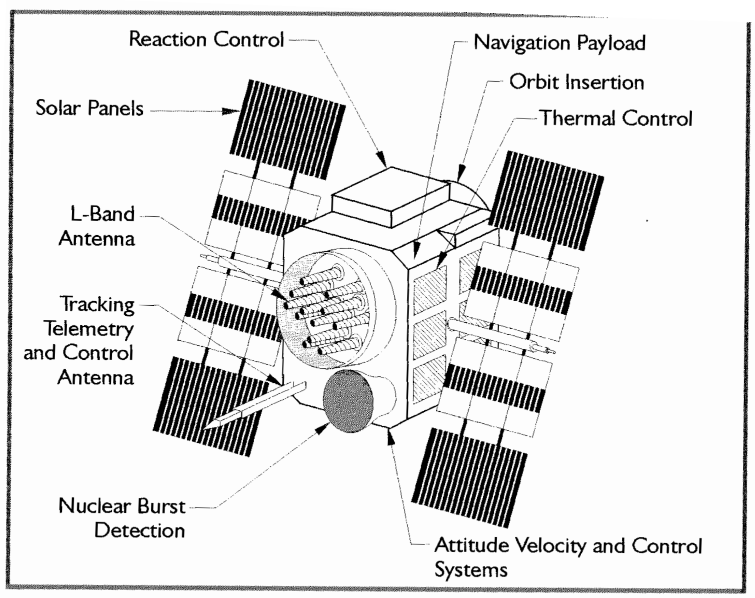

FIGURE 1. This is a picture of a Block II Satellite. The antennae are pointed toward the earth while the solar panels are pointed toward the sun during operation.[1]

GPS started with 11 Block I satellites. The first satellite was launched on February 22, 1978 and the last Block I satellite was launched on October 9, 1985. Block I satellites were used primarily for experimental purposes with a life expectancy of 4.5 years. After Block I satellites came Block II/IIA satellites which can function without ground support for 14 or 180 days, respectively. A total of 28 Block II/IIA satellites have been launched and include new security features such as selective availability and antispoofing. Block II/IIA satellites have a lifetime of 7.5 years, but almost all of the satellites exceeded that lifetime. Block IIR satellites are currently being launched, are expected to have a higher accuracy, and are expected to be able to go at least 180 days without ground corrections or accuracy degradation.[1] The current configuration of satellites in the sky is 5 Block II, 18 Block IIA, and 6 Block IIR for a total of 29 satellites.[4]

The control segment consists of the master control station (MCS), monitor stations, and ground control stations. The five monitor stations are in Colorado Springs, Hawaii, Kwajalein (an island west of Hawaii), Diego Garcia (an island south of India and northeast of Madagascar, and Ascension Island (an island in between South America and Africa, east of Brazil), and their locations are known precisely. These monitor stations are responsible for tracking all of the GPS satellites in view and three of them have ground antennas so that information can be uploaded to the orbiting satellites. Each of these monitor stations and ground control stations (not the MCS) is unmanned and operated remotely from the MCS.[1]

There are several additional satellite navigation systems created and operated by countries other than the United States. Russia built and developed a system called GLONASS which is very similar to GPS. GLONASS should have 21 operational satellites and 3 spare satellites that orbit at 19100 Km. The orbits of GLONASS satellites are approximately circular, taking 11 hours and 15 minutes to orbit, and are arranged in three orbital planes. GLONASS uses two L-band carriers, a C/A code on L1, P-code on both L1 and L2, and a navigation message. The difference is that each GLONASS satellite has its own frequency but this is going to change so there is only one frequency, making GLONASS even more similar to GPS. GLONASS uses the frequency to determine which satellite is sending the signal. In January 1996, GLONASS was completed with 24 working satellites but in May 2001 there were only seven satellites. A new generation of satellites are expected to be launched and feature improved clocks and capability of sending the C/A code on both L1 and L2. It is possible to use both GPS and GLONASS together, but GLONASS uses PZ-90 whereas GPS uses WGS 84 as discussed last week. This results in as much as a 20 meter difference. [1]

China has launched two navigation satellites that are placed in geostationary orbits at about 36000 Km above Earth's surface. Right now this system is used in land and marine transportation and China is planning to build a second positioning and navigation system with more satellites and more coverage.[1]

GLONASS and GPS are not able to meet all of the civil aviation requirements and are augmented by regional systems. These regional augmentation systems usually combine GPS or GLONASS with geostationary satellites with navigation transponders and ground reference stations.[1]

Europe is in the process of building Galileo, a global satellite navigation system. Galileo will be civil-controlled delivered through a public-private partnership. After investigation, it was decided that Galileo would consist of 30 Medium Earth orbit satellites in three orbital planes at about 23000 Km. Galileo will be compatible with GPS and GLONASS, but will have two levels of service, the basic, free service and a subscription service with additional features. A European political body not associated with Galileo management will have the authority to take over the system during a crisis and for security purposes.[1]

3.1. **GPS: Signal.** The satellite transmits a microwave signal that contains 5 parts: two carrier frequencies (1575.42 MHz L1 and 1227.6 MHz L2), 2 digital codes and one navigation message. The two frequencies allow the user to adjust for ionospheric delay if using a receiver capable of receiving both the signals. Many civilian receivers only receive the L1 frequency. All GPS satellites transmit the same L1 and L2 signals, but have different code modulation and navigation messages. The two digital codes are C/A-code (coarse acquisition) and P-code (precision). The C/A code is transmitted only on L1, while the P-code is transmitted on both L1 and L2. These digital codes are a stream of binary digits with the C/A code having 1023 binary digits that repeat every millisecond (1.023 Mbps), and the P-code being much longer and involving a more complicated process.[1]

Each satellite has a unique C/A code so that the receiver knows which satellite is sending the signal. Using only the C/A code is not as accurate but it is accessible to all and is less complicated. The P-code is very long, 266 days at 10.23 Mbps which is 10 times faster than the C/A code, and divided into 38 segments each one week long. Thirty two segments are assigned to various satellites with the other 6 segments being reserved for other uses. As the segments are one week long, they start over Saturday/Sunday at midnight. The satellites are commonly identified by which part of the signal they transmit: ID of PRN 18 has 18th week segment. The P-code is primarily for military purpose. All users were allowed access to the P-code until January 31, 1994 when a W-code encryption was added making it the Y-code. This encryption is antispoofing.[1]

The navigation message is binary biphase modulation at 50kbps which takes 12.5 minutes to go through. The navigation message contains the coordinates of the satellite as a function of time, the health status of the satellite, clock correction, almanac, atmospheric data, and information on other satellites including approximate location and health of different satellites.[1]

Another important detail about the signal is cycle slips. Cycle slips occur when a satellite signal is lost for any amount of time, it can be for just one cycle or millions of cycles. Cycle slips can occur when the signal isn't fully received due to buildings, bridges, trees, radio interference, severe ionospheric disturbances and a faulty receiver. The cycle slips occur when the receiver does not receive the signal, and if not identified can cause large errors in the measurement. Like most measurement corrections, the methods for determining cycle slips involve two receivers.[1]

3.2. **GPS: Receiver.** The quality and price of GPS receivers has changed drastically in the 20 years they have been commercially available. In 1980 there was one receiver which cost $250,000 and now there are over 500 commercial models that range from $100 for the low-end commercial model to $15,000 for the sophisticated geodetic quality receiver. The price will continue to drop as more advanced technology becomes available.[1]

The antenna receives the signal and converts the energy into a current which the receiver can then analyze. There are four types of receivers: single-frequency code receivers, single-frequency carrier-smoothed code receivers, single-frequency code and carrier receivers, and dual-frequency receivers. All of the receivers receive L1 while only a dual-frequency receiver receives both L1 and L2. The single frequency code receiver measures the pseudorange with C/A-code only. This is the cheapest, least accurate receiver, and is mostly used for recreational purposes. The single-frequency carrier-smoothed code receiver also receives only C/A, but a higher resolution carrier frequency is used internally to improve the resolution of the code pseudorange, resulting in high precision measurements. The single-frequency code and carrier receivers get the same information that the previous two do, but also can output raw C/A-code pseudoranges, L1 carrier-phase measurements, and navigation message. The dual-frequency receivers are the most-sophisticated and most expensive receivers and were capable of getting everything before antispoofing. They are not able to get the Y-code, although they are able to recover a full but weaker L2 signal.[1]

3.3. **GPS: Modernization Program.** The GPS system was designed in the early 1970's and is ready for a modernization period. The aims are to provide signal redundancy and improve upon positioning accuracy, signal availability, and system integrity. The plan is to add C/A-code to L2 and add two new military codes (M-codes) on L1 and L2. The 12 Block IIR satellites that have been/will be launched since 2003 should have these new codes. Two C/A codes will allow for a stand-alone GPS receiver to account for ionospheric interference and is expected to improve accuracy to within ≈ 8.5 meter accuracy. There are also plans for a third civilian signal L5 at 1176.45 MHz primarily for aviation since L2 is near the same frequencies as ground radar and causes interference. It will also have higher power and a wide bandwidth of 20 MHz and will be added to Block IIF satellites. This will allow higher accuracy under noisy or multipath conditioners and will be longer than current C/A codes to reduce self-interference and become more accurate. Block III satellites are planned for future modernization in 2030 and the ground control is also expected to be updated to move the accuracy to ≈ 6 meters.[1]

3.4. **GPS: Time Systems.** There are two time systems used, Coordinated Universal Time (UTC) and GPS time. UTC is based on the International Atomic Time (TAI) which is computed based on independent time scales generated by atomic clocks spread throughout the world. In surveying, a time is wanted based on the relation of Earth's rotation. We get this by introducing leap seconds that adjust UTC that keep UTC within 0.9 seconds of Universal Time. The leap seconds occur on either June 30 or December 31 and as of 2001 TAI is ahead of UTC by exactly 32 leap seconds. GPS Time is used for referencing (time tagging) the GPS signals computed based on time scales generated by atomic clocks at monitor stations and aboard GPS satellites.[1]

There is another complication involved in the timing systems known as sidereal time. Each day, the satellites will be in the same spot they were the day before, but 3 minutes and 56 seconds earlier than they were the previous day. This is because the satellites go about their orbits twice every sidereal day whereas Earth uses solar days which are different than sidereal days by 3 minutes and 56 seconds. This is important when a time is found in which the satellite configuration is very good, the time that the satellites will be in the same position is different on other days. This could be fixed if the satellites were pushed another 50 Kilometers higher(from 20,000 to 20,050), but that's not likely to happen and it's not that important.[3]

3.5. **GPS: Coordinate Systems.** The Earth is not a perfect sphere, but is treated as a biaxial ellipsoid. This biaxial ellipsoid is called a geodetic datum, a reference ellipsoid with a well defined center and orientation. This is known as the geoid and best approximates the mean sea level on a global basis. Before GPS or similar satellites were invented, the horizontal and vertical positions on the Earth were measured independently, but with this satellite technology, the position is represented based on the 3-D geodetic coordinate system. This coordinate system required the geodetic latitude ϕ, geodetic longitude λ and the height h. The Conventional

Terrestrial Reference System (CTRS) is a 3-D geocentric coordinate system whose origin coincides with the center of Earth and rotates with Earth, making it also known as an Earth Center Earth Fixed (ECEF) coordinate system. The z-axis points toward conventional terrestrial pole, while the x-axis is defined as the intersection of the terrestrial equatorial plane and the meridional plane that contains the mean location of the Greenwich Observatory.[1]

The official GPS reference system is WGS84 which means it is the World Geodetic System of 1984. Another common coordinate system is the NAD83 which stands for North America Datum of 1983. If one of these coordinate systems is supposed to be used, but the other one is used instead, the error is in the millimeter range, so usually considered negligible. [1]

3.6. **Satellite Orbits: An Introduction.** One of the most important pieces of information required to use GPS to measure a location is to know where the satellites are located. Most of the error in position (while using a single receiver) is attributed to not knowing where the satellites are located. The position is sent by the satellite and this position is pretty accurate, while it could be better. The position of the satellites can be found days later from several sources, although the military is no longer allowed to display their most accurate locations of the satellites. The orbit description is described by a first order approximation of the Keplarian model treating the Earth and the satellite as point masses.

$$\vec{a} + \frac{G\,(M+m)\,\vec{r}}{r^3} = 0.$$

The mean angular speed of the satellites in orbit is described by

$$\sqrt{\frac{GM}{a^3}},$$

where M is the mass of the Earth and a is the semi-major axis. The instantaneous position of the satellite within its orbit is described by an angular quantity known as anomaly. There are three types of anomaly; mean, eccentric, and true. [2]

3.7. **GPS: Pseudoranges.** The pseudorange measures the distance between the satellite antenna and the receiver antenna. The distance is needed to compute the position of the receiver. To picture how pseudoranging works, first assume that the clocks of the receiver and satellite are perfectly synchronized. When the satellite sends out its PRN code, the receiver generates the exact code at the exact time. When the code from the satellite makes it to the receiver, there will be an offset of the two codes caused by the separation distance. The travel distance can then be computed using the speed of light and because the travel time is now known. However, the clocks are not synchronized and that is why this is not the range but the pseudorange. This is also why four satellites are needed to accurately measure position.[1]

4. GPS: Positioning and Techniques Used in Surveying

GPS can be used in two ways to measure position; either point positioning or relative positioning. These methods can be broken down further into static and kinematic, making a total of four ways to use GPS. In point positioning one receiver is used to measure position, while in relative positioning, two receivers measuring the same satellites are used. One of the receivers is the reference and is located in a position whose coordinates are known. The other receiver, rover or remote, is placed at a location whose coordinates are unknown. The reason for this is that when the receivers are relatively close, within a couple dozen kilometers, the total error (including ionospheric, clock errors, ephemeris errors, and tropospheric errors) should be about the same for both receivers. This says that the vector connecting the two positions will be the same as the vector that actually connects the reference and the rover. The position of the rover can then be determined by using the precise position of the reference and the found vector. Using two receivers and this differencing process will also give very accurate length measurements because both receivers should have very similar error. [1]

In addition, there are several hybrid techniques in GPS that make use of kinetic and static positioning. Pseudo-kinematic positioning makes use of short static observations and revisits to positions to measure again. Semi-kinematic positioning make use of short static observations and the receiver that moves alternates between receivers (can be more than two receivers). Another hybrid technique is rapid static in which the receivers are stationary during their observations but the observations are very short (both code and carrier observations on L1 and L2 are needed for rapid static observations).[3]

RTK surveying is a carrier phase-based relative positioning technique that is suitable when there are a large number of unknown points within a couple of dozen kilometers of a known point, the coordinates of the unknown points are required in real time, and the propagation path is relatively unobstructed so the signal will not be lost. [1]

5. GPS: Applications

GPS has many uses including industrial, recreational, marine, and airborne applications. GPS is also useful in hazardous areas where human lives are at risk. As GPS gives a precise location, it has been very helpful in providing a cost-effective, efficient, and accurate tool in creating utility maps. By using pipe/cable locator's attached to a GPS controller, accurate location and depth of pipes can be established in a low-cost and efficient manner. GPS is used in forestry and natural resources including fire prevention and control, harvesting operations, insect infestation, boundary determination, and aerial spraying. GPS provides the efficient resource-management system needed to identify and monitor the exact locations of the forestry resources. Previously, aerial photography was the only way to provide the shape and location of cut blocks during harvest season, but now GPS is able to do this in real time. GPS is also able to provide precise positions of wildlife activity centers which can be easily returned to by using way-points. [1]

Additionally, GPS has many uses in the agricultural industry and precise farming including soil sample collection, chemical applications control, and harvest yield monitors. During soil samples, GPS is used to take samples at predetermined locations on a grid to help in accurate mapping of information such as nitrogen and organic material contents. During field spraying, GPS is integrated with an aerial guidance system and together are able to make it so the right amount of chemicals are applied to the intended spot. This minimizes overlap and chemicals while maximizing both fuel efficiency and productivity. In crop fields, GPS is used to help map yield rates which then show which locations had the best yield.[1]

In civil engineering applications, GPS is used during road construction, Earth moving, and fleet management. During Earth moving, GPS is able to make sure that the desired grade of the land is achieved. This technology is also used in laying down foundations in making sure that the material is placed in the desired positions. GPS is integrated with other technology and is used to keep track of equipment, maximize deployment efficiency of equipment, and to help guide vehicles to their destinations.[1]

Since early development, GPS has been used to monitor the stability of structures. This requires very high accuracy as the deformation of buildings like dams, bridges, and TV towers does not occur very quickly, but is very important. Deformation is measured by taking measurements at the same place over different time intervals. In monitoring these structures, it is best to integrate GPS with other technologies. When monitoring bridges, dual GPS receivers are used. GPS has found use in open-pit mining. It is much more efficient in measuring drill bit depth and the stakes that were previously used were often buried or misplaced. RTK GPS has improved drilling, shoveling, vehicle tracking, and surveying in open-pit mines. RTK provides centimeter accuracy and only needs one reference receiver which can be used with many rovers. Shovel operators are able to keep the correct grade when loading ore into haul trucks by using GPS.[1]

GPS is also used for land seismic surveying and marine seismic surveying. Seismic surveying is used to map the subsurface geology of the planet which can then be used to find oil and gas. Low-frequency acoustic energy is the key. The signal wave is commonly produced by a large metal plate that can vibrate being placed on the ground. The energy is sent into underground rock layers and is affected by the physical properties of the rock and parts of the signal are reflected. The reflected wave can be detected by geophones. When the geophones detect the energy, they output signals that are proportional to the intensity of the reflected wave which are recorded for analysis. GPS is used to know the positions of the geophones and of the signal source, without these positions, the information is useless.[1]

In marine seismic surveying, the basic procedure is the same. However, different methods are required at different water depths. In deep waters, four to eight several kilometer long seismic cables containing hydrophones are dragged behind the seismic vessel. The acoustic energy is produced by air guns that are towed behind the vessel at depths of approximately six meters. In shallow waters, both land and marine methods are used. A new technology has been used for waters up to 200 meters in depth. It is called an ocean bottom cable survey and uses both hydrophones and geophones that are combined into a single receiver. This method helps to avoid water column reverberation. As with land surveying, the locations of the energy source and the hydrophones must be known very accurately and GPS accomplishes this the most efficiently and at the lowest cost.[1]

GPS has been used alone for topographic mapping of small areas. However, it becomes cost ineffective for large areas, inaccessible areas, areas that take a long time to travel through like forests, and in areas where many locations need to be determined like the coast. Previously, large areas were mapped using aerial photogrammetry. In this method, the plane flies by and takes pictures which are used to construct the map after the pictures are developed. The images must first be related to a coordinate system for the map to be of any use. Ground control stations with known coordinates and known aerial images are used as references to properly place the coordinate system on the gathered photographs. Now, GPS is used to take the position of the camera as well as the precise time that the image was taken. This is a much more cost effective way as less ground control points are needed and there is not as much time spent in post production setting up the coordinate system on the photos.[1]

GPS has been used to map the seafloor. It is important to know accurate water depth and shape of the sea bottom. When accurate water depth is known, the ships can be loaded so to get the maximum use out of them. Previously water depth was measured using a single-beam echo sounder and measuring how long it took for the reflected signal to be measured at the surface. It is then straightforward to figure out the water depth. The process was simple, but time consuming and didn't cover the entire ocean floor, just preset grid lines. A new technology of seafloor mapping has been developed which integrates multi-beam echo sounders and GPS. The multi-beam sounder is better than the single-beam sounder because it is able to cover the entire seafloor whereas the single-beam sounder did not cover the entire seafloor. GPS is used to make sure the vessels follow the designated paths to properly map the seafloor. GPS is also used to provide accurate positioning and altitude of the mapping vessel which is also needed to properly map the seafloor. [1]

Vehicle navigation makes use of GPS in the way of digital road maps integrated with a computer system equipped with GPS. GPS is used to find where the vehicle is at any moment and then that position is superimposed on the digital map. With this technology, it is possible to get turn by turn directions, sites of interest (airports, attractions, hotels, etc), and the best routes to get to the destination. The program incorporates information like one-way roads, illegal turns, and rush-hour restrictions into the shortest route and gives a time to destination as well. Similarly, GPS is used in transit systems to locate vehicles if they go off-route. However, transit systems are common in cities with high-rises and GPS needs to be integrated with other systems for accurate positioning.[1]

A common use for GPS is in way-point navigation (called stakeout by surveyors). A way-point is basically a location that has either been recorded and the coordinates stored, or a point in which the coordinates were downloaded into the receiver. Way-points are useful because they allow the user to go back to locations or to find locations they had not been to before with ease. Geo-caching makes extensive use of way-points. Geo-caching is when a person using a GPS goes out and tries to find something with known coordinates. Way-points are a very convenient tool that has use in many of the previously mentioned applications. [1]

6. GPS: Integration of Technology

While GPS is a very useful technology, it does have its limitations. These limitations primarily occur when the receiver is not able to receive signals from at least four satellites. This occurs in densely wooded areas, in buildings, canyons, in cities with lots of tall buildings, and in deep open-pit mining. There are many situations in which GPS is useful where the signal is not able to be fully received. To make use of GPS in these situations, it needs to be integrated with other technology.[1]

One example is GPS/GIS integration. A geographic information system (GIS) is a computer-based tool that can acquire, store, manipulate, analyze and display data that is referenced according to its geographic location. GPS is one way in which locations can be determined and is done so efficiently and accurately. GPS/GIS integration is commonly used in utilities management, forestry, agriculture, public safety, and fleet management.[1]

GPS can also be integrated with laser range finders (LRFs). This integration is useful when GPS receivers will lose their satellite locks such as in densely wooded areas. In this situation, a GPS antenna is situated in an open area where it will operate normally (Figure 2). Then an LRF is used to find the position of a point that is inaccessible that is in the trees. This information can then be used to find the location that is in the woods. [1]

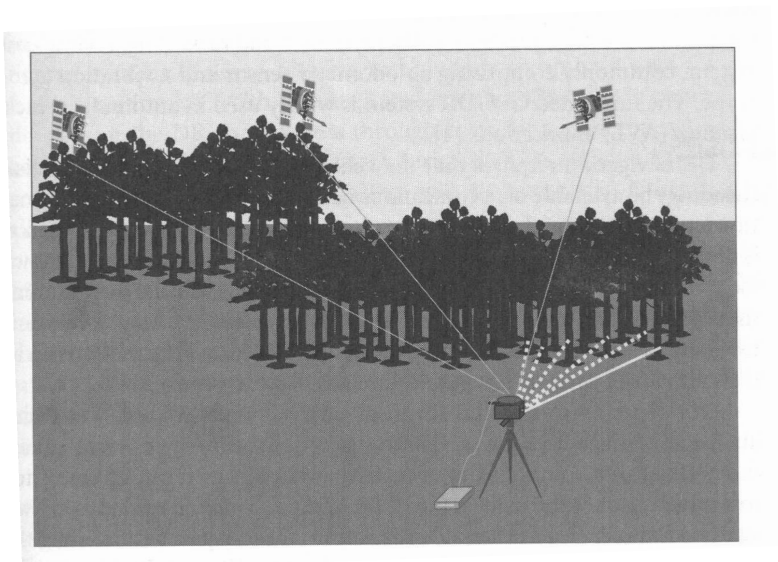

FIGURE 2. This is a picture using GPS/LRF. The antenna is situated in an open area, and then the locations in the woods are determined using a laser range finder and a compass to determine which direction the measurement is taking place at.[1]

Another system that supplements GPS when the signal can't be received is dead reckoning. A dead reckoning system is made up of a low-cost odometer sensor and a vibration gyroscope. The odometer is able to determine the distance travelled while the gyroscope can determine the direction traveled. This integrated system is commonly used in vehicles. If the vehicle moves from a known location, it can then determine how far it has traveled and in what direction it has traveled by using dead reckoning. This system will work accurately, but over time will become less accurate mostly because of changes in the tires like pressure, tire wear, slipping and skidding but also vehicle speed. A new inertial positioning system is being looked into, microelectro mechanical system technology (MEMS), that will provide the same function as odometers and gyroscopes and replace the traditional dead reckoning system. [1]

A device called a pseudolite is commonly used in combination with GPS in open-pit mining (Figure 3). The pseudolite is needed because the signal may not be received while in the pit so GPS technology is integrated with a pseudolite

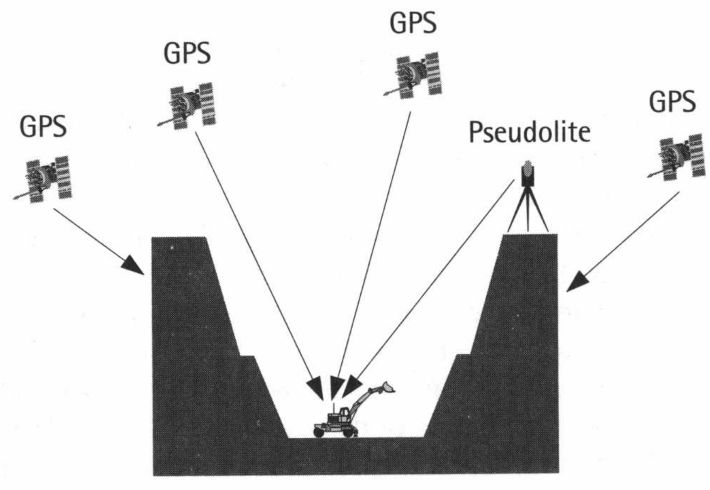

FIGURE 3. This is a picture of GPS satellites integrated with a pseudolite for positioning in an open-pit mine.[1]

for accurate positioning. A pseudolite is a ground-based electronic device that transmits a code similar to a GPS signal(code, carrier frequency, and data message) which can be received by a GPS receiver. This combination of technology can improve the position dramatically. There are some downsides to this integration. The first of which is called the near-far problem which occurs because the signal from the pseudolite is very strong when the receiver is close causing it to jam and overwhelm the other signals and the signal is weak when the receiver is far away so the signals from the satellites are much strong and overwhelm the pseudolite signal. This problem does not occur when just using satellites because the satellites are always far away and the signal is approximately constant at every point. The pseudolites also use only low-cost crystal clocks so there is inaccuracy there. Multipath error again comes up as a result of the signals being reflected and is more common when using pseudolites. Pseudolite integration is also useful in precise aircraft landing and deformation monitoring. It is possible to use pseudolite-only positioning, but it is important to know the positions of each pseudolite. Pseudolite-only positioning would be useful for indoor applications as well as underground mining. [1]

Recently GPS has been integrated with cellular phone technology. This is because of FCC requirements that make it so that wireless emergency calls need to be located within 125 meters. A GPS receiver is integrated into the cellular handset and the position can be determined to much greater accuracies than 125 meters. The downside to this approach is that only new phones can be equipped with GPS and that the signal is weak inside of buildings. There are many more applications of GPS and wireless technology that are just now starting to be discovered. [1]

7. GPS: Error Approximation

Taking a position using GPS will result in some error of the coordinates. There are many known reasons for this error, some of which were discussed previously like Selective Availability and antispoofing, while many have not been discussed. The first of these is ephemeris (orbital) which comes from the fact that the exact position of the satellite is not known because the model of the forces acting on the satellite is not perfect and results in a satellite position error of 2-3 meters (50 meters when Selective Availability was active). While the ephemeris error is a property of the satellite, it affects different locations differently as the amount of the error is related to the angle of the receiver. If the receiver is directly below the satellite, the error is smaller than when the satellite is at a small angle with the satellite relative to the ground. The combined error of ephemeris and the clock error results in about a 3 meter range error. Although the satellites have atomic rubidium and cesium clocks (only rubidium in Block IIR), it is found that the clock error is 8.64-17.28 ns/day which causes the range to be off by 2.6-5.2 meters. The receivers do not have atomic clocks and are not as accurate in keeping time as the satellites which causes more error.

Another error is multipath error which results from the antenna receiving more than one signal. One of these signals is the actual signal, while the others are reflected off a surface like water. When this error occurs, it affects both carrier-phase and pseudorange measurements on the order of 10's of meters and new advances to get rid of multipath error are not likely. The antenna-phase center error results in errors in the centimeters and comes from the signal being received at a location other than the geometric center of the antenna. There is also electrical noise from the circuits making up the receiver which affects position by around 0.5 meters. As the signal is sent through the atmosphere, there is ionospheric delay and tropospheric delay. While the ionosphere speeds up the propagation of the carriers, it slows down the propagation speed of the PRN code, this causes too short of a measurement using the carrier and too long of a measurement using the code, but since the propagation speed changes at the same rate, this error can be gotten rid of by differencing the two measurements. If the receiver is not capable of doing this, an error of 5-15 meters may occur. The troposphere slows the propagation speeds at the same rate so they can not be canceled by differencing them.

In addition to these types of error, there was an additional source of error in GPS. There were two methods in use (1990) to deny civilians full use of the system, selective availability and Anti-spoofing. In selective availability (SA), the transmitted signal from the satellite is truncated so that the satellite position cannot be accurately computed, causing error in measurements of 30-50 meters. Anti-spoofing affects many of the high accuracy survey uses by only allowing authorized receivers to get the most accurate signal. In the first Gulf War, SA was turned off so that the US troops could use civilian model receivers that were more readily available at the time. Selective availability is no longer in use by the US government.[2]

7.1. **The bar-bell.** When taking a reading with a GPS unit, the given coordinates have an error associated with them. That is, the given coordinates do not give the precise location, but the measured coordinates are within a radius δ of the actual location. In using this idea, we first look at a line segment constructed from two points. If we measure the distance between the two points to be l, then the actual distance, L, between the two points is

$$l - 2\delta \ \leq L \leq l + 2\delta \tag{1}$$

and looks like Figure 4. Equation 1 shows that the actual length is given by taking the actual points where the measurement took place and connecting them with a line which will have length L, and then putting the error circles around these points. The error circles correspond to the error of the GPS and have radii of δ. Figure 4 looks like a bar-bell so that is what we will call it. When the measured

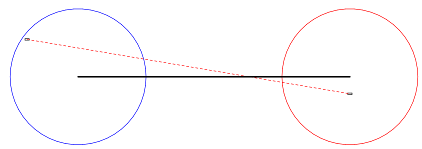

FIGURE 4. This is a picture of the bar-bell that is being used. This bar-bell shows the measured distance (dashed line) when the two measured points are within the error circles. When the measured points are on the opposite outer sides of the circle, the measured distance is at a maximum.

points are on the axis at the outer sides of the bar-bell, the measured distance is maximized. When the two points are on the axis at the inner sides of the bar-bell, the measured distance is minimized. These are the two worst-case estimates for the actual distance. When we apply this idea to two connected path segments, the error cannot be maximized or minimized for both path lengths. This can be seen in Figure 5. We notice this because the middle point is a measurement that corresponds to two connected bar-bells which cannot both give maximum or minimum error. In relating this to GPS, this is equivalent to the fact that interior waypoints share the same measurements so they can not be worst case measurements for both line segments.

This idea applies for any number of segments which corresponds to any number of bar-bells. If one bar-bell gives a maximum distance, the bar-bells that are directly connected to it cannot give maximum distances. In the first example, we will look at methods to estimate the worst-case error for a three measurement, two-segment path. We will then use the bar-bells in a slightly different method by assuming that each bar-bell gives the maximum length. We will do this by assuming that two measurements are taken at each interior point. One of these measurements will be for the first bar-bell, and the other measurement will correspond to the second measurement so that we have two maximized bar-bells. When using GPS, we will only take one measurement at each point. By doing this, our task changes to finding out where to take measurements to get the best approximation for the perimeter of a figure.

7.2. **Triangle.** The first example we will look at is of a triangular path made of two segments and three points (Figure 6). In looking at the triangle example, we first want to note a few things about the bar-bell discussed in the previous section. We know that our measured point will be within a radius δ of the actual point. When we plot this, the measured point will be at an angle in relation to an axis that we choose (see Figure 7). Using this idea, we are now ready to look at a specific

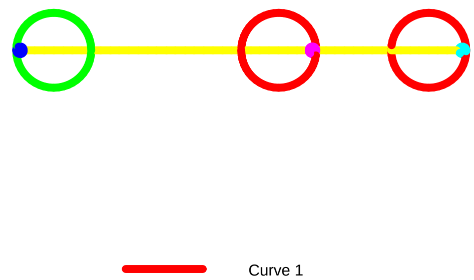

FIGURE 5. This is a picture of two bar-bells connected together. The dots on the circles represent where the measurement is taken. This shows that when one of the distances is a maximum, the other one cannot be a maximum. The left bar-bell is maximized which makes it so that the right bar-bell cannot be maximized, but does not stop the right bar-bell from being a minimum.

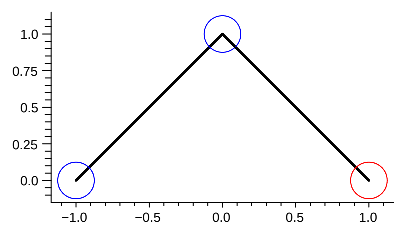

FIGURE 6. This figure shows the triangular path that is discussed in this section.

path made up of three points in the shape of a triangle. As we see in Figure 8, the actual measurements are taken at the three points $(-1,0), (0,1)$, and $(1,0)$ with the angles of the measured positions α, β, and γ being taken with respect to the x-axis. We also notice that the length of this path from $(-1,0)$ to $(0,1)$ to $(1,0)$ (which we will now call a, b, and c, respectively) is $2\sqrt{2}$ if there are no error circles. This means that the error range for our measured distance should include this number. Next, we construct the lengths of these two paths, calling the path going from a to b P_1 and the path going from b to c P_2. Using the distance formula, we see that the length of P_1 is

$$\sqrt{(\delta\,\cos(\beta)+1-\delta\,\cos(\alpha))^2+(1+\delta\,\sin(\beta)-\delta\,\sin(\alpha))^2}, \tag{2}$$

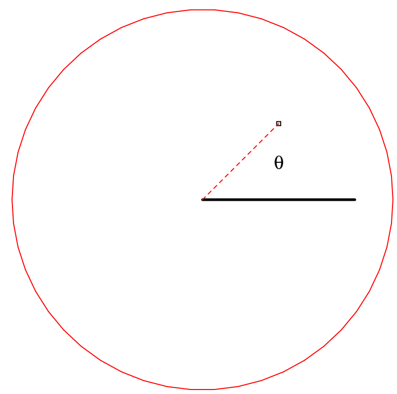

FIGURE 7. This figure takes the actual point of the measurement to be at the center of the circle and shows how the angle θ of the measured point is defined.

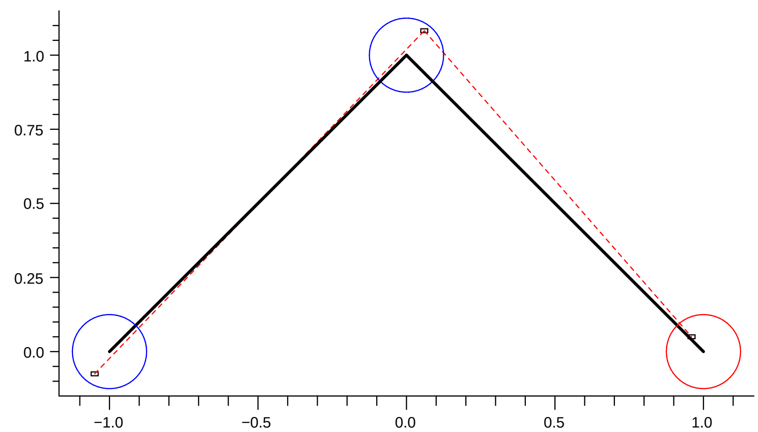

FIGURE 8. The path follows two sides of a triangle that has vertices at $(-1,0)$, $(0,1)$ and $(1,0)$. The error circles around the vertices are each of radius δ and the angles are α, β, γ going from left to right. The dashed line is of a possible measured path, while the solid lines are the actual path between the two actual points.

and the length of P_2 is

$$\sqrt{(1+\delta\,\cos(\gamma)-\delta\,\cos(\beta))^2+(\delta\,\sin(\gamma)-1-\delta\,\sin(\beta))^2}. \tag{3}$$

The total length of this triangular path is then $P_1 + P_2$. In order to determine the possible measured minimum and maximum perimeter lengths, the error range, we take the partial derivatives of the length measurement and need to optimize the system of equations for the three angles, α, β, and γ. Since we are not dealing with a system of linear equations, the best method to finding the angles is through approximation using Newton's Method for nonlinear problem solving. We will use Maple's command NLPSolve to optimize these equations for us. To use NLPSolve, we need to pick a value for δ and a range for each of the angles. When we pick a value for δ, we assume that the measured point lies upon this circle. This is a fair assumption because the maximum and minimum lengths will occur along this circle. We choose δ to be $\frac{1}{8}$ because this corresponds to about a 6 meter error for length measurements of about 50 meters. We choose a range for the angles based upon where we think the angles are.

When we use NLPSolve with these parameters, we find that for a maximum path length α is 228.3664 $^\circ$, β is 90.0000 $^\circ$, and γ is -48.3664 $^\circ$ and the path length is 3.2604. Using NLPSolve to find the minimum length, we see that α is 36.8699 $^\circ$, β is -90 $^\circ$, and γ is 135.5787 $^\circ$ and the path length is 2.7602. This specific case is shown in Figure 9. We notice that these angles are not 180 $^\circ$ from each other and

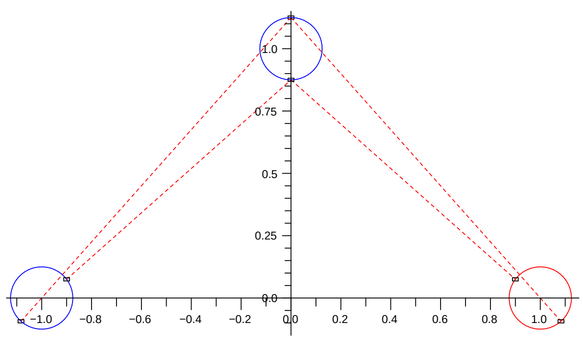

FIGURE 9. This figure shows the two optimized paths for the triangle in the case that $\delta = \frac{1}{8}$. We see that one of the paths goes through the error circles until it hits the other side of the circle (the maximized path) and that one of the paths stops as soon as it hits an error circle (the minimized path).

that the range for the path length includes the value if we did not have any error in our measurements. We also use NLPSolve to make sure that as δ goes to zero, the measured path length goes to $2\sqrt{2}$ and that the angles go to α=45 $^\circ$, β=-90 $^\circ$, and γ=135 $^\circ$ for the minimum length while the angles for the maximum length differ by 180 $^\circ$ as expected from the figure.

When we close this path and make it the perimeter of the triangle (Figure 10), we again use Maple and a slightly modified path that includes adding the third side into the equation. Again, the behavior is as expected when δ goes to zero. The perimeter in this case went to the perimeter of the triangle without any error, $2 + 2\sqrt{2}$. However, the angles are not as expected. When minimizing the problem

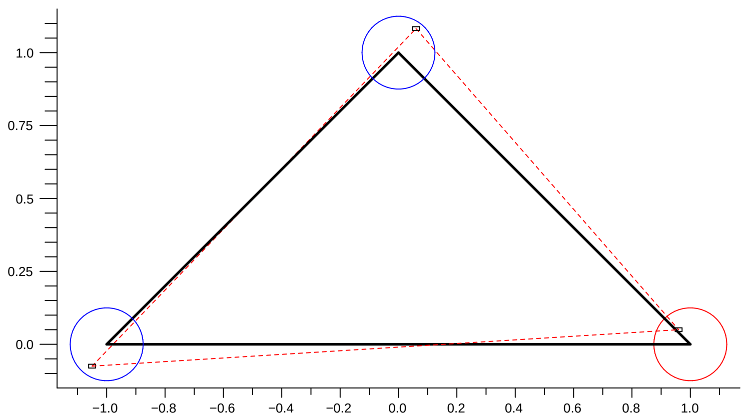

FIGURE 10. This time the path is of a closed triangle that has vertices at $(-1, 0)$, $(0, 1)$ and $(1, 0)$. The error circles around the vertices are each of radius δ and the angles are α, β, γ going from left to right.

and sending δ to zero, α went to 22.5 °, β went to -90 °, and γ went to 175.5 °. We expect the angles go to the same value as in the two-side path problem as that would put the angles as the same angles that connect the points. The same situation arises when trying to maximize the perimeter. When δ goes to zero, α goes to 202.5 °, β goes to 90 °, and γ goes to -22.5 °.

7.3. **The Semi-Circle Problem.** We will now look at estimating the perimeter of a semi-circle of radius 1 using a finite number of marks and adding straight line distances. We will first look at estimating the Semi-Circle using three points and determine the distance along that path. We will then add more data points and determine the distance along each of these paths. Note that using more data points should give a more accurate answer. Additionally, we show how many data points are needed to get an accurate perimeter measurement. This method is useful while using the GPS because we then have an idea of how many data points are needed for an accurate perimeter measurement. We will start with $\delta = 0.1$. As mentioned in the bar-bell section, we will treat this as taking two marks at each point. Our discussion will focus on the maximized bar-bells although it applies in both situations.

We begin by taking data points that are symmetrically placed about the semi-circle. When we use three data points (Figure 11), our path is that of a triangle. The three points are then connected using the bar-bell that maximizes the distance. This means that for each segment, we change where we put the error. We do this because two consecutive segments cannot have the maximum error for both lengths so we know that the length will be less than this. We also know that the length will be more than if we used to minimum error bar-bells. In Figure 11 we see the semi-circle that is being estimated as well as the bar-bells.

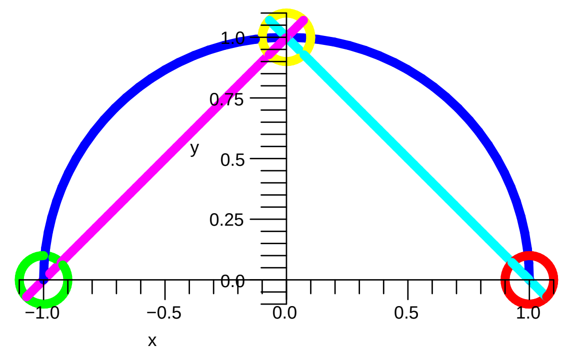

FIGURE 11. This is a picture of how three GPS points might be used to estimate a semi-circle using the bar-bell method.

While using this technique,as we increase the number of data points the error balls start to overlap. We see that Figure 12 is using 15 sample points and they are

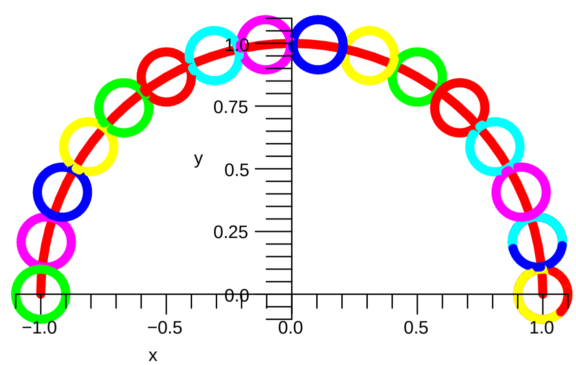

FIGURE 12. The error circles are just about to overlap when using 15 circles of radius 0.1.

almost overlapping. In Figure 13 the error circles have now begun to overlap. The minimum measured distance can now be zero because once the error circles start to overlap, the minimum distance between them is zero.

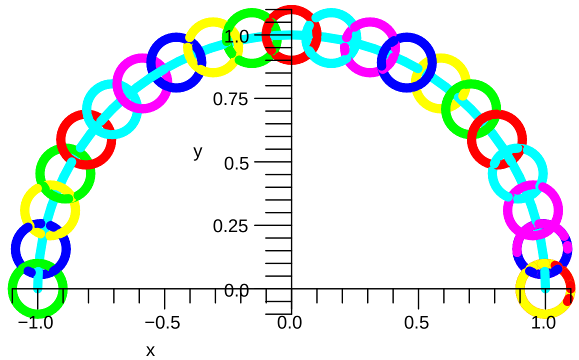

FIGURE 13. The error circles are now overlapping when using 20 circles of radius 0.1.

Number of Points	Min	Max
2	1.8	2.2
3	2.4284	3.2284
4	2.4	3.6
5	2.2615	3.8615
6	3.0	4.0917
7	1.2139	3.788
8	0.833	3.1833

TABLE 1. The estimated perimeter of the semi-circle using from 2 to 8 points. When more than two data points are used, the maximum values are above π and the minimum values are below π. The maximum values approach π while the minimum values go to zero.

7.4. **Length Measurements Using Bar-Bells.** Using the method discussed above, we have computed the estimated differences for both maximum and minimum bar-bells for 2 points up to 8 points. This data can be seen in table 1. The first column is the number of points, the second column is the minimum estimated distance, and the third column is the maximum estimated distance. Note that the maximum values are getting close to π while the minimum values look as if they are heading to zero. When eight points are taken, the figure that we are using is Figure 14.

We will now look at how changing δ impacts are results (Figure 15). This figure shows how the value of δ effects the measured length. The first column is the number of sample points used. As can be seen, the maximum values and minimum values are getting close to π.

7.5. **A Reasonable Value for δ.** One of the main reasons we use different values of δ is because $\delta = 0.1$ is a rather large δ when the actual distance is as small as

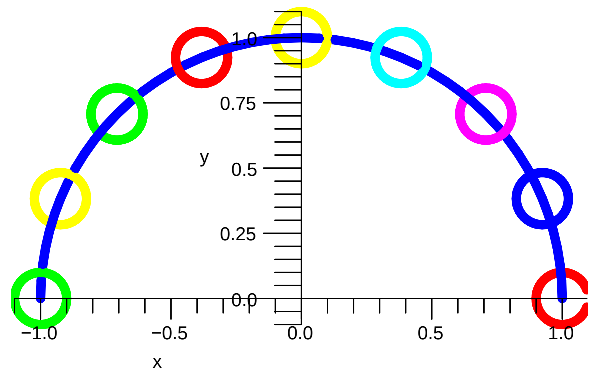

FIGURE 14. This is the the figure that is being used to estimate the arc-length of the semi-circle using eight sample points.

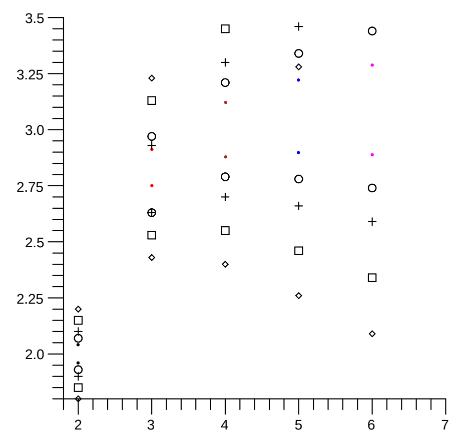

FIGURE 15. In this figure, δ is being varied. The $\Diamond$ is for $\delta = 0.1$. The $\Box$ is for $\delta = 0.075$. The $+$ is for $\delta = 0.05$. The $\circ$ is for $\delta = 0.035$. The $\cdot$ is for $\delta = 0.02$.

it is. In practice, common GPS measurements are of kilometer long baselines or measuring acre lots. As GPS is accurate to within 20 meters, we will use this to determine what size δ we use. To do this, we find what percentage the error is on an average measurement. Consider, for example, a square lot with sides of length 2 kilometers long using GPS. Since each measurement can be off by 20 meters, the total error could be at most 40 meters with the length being 2,000 meters. This results in at most a 2 % error. Although this is a good value for large distances, this value is very small for something like the semi-circle so at times we use large values of δ to see the effect of the error. For example, with the semi-circle approximation using a small value of δ is hard to see (Figure 16).

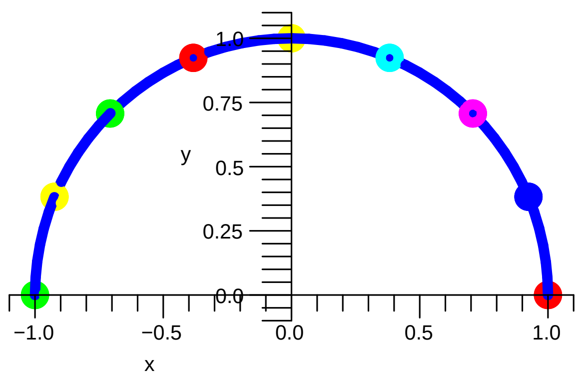

FIGURE 16. This is the semi-circle using 8 bar-bells with $\delta = 0.035$

7.6. **More General Paths.** Using a circle, it is best to take measurements after having traveled a fixed distance around the circle and taking a measurement (as in Figure 14). The next question is when to take measurements for a non-circular path. When the path is a straight line, only the end points need be measured to minimize error while getting the appropriate shape. However, when the path makes large bends or curves, intermediate points will need to be taken to appropriately measure this path distance. If we were to measure something like the path in Figure 17, it should be apparent that we would not want to take points every time we have travelled a certain distance on the path, but would want to take more data points where the path curved and less data points where the path was relatively straight. From calculus, we know that the curvature of an ellipse is

$$\frac{ab}{\left(b^2\left(\cos\left(\theta\right)\right)^2 + a^2\left(\sin\left(\theta\right)\right)^2\right)^{3/2}}. \tag{4}$$

We will now determine the total amount of curvature along different paths. By integrating curvature along the path we accumulate curvature. For example, Figure 18 shows the total curvature of a circle written in parametric form with the x-axis being the angle. This shows that the curvature of a circle is constant and that we would want to take points distributed equally around the circle to best estimate the circumference. As the circle is not that exciting of an example, we move to looking at the curvature of different ellipses. When the ratio of semi-major axis to semi-minor axis is large, the plot looks like Figure 19. This shows that when the ellipse is making a sharper curve (near the left end and the right end),

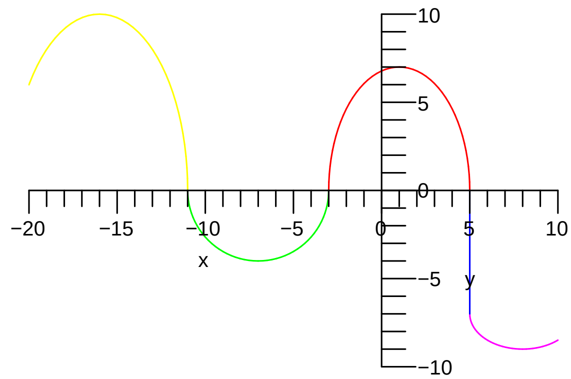

FIGURE 17. This is a path in which we would want to take more measurements around points with higher curvature. For example, we would want more points around $x = 0$ and less points around $x = 5$

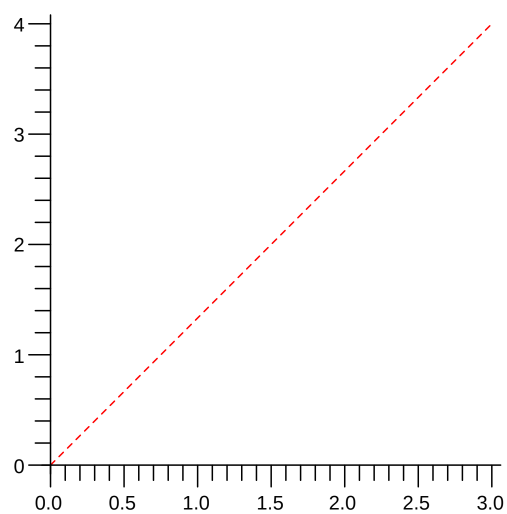

FIGURE 18. This shows how curvature accumulates as we travel from the start of a circular path to its end.

the curvature is accumulating very rapidly, and when the ellipse is more flat the curvature accumulates slowly. As mentioned above, to be more accurate, we want

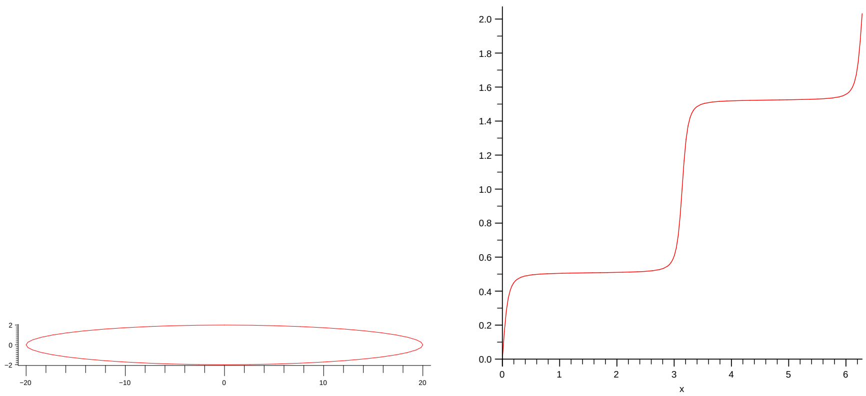

FIGURE 19. This is a graph of an ellipse with a ratio of 10-1 semi-major to semi-minor axis alongside the total curvature of the ellipse.

to take more data points where the curvature is large and fewer data points where the curvature is small. A convenient way to do this is by looking at a graph of curvature. Looking at the y-axis of a plot of curvature versus angle, each increment shows a change in curvature that is the same for every increment. That is, if we go from 2 to 2.5 on the y-axis, our change in curvature is the same as if we went from 0.75 to 1.25. Using this idea, we break up the ellipse into 22 segments that each have the same curvature. The ellipse used is shown in Figure 20.

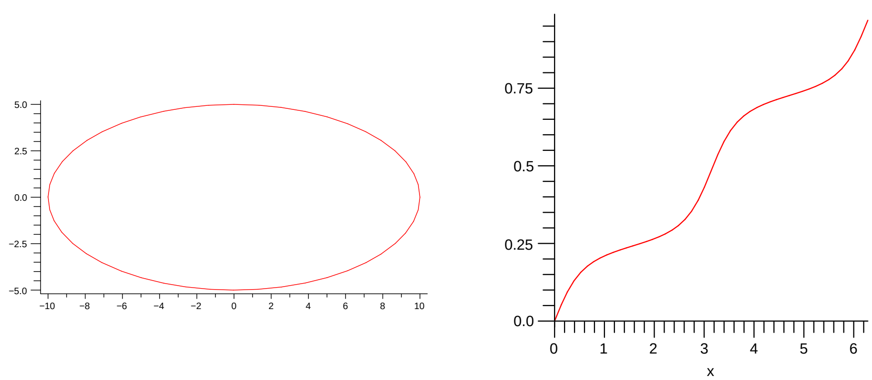

FIGURE 20. This is a graph of the ellipse with a ratio of 3-1 semi-major to semi-minor axis alongside the total curvature of the ellipse. This ellipse used an example of estimating the perimeter using the curvature technique.

The ellipse is divided into equal segments by going up on the curvature plot by 0.02 and recording the angle that corresponded with that curvature. The angles are then used to find the locations of the sample points. Next, find the distance between each of these points and add them up to get an approximation of the perimeter. Using this technique, the perimeter is 132.56. The actual perimeter of this ellipse is 133.65, so the approximation is close.

Another technique to determine when to take measurements is to take data points after travelling through a certain angle, and using these data points to measure the distance travelled. When 22 points are used with this technique, the increment for θ is $\frac{\pi}{11}$ radians, and the perimeter is 133.2 which is closer to the actual perimeter. However, this technique would not be the best for use with GPS because it is hard to know when we have gone through a specific angle.

The final technique we will look at is to go a fixed distance on the ellipse and then take a data point. This way the data points are distributed equally about the ellipse. It is expected that the curvature method gives a more accurate solution with the same number of data points. The total perimeter of the ellipse is divided by the number of segments used to find the distance travelled between data points. To find the arc length of a segment of the ellipse we need to use elliptical integrals. We use Maple to do this. When 22 data points are used, the perimeter is 132.6. We note that this is slightly better then the curvature approach. However, this might be due to the fact that the number of data points is pretty large, so we will decrease the number of data points and observe what happens. This data is contained in Figure 21.

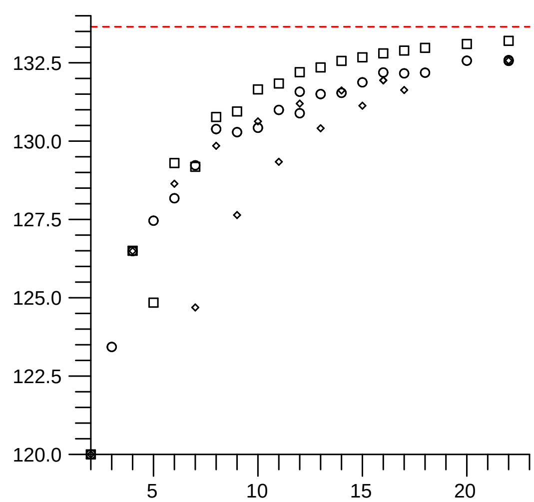

FIGURE 21. This plot shows the estimated perimeter versus the number of path segments used in this approximation. One set of data points,○, is using the curvature approach, another set, ♢, is using the fixed path length approach, and the last set ,□, is using the fixed angle approach. The dashed line is the actual perimeter of the ellipse.

The graph in Figure 21 shows that the fixed angle approach is the best. Additionally, the curvature approach is better than the fixed path length approach.

However, the fixed path length approach will be the easiest method to use with typical GPS measurements. This plot also shows how the measured distance oscillates depending on whether an even number or an odd number of path segments are used. While using the fixed path length method, as the number of path segments increases from even to odd, the path length goes down, but then is back up for the next even number of segments. This is because an odd number of segments is not going to include a point at one end of the semi-major axis while an even number of path segments will. When using less than 5 data points, the fixed angle approach is the worst and also has the biggest variation when increasing from one data point to the next when using less than 5 data points. The fixed path length approach is better when using an even number of data points as opposed to an odd number of data points, this is best seen in Figure 22. Figure 22 shows each estimation technique plotted by itself to see the data better.

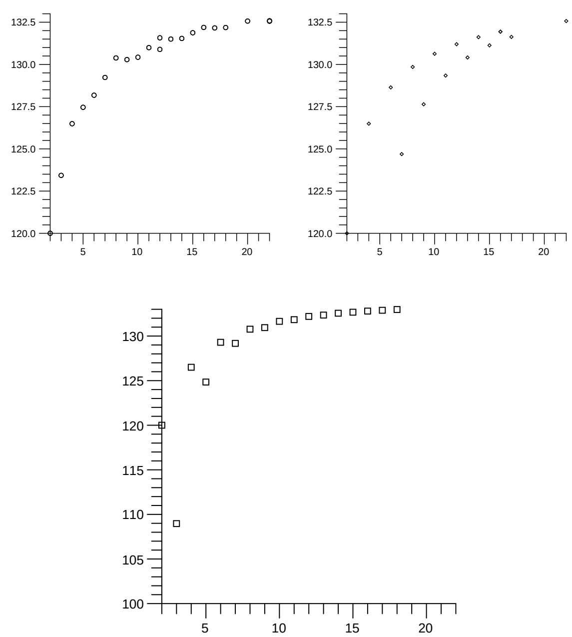

FIGURE 22. The plot with ○ is the curvature approach. The plot with ◊ is the fixed path length approach. The plot with □ is the fixed angle approach.

Using Figure 21, it is seen that it is best to use the fixed curvature approach when using less than 5 data points. When using between 5 and 20 data points, the fixed angle approach is best, then the fixed curvature approach, and then the fixed path length approach. When using more than 20 data points, there is not very much difference between the fixed path length and the curvature approach. Therefore, it is best to use the fixed angle approach when approximating an ellipse unless using less than 5 data points when the curvature approach is better. When using an even number of data points, there is not much difference between the fixed path length and the curvature approaches. A reason for this is that the only time there is a big change in curvature, the distance must be small so it really does not matter in the end. When using GPS to measure a path that is somewhat elliptical, the fixed path length approach would typically be the most efficient.

8. Conclusion

In this paper, we focused on two main ideas. We first looked at the Global Positioning System in detail. The configuration of the system was discussed, as were the uses and applications of GPS. The other main idea was in determining the best positions to take data points to measure a path length or a perimeter. When we know where to take measurements to best estimate an object, this information can be applied to taking way-points with a GPS to get accurate and efficient path lengths.

There are many ideas that were discussed which can still be expanded. The largest problem to be worked on is in determining how many data points to take and where to take these data points when given a particular path shape. To do this, many different shapes will need to be examined and it should be determined which technique is the best to estimate the path of that figure. The ideas and methods applied to specific fundamental shapes could be used collectively to help determine where to take data points for more complicated shapes. Additionally, bringing GPS back into the problem and estimating perimeters and determining experimentally which measurement method is the most accurate would see if the results are applicable to using GPS to measure distances.

References

[1] Introduction to GPS: The Global Positioning System. Ahmed El-Rabbany. Artech House, Boston. 2002

[2] Global Positioning System: Theory and Practice. Bernard Hofmann-Wellenhof and Herbert Lichtenegger. Adolf Holzhausens Hachfolger, Austria. 1993.

[3] GPS for Land Surveyors. Jan Van Sickle. Ann Arbor Press, Chelsea, Michigan. 2001

[4] `http://tycho.usno.navy.mil/gpscurr.html` Here I found the current configuration of the orbital satellite configuration.