GPS Basics

Using GPS signals to find where you are

JONATHAN OLDS

jontio@i4free.co.nz

http://jontio.zapto.org

Contents

Chapter 1

GPS basics

Here we give a breif introduction to the basics of legacy GPS signals and how code based position solutions can be obtained using them. This is the standard observable that is used by consumers to obtain postion solutions. For compleateness we also mention the phase observable but do not attempt to describ how such an observable can be used.

1.1 GPS signal

As of writing GPS is currently undergoing a modernization to improve both civilian and military use. Between 1990 and 2004 legacy satellites were launched while from 2005 modernized satellites have been launched. According to the National Coordination Office for Space-Based Positioning, Navigation, and Timing [2] this is in an effort to upgrade the features and performance of GPS. Currently GPS transmits on three different RF links from the satellites to end-users. The RF links are called L1, L2 and L5 and are named after the bands that they transmit in. *Code Division Multiple Access* (CDMA) is used as the channel access method so all satellites used the same carrier frequencies. L1 has a nominal frequency of 1575.42Mhz as seen from Earth, L2 1227.60Mhz and L5 1176.45Mhz. These nominal frequencies are modulated with various signals to aid navigation. The current GPS modernization consists of generally improving the hardware as well as adding more signals that are sent over the RF links. The GPS modernization currently underway will take many years and satellites producing signals such as

L1C on L1 are not expected to be launched until 2016 with 24 satellites expected by around 2026 [4]. We restrict our investigation here to that of legacy L1 signals.

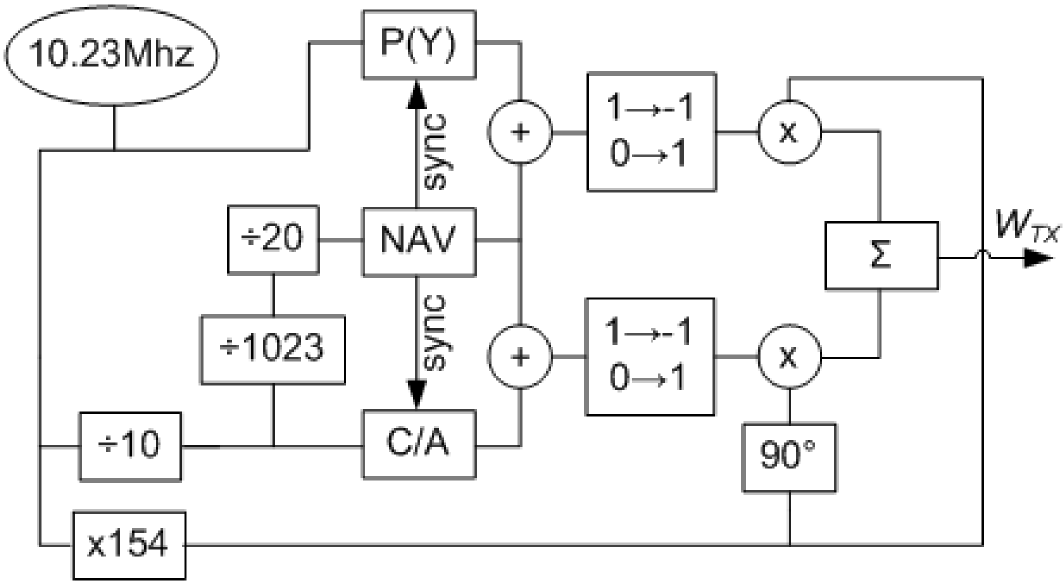

Figure 1.1: Legacy L1 signal generation block diagram

GPS legacy signals on L1 consist of a coarse acquisition *Pseudo Random Number* (PRN) code called C/A sent at 1.023Mb/s, a precise PRN code P sent at 10.23Mb/s which is called Y if encrypted, and navigation data called NAV sent at 50b/s. The two PRN codes are unique to each satellite and each spread the navigation signal. Bits of the PRN codes are also called chips. L1 consists of in-phase and quadrature components. Each component is separately modulated using *Binary Phase Shift Keying* (BPSK) as their modulation technique. One modulator is supplied with a bit train from modulo-2 addition of P(Y) and NAV while the other is supplied with a bit train from modulo-2 addition of C/A and NAV. The signals and the carrier frequencies are derived from a single time source and synchronized with one another thus causing the signals to be bit phased with one another in addition to the signals being synchronized with one another at a higher level. Figure 1.1 depicts a block diagram of the generation of such signals. The instantaneous L1 wavefront a satellite sends $W_{TX}(t)$ at a time t, can be written as follows assuming the satellite clock keeps perfect frequency.

$$W_{TX}(t) = A_{TX}\Re\left\{S\left(t+\Delta T_{TX}(t)\right)e^{2\pi i\left(tf_{TX}+\phi_0^{TX}\right)}\right\} \quad (1.1)$$

Where $S(t) = NAV(t)\left(P(Y)(t)\,e^{i0} + C(t)\,e^{i\pi/2}\right)$ is the composite signal, $C(t)$

C/A code, $P\left(Y\right)\left(t\right)$ P(Y)-code, $NAV\left(t\right)$ navigation data, ϕ_0^{TX} satellite oscillator phase at time zero, A_{TX} transmission amplitude, ϕ_0^{TX} satellite oscillator phase at time zero, f_{TX} nominal satellite oscillator frequency as seen from Earth so as to account for general relativistic effects, f_0 1575.42Mhz as seen from Earth, and τ_0^{TX} is the satellite composite signal offset at time zero. $C, P\left(Y\right), NAV \in \{-1, 1\}$ and are functions so as to produce the correct composite signal where a mapping of $0 \rightarrow 1$ and $1 \rightarrow -1$ has been applied. At the receiver the instantaneous L1 wavefront a receiver receives $W_{RX}\left(t\right)$ at a time t, can then be written as follows assuming no hindrance by the atmosphere.

$$W_{RX}\left(t\right) = A_{RX}\Re\left\{S\left(t + \Delta T_{TX}\left(t\right) - \Delta t\left(t\right)\right)e^{2\pi i\left(\left(t-\Delta t(t)\right)f_{TX}+\phi_0^{TX}\right)}\right\} \tag{1.2}$$

Where A_{RX} is reception amplitude and $\Delta t\left(t\right)$ is the transmission flight time from the satellite at transmission time to the receiver at reception time t. It's important to realize that $\Delta t\left(t\right)$ is not the flight time between the satellite and receiver at time t, but rather the flight time based as the receiver sees it. It is similar to when a airplane passes by and one hears the sound of the plane lagging where the plane actually is. The flight time of the sound from the plane as determined by the listener is different from the flight time one would get from calculating where the plane actually is to the user with respect to the same reception time.

1.2 What the GPS receiver does to the received $W_{RX}(t)$ signal

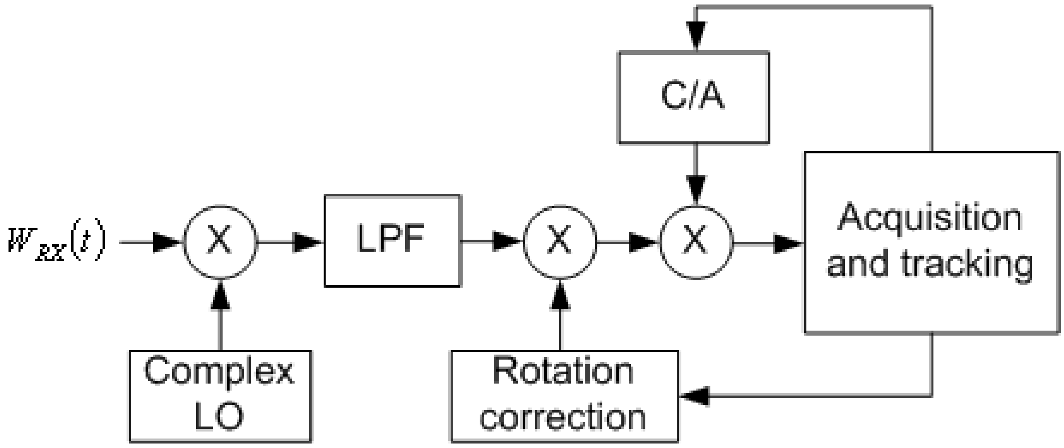

Figure 1.2: Simplified version of a possible GPS satellite acquisition and tracking scheme

Figure 1.2 shows a simplified version of what a GPS satellite receiver can do to acquire the satellite and track the signals that it produces. Upon reception the receiver tries to separate the signals from L1 and track the C/A code and the carrier phase. This can effectively be performed by mixing $W_{RX}(t)$ with a complex local oscillator $LO(t) = e^{2\pi i\left(tf_{RX}+\phi_0^{RX}\right)}$ where f_{RX} is the frequency of the oscillator and ϕ_0^{RX} the phase of the oscillator at time zero, filtering using a low pass filter LPF, then phase and/or frequency tracking to stop rotation along with correct C/A timeing to match the C/A code as sent by the satellite so as to acquire access to the data being sent by the satellite.

1.2.0.1 Obtaining the baseband signal $R(t)$

After mixing with the local oscillator and filtering using the low pass filter, the receiver obtains the following baseband, where, $\Delta f = f_{RX} - f_{TX}$ and $\Delta\phi_0 = \phi_0^{RX} - \phi_0^{TX}$.

$$\begin{aligned} R(t) &= \frac{2}{A_{RX}} LPF\left\{W_{TX}(t)\, LO(t)\right\} \\ &= S\left(t + \Delta T_{TX}(t) - \Delta t(t)\right) e^{2\pi i\left(t\Delta f + \Delta t(t) f_{TX} + \Delta\phi_0\right)} \end{aligned} \tag{1.3}$$

We define $t\Delta f + \Delta t\left(t\right) f_{TX} + \Delta\phi_0$ as the received beat carrier phase (carrier phase).

1.2.0.2 Stopping rotation

We see that this is a constellation of four points that rotates due to the frequency difference between the receiver's local oscillator and the satellite's oscillator, and also rotates due to the radial motion of the satellite itself with respect to the receiver.

If we let $\Phi\left(t\right) = t\Delta f + \mathit{\Delta} t\left(t\right) f_{TX} + \Delta\phi_0$ then we can correct for rotation by multiplying $R\left(t\right)$ as follows.

$$\begin{aligned} R\left(t\right) e^{-2\pi i(\Phi(t))} &= S\left(t + \Delta T_{TX}\left(t\right) - \Delta t\left(t\right)\right) e^{2\pi i(\Phi(t))} e^{-2\pi i(\Phi(t))} \\ &= S\left(t + \Delta T_{TX}\left(t\right) - \Delta t\left(t\right)\right) \end{aligned} \tag{1.4}$$

This stops the constellation from rotating and removes any constant constellation rotation offset. This then resolves the composite signal. Hence we define $\Phi\left(t\right)$ as the estimated received beat carrier phase by the receiver (estimated carrier phase). The constellation's phase (or equivalently the constellation's rotation offset) is defined as the difference between the carrier phase and the estimated carrier phase $t\Delta f + \mathit{\Delta} t\left(t\right) f_{TX} + \Delta\phi_0 - \Phi\left(t\right)$. More generally as long as the estimated carrier phase is in phase with the carrier phase the constellation stops rotating and there is no constant constellation rotation offset.

1.2.0.3 C/A Code alignment

After the constellation rotation has been stopped by letting the estimated carrier phase be in phase with the carrier phase, a local replica of the C/A code has to be mixed with the composite signal and phase shifted in time until the local replica of the C/A code is in phase with the one that is in the received composite signal. We then define $\tau\left(t\right)$ as the C/A code alignment offset and is how much the incoming C/A is misaligned with the local replica. We define the local replica of the C/A code as $LCA\left(t\right)=C\left(t+\Delta T_{RX}\left(t\right)\right)$ where $\Delta T_{RX}\left(t\right)$ is the receiver's clock offset. First we notice that if we offset the local C/A replica by $\tau\left(t\right)$ and let $\tau\left(t\right)=\Delta t\left(t\right)+\Delta T_{RX}\left(t\right)-\Delta T_{TX}\left(t\right)$ and multiply this offseted C/A with $R\left(t\right)$ we expect the following for a random time t.

$$
\begin{aligned}
& E\left[LCA\left(t-\tau\left(t\right)\right)R\left(t\right)e^{-2\pi i\left(\Phi\left(t\right)\right)}\right] \\
& = E\left[LCA\left(t-\tau\left(t\right)\right)S\left(t+\Delta T_{TX}\left(t\right)-\Delta t\left(t\right)\right)\right] \\
& = E\left[C\left(t+\Delta T_{TX}\left(t\right)-\Delta t\left(t\right)\right)S\left(t+\Delta T_{TX}\left(t\right)-\Delta t\left(t\right)\right)\right] \\
& \approx iE\left[NAV\left(t+\Delta T_{TX}\left(t\right)-\Delta t\left(t\right)\right)\right]
\end{aligned}
\tag{1.5}
$$

This is due to the fact that $P(Y)$ and C are not well cross-correlated, while of course C is perfectly correlated with itself.

$LCA\left(t-\tau\left(t\right)\right)R\left(t\right)e^{-2\pi i\left(\Phi\left(t\right)\right)}$ is the input to the acquisition and tracking block in figure 1.2. Visually $LCA\left(t-\tau\left(t\right)\right)R\left(t\right)e^{-2\pi i\left(\Phi\left(t\right)\right)}$ is a constellation without any rotation of four points. The two points lying on the imaginary axis move slowly at no more than 50 times a second and contain the navigation data, while, the two points that lie on the real axis move very rapidly at up to 10.23 million times a second in a seemingly random way with a mean value of zero. More generally when this happens the incoming C/A code is aligned with the local replica and the carrier phase is in phase with the estimated carrier phase.

1.3 Acquisition, tracking and NAV data extraction

1.3.0.1 Extraction of NAV data using filtering

When the incoming C/A code is aligned with the local replica and the carrier phase is in phase with the estimated carrier phase, navigation data can easily be obtained simply by using a low pass filter

$$NAV\left(t+\Delta T_{TX}\left(t\right)-\Delta t\left(t\right)\right)=-iLPF\left[LCA\left(t-\tau\left(t\right)\right)R\left(t\right)e^{-2\pi i\left(\Phi\left(t\right)\right)}\right] \tag{1.6}$$

This allows the GPS receiver access to the satellite's navigation data which includes a wealth of data including the satellite's estimate of $\Delta T_{TX}\left(t\right)$ and time of transmission of specific navigation data transitions. Along with data to calculate the satellites positions.

1.3.0.2 A metric for C/A code alignment and rotation

We restrict t to be within a small period of time 2δ which is less than NAV's period while still ensuring a long enough period such that $P(Y)$ and C are still not well correlated over that period. To find such a time period is possible as NAV's period is 204600 times longer than $P(Y)$'s period. We assume the C/A codes are perfectly alinged and the the carrier phase is in phase with the estimated carrier phase. We then expect the following.

$$E\left[LCA\left(t-\tau\left(t\right)\right)R\left(t\right)e^{-2\pi i\left(\Phi\left(t\right)\right)}\right]\approx iNAV\left(t+\Delta T_{TX}\left(t\right)-\Delta t\left(t\right)\right) \tag{1.7}$$

$$\begin{aligned}\therefore NAV&\left(t+\Delta T_{TX}\left(t\right)-\Delta t\left(t\right)\right)\\&\approx\frac{-i}{2\delta}\int_{k=t-\delta}^{k=t+\delta}LCA\left(k-\tau\left(k\right)\right)R\left(k\right)e^{-2\pi i\left(\Phi\left(k\right)\right)}dk\end{aligned} \tag{1.8}$$

As the navigation data has a constant magnitude of 1, any imperfections in the correlation between $LCA\left(t-\tau\left(t\right)\right)$ and $C\left(t+\Delta T_{TX}\left(t\right)-\Delta t\left(t\right)\right)$ due to incorrectly estimating $\tau\left(t\right)$ will affect the magnitude of the acquired navigation data. In addition, incorrectly estimating $\Phi\left(k\right)$ causing a rotating constellation motion, will, after integration also negatively affect the magnitude of the acquired navigation data. Therefore, we can say the following where γ is defined as the correlation coefficient between a local time shifted replica of the C/A code and the one being received that may be rotating.

$$\gamma = \frac{-1}{4\delta^2}\left(\int_{k=t-\delta}^{k=t+\delta} LCA\left(k-\tau\left(k\right)\right)R\left(k\right)e^{-2\pi i(\Phi(k))}dk\right)^2 \tag{1.9}$$

This coefficient more generally can be calculated more accurately with averaging; therefore a GPS receiver could calculate it as follows.

$$\gamma = \frac{-1}{4n\delta^2}\sum_{q=1}\left(\int_{k=t_q-\delta}^{k=t_q+\delta} LCA\left(k-\tau\left(k\right)\right)R\left(k\right)e^{-2\pi i(\Phi(k))}dk\right)^2 \tag{1.10}$$

Squaring eliminates the sign of the BPSK. The magnitude of γ is related to how well the C/A codes are aligned and how well the rotating motion of the constellation has been stopped, while the angle of γ is related to the constellation's constant rotation offset with a half cycle ambiguity. A constant constellation rotation offset does not affect the magnitude of γ as a constant rotation offset is just a constant that can be taken out of the integral. Therefore, gamma will be maximized when the C/A codes are perfectly aligned and the constellation is not rotating while it is invariant for constant rotation offset of the constellation.

- $|\gamma|$is maximized when C/A codes are aligned and the constellation is not rotating.

Treating Φ and τ as variables a GPS receiver can vary Φ and τ to maximize γ. If γ is above a certain threshold the GPS receiver can assume that the satellite is acquired and to commence tracking τ, Φ, and decoding NAV data.

We are interested in maximizing γ because when it is a maximum, with the addition of some ambiguity both Φ and τ are good estimates for $t\triangle f + \Delta t\left(t\right)f_{TX} + \triangle\phi_0$ and $\Delta t\left(t\right) + \Delta T_{RX}\left(t\right) - \Delta T_{TX}\left(t\right)$ respectively which turns out to be useful in finding positions solutions. In addition when γ is maximized we are able to obtain navigation data which is also useful for finding position solutions.

1.3.0.3 First-order linear approximations of unknown functions τ and Φ

We wish to maximize γ , the correlation coefficient. Due to the surface shape of the magnitude of γ with respect to Φ and τ initial estimates for both Φ and

τ are required. Without good initial estimates, γ is dominated by noise making standard tracking schemes such as *Phase Locked Loop* (PLL), *Frequency Locked Loop* (FLL), *Delay Locked Loop* (DLL) and early/late time useless. It's like trying to track an ant crawling in long grass; you have to find it first before you can track it as the grass makes it difficult to see the ant from afar.

First we create a first order linear approximation model of how Φ and τ change. We have already seen that when $\Phi(t) = t\triangle f + \mathit{\Delta} t(t) f_{TX} + \triangle\phi_0$ and $\tau(t) = \Delta t(t) + \Delta T_{RX}(t) - \Delta T_{TX}(t)$ we are able to stop rotation and align our local C/A code replica with the incoming one. Therefore, these are the Φ and τ that we are looking for. Linear approximations of these two equations are written below where $F(t_m) = (\Delta f + \Delta f_{TX}(t_m))$, $\Delta f_{TX}(t_m)$ is the change of frequency of f_{TX} due to Doppler at time t_m, where a positive value is for the satellite moving away from the receiver and $\theta(t_m)$ and $\mathit{\Xi}(t_m)$ are some constants. Proofs can be found in 1.5.1 and 1.5.2.

$$\Phi(t) \approx tF(t_m) + \theta(t_m) \tag{1.11}$$

$$\tau(t) \approx (t - t_m) F(t_m) \frac{-1}{f_0} + \mathit{\Xi}(t_m) \tag{1.12}$$

These approximations are only valid if $F(t_m)$ does not change to rapidly around time t_m. The maximum rate at which velocity will change is about $0.1178\, ms^{-2}$ and is when the satellite is directly overhead (1.5.4). On the L1 band this implies that the constellation rotation speed will change by less than about $0.9\, Hzs^{-1}$ if $F(t_m)$ is left unchanged in equation 1.13 (see 1.5.3 and 1.5.4). Now compare this to the range of $F(t_m)$. $\Delta f_{TX}(t_m)$ can be as large as about $\pm 5\, kHz$ (1.5.4), and depending on the receiver clock accuracyΔf could be out by another $5\, kHz$ if we assume a receiver clock accuracy of $3.5\, ppm$. This means the range of $F(t_m)$ is in the order of $20\, kHz$. If we restrict our time of interest to $1\, ms$, then, $F(t_m)$ will change by less than $0.0009\, Hz$ which is far less than the range of the $20\, kHz$ of $F(t_m)$. Therefore, for short periods of time this is a valid approximation.

Equations 1.11 and 1.12 form an approximate model of how Φ and τ will change in the short term.

1.3.0.4 Further simplifications to the first-order linear approximations of unknown functions τ and Φ in regard to acquisition

In figure 1.2 the input of the acquisition and tracking block as we have already seen is written as in equation 1.13.

$$LCA\left(t-\tau\left(t\right)\right)R\left(t\right)e^{-2\pi i\left(\Phi\left(t\right)\right)} \tag{1.13}$$

To acquire Φ and τ initially we can make further simplifications to equations 1.11 and 1.12. $\theta\left(t_m\right)$ in equation 1.11 when placed in equation 1.13 has no effect on changing constellation rotation with respect to time. 1.13 will still be a non-rotating four-point constellation but just with a constant rotation offset of $\theta\left(t_m\right)$ cycles. As we have already mentioned in equation 1.10, any constant rotation offset has no effect on the magnitude of the correlation coefficient. Therefore we can ignore $\theta\left(t_m\right)$ when initially acquiring Φ.

The carrier wave frequency is 1540 times greater than the bit rate of the C/A code. As frequency times time is phase, if $F=10,000\,Hz$ then it takes $0.025\,ms$ for Φ to change by a quarter of the cycle, while it takes $38.5\,ms$ for the C/A code to change by a quarter of a chip. If we then assume a digitalization of R at the rate of $4.092\,Mb/s$ and sampling $1\,ms$ worth of R, more often than not we couldn't even detect the difference between $F=0$ and $F=10,000$ in the C/A code directly while it would be easy to detect in the carrier wave.

Due to these two points we make the following two approximations when considering initial acquisition of Φ and τ. Here we acknowledge that the constellation will have an arbitrary constant rotation offset, are only valid for periods of time of a few milliseconds, and baseband sampling rate is no more than a few times per chip.

$$\Phi\left(t\right)\approx tF\left(t_m\right) \tag{1.14}$$

$$\tau\left(t\right)\approx \Xi\left(t_m\right) \tag{1.15}$$

1.3.0.5 Acquisition

With these two approximations 1.14 and 1.15 the receiver can do a two-dimensional search, F with a frequency dimension and the other Ξ with a time dimension to find the point that maximizes the correlation coefficient in equation 1.10. Assuming a receiver clock accuracy of $3.5\,ppm$, the receiver would have to search from $-10\,kHz$ to $10\,kHz$. As the C/A code is a periodic function with a period of 1 ms, the receiver would have to search from 0 ms to 1 ms.

Estimating F and Ξ by trying to maximize the correlation between the local C/A code replica with the incoming one turns out to be computationally demanding using more energy than tracking, and is a major concern for GPS receivers. Because of this much research has been directed towards this problem to reduce the computational effort to estimate these two parameters [7]. A parallelized 2D search by using *Fast Fourier Transform* (FFT) is a conventional method currently used in software defined receivers [7].

As an example we cross-correlated a local C/A code replica with an incoming one on L1. We used a sample rate of $4.092\,Mb/s$ being four times the nominal frequency of the chip rate and searched by varying the frequency term by $\pm 10\,kHz$ in steps of $125\,Hz$ and then using cross correlation varying the time term by no more than $\pm 511.5\,chips$ ($\pm 511.5\,chips$ covers the entire $1\,ms$). Navigation data was simulated by using random data, while no P(Y) data was added. An offset of $5000\,Hz$ was applied to the $1.57542\,GHz$ carrier frequency and a phase offset of $250.15\,chips$ was applied to the C/A code of the incoming signal. Using $1\,ms$ worth of sequential incoming data, figure 1.3 was obtained.

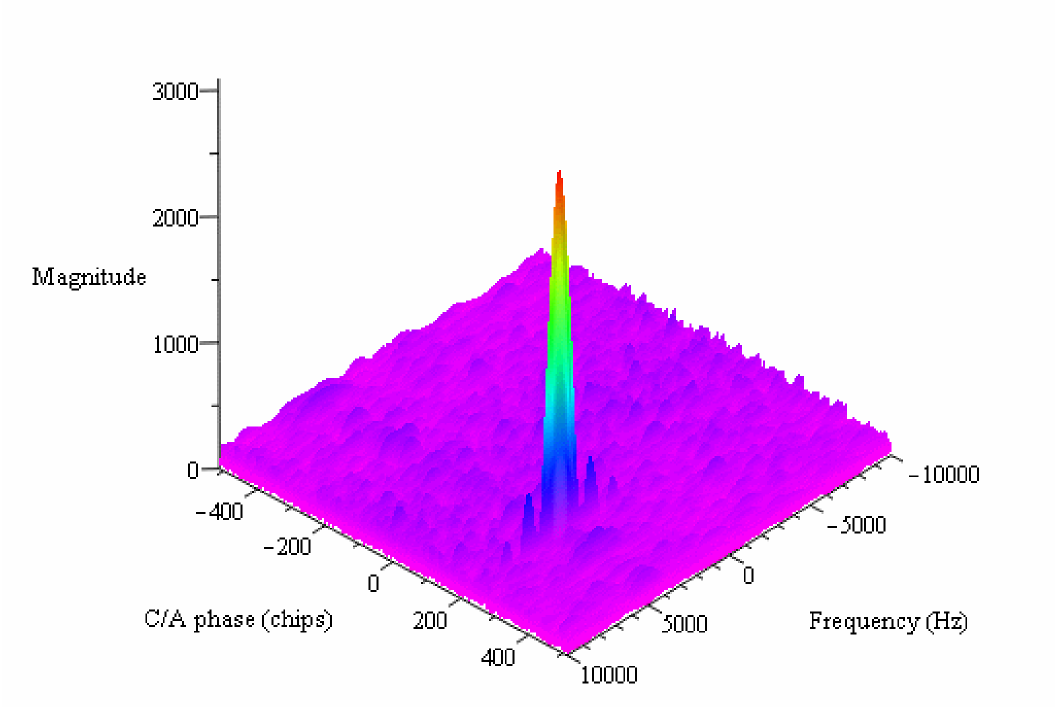

Figure 1.3: Correlation coefficient versus phase and frequency offsets. 1 ms of sequential data of C/A PRN 16 cross-correlation. 4.092Mb/s sampling rate. 16 kb of data, 658,651 evaluation points.

As can be seen there is a clear peak in the graph representing the estimates of F and Ξ . The estimated phase offset using interpolation around the highest peak was $250.18\,chips$ while that of the frequency was $5000\,Hz$; this matches well with the exact values. From this figure it is clear to see why tracking will not work without good initial estimations of τ and Φ. It's interesting to note in passing that the 0.03 chips that the interpolation value was out equates to the time light takes to travel 9m. This is roughly the correct order of accuracy cheap consumer grade GPS receivers have.

The number of points needed to be evaluated in figure 1.3 was 658,651.While less points could have been used and still be able to obtain reasonable estimates for F and Ξ , by the 2-D search method there are inherently always going to be a large number of points needed to be evaluated. While other methods such as the parallelized 2D search by using FFT exist and are less computationally demanding and hence less energy demanding, acquisition to the best of our knowledge still uses more energy than tracking. As an example the Ublox NEO-7N which is a modern consumer grade GPS receiver uses 23% more energy during its acquisition state than it's tracking state [3].

1.3.0.6 Tracking

Once the acquisition has been performed and the point of the maximum correlation coefficient has been found, the receiver can then track τ and Φ using standard techniques such as PLL, FLL, DLL and early/late tracking methods. The linear approximations do not stop the maximum correlation coefficient point from moving but it slows it down sufficiently that one can treat it as a stationary point until it has been found then one can simply track it as it moves.

From acquisition, estimates for F and $\varXi$ are obtained, which via equation 1.12 gives an estimate for $\tau(t)$. To track $\tau(t)$ then early/late tracking can be used. Such a method usually consists of three locally produced C/A replicas, one slightly ahead of what is expected from the satellite, one as expected from the satellite, and one slightly behind what is expected from the satellite.

$$\text{Early } LCA\left(t-\tau(t)+\xi\right)$$

$$\text{Prompt } LCA\left(t-\tau(t)\right)$$

$$\text{Late } LCA\left(t-\tau(t)-\xi\right)$$

These three C/A replicas are then each correlated with $R(t)\,e^{-2\pi i(\Phi(t))}$ to produce three correlation coefficients (γ_E ,γ_P, γ_L) and using interpolation a new estimate for $\tau(t)$ can be obtained to keep the code aligned. Keeping the code aligned is the one of the two requirements for maximizing γ.

$LCA\left(t-\tau(t)\right)R(t)$ has the effect of removing the C/A code from R. This removal is called wiping the code. Once it is removed a carrier tracking scheme such as a PLL can be used on $LCA\left(t-\tau(t)\right)R(t)$ because in one direction it appears as a standard BPSK signal. A costas PLL could be performed on $LCA\left(t-\tau(t)\right)R(t)$ to estimate $\Phi(t)$ as it is invariant to the navigation transitions. Assume that we designed the costas loop to align on the imaginary axis. Then, the costas loop will align NAV's BPSK signal along the imaginary axis with an ambiguity as to which way around it is aligned. The costas loop will also stop the constellation from rotating which is one of the two requirements for maximizing γ.

So, using code tracking and carrier tracking simultaneously the maximum point of correlation can be continuously tracked. It is not sufficient for a receiver solely to track only one of $\tau(t)$ or $\Phi(t)$; both need to be tracked simultaneously.

We have already seen that in 1.3.0.2 $\tau(t) = \Delta t(t) + \Delta T_{RX}(t) - \Delta T_{TX}(t)$ implies maximum correlation. Due to the $1\,ms$ C/A ambiguity the converse is not true. Therefore we can say that maximum correlation implies $\tau(t) = \Delta t(t) + \Delta T_{RX}(t) - \Delta T_{TX}(t) + M/1000$ where M is some fixed integer.

Likewise $\Phi(t) = t\triangle f + \mathit{\Delta} t(t) f_{TX} + \triangle\phi_0$ implies maximum correlation but due to carrier phase cycle ambiguity and that gamma is maximized for any constant rotation offset the converse is not true. The costas loop removes the constant rotation offset with an ambiguity of half a cycle, and if the PLL accumulates its phase offset rather than resetting it as it passes through an angle of zero, then maximum correlation implies $\Phi(t) = t\triangle f + \mathit{\Delta} t(t) f_{TX} + \triangle\phi_0 + N/2$ for some fixed integer N.

τ and Φ when these ambiguities are considered become the two observables used by almost all low end consumer grade GPS receivers for position solution calculations.

1.4 Observables

Observables are measurements taken by the GPS receiver of quantities that the GPS receiver can directly measure. Observables do not directly tell you where the GPS receiver is situated but with using various techniques will allow you to calculate position solutions that do tell you where the GPS is situated. The two observables we consider are the code observable and the phase observable.

We have seen by tracking the maximum point of γ using early late timing and a costas PLL that we have found $\tau(t) = \Delta t(t) + \Delta T_{RX}(t) - \Delta T_{TX}(t) + M/1000$ for some fixed integer M and $\Phi(t) = t\triangle f + \mathit{\Delta} t(t) f_{TX} + \triangle\phi_0 + N/2$ for some fixed integer N. These are the code observable and the phase observable respectively so far. However, there are some added complications and the form they take can differ.

1.4.1 The code observable

Once γ is tracked the receiver has access to the navigation data. The navigation data is sent as 30 bits per word. There are 10 words in a subframe taking 6 seconds to transmit. Each subframe contains a *Hand Over Word* (HOW) word that indicates the exact time when the leading-edge of the first bit of the navigation data was transmitted from the satellite. In the satellite the first bit of every NAV data transition is aligned to first chip of the the C/A code. This is possible as they are derived from the same oscillator (see 1.1). Figure 1.4 shows the C/A NAV timing relationship.

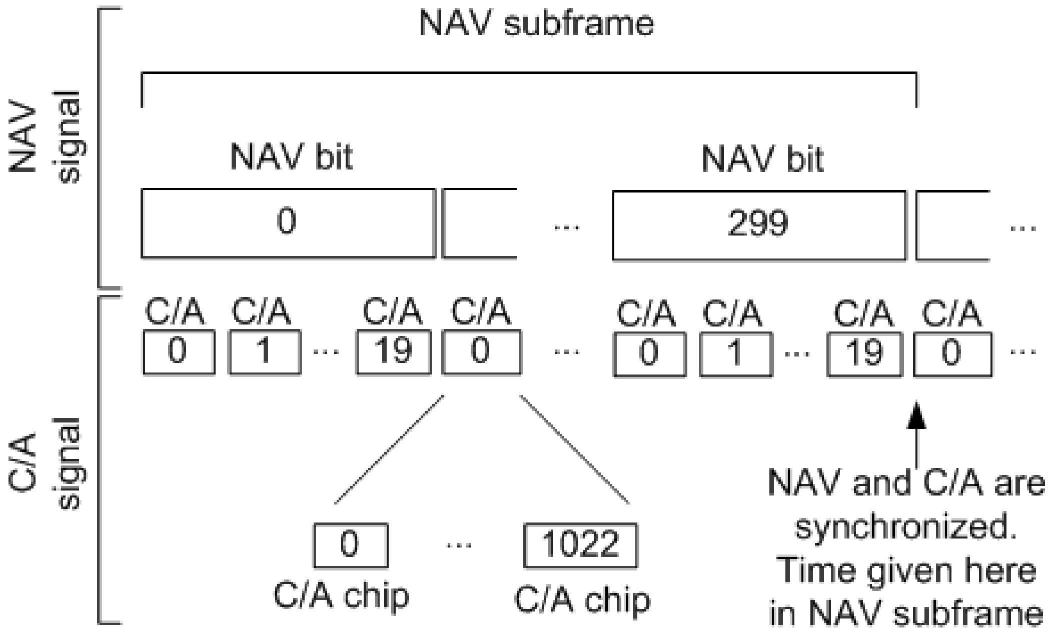

Figure 1.4: C/A NAV timing relationship

Because of this unique time stamp every 6 seconds and that the receiver is continuously tracking the C/A code of the satellite, each chip of a C/A code can be uniquely identified with an exact time of transmission. Therefore, the receiver can resolve the ambiguity M in the code observable and hence can estimate $\tau(t)$ such that $\tau(t) = \Delta t(t) + \Delta T_{RX}(t) - \Delta T_{TX}(t)$.

Usually the code observable is in units of meters and is called the pseudorange. Converting $\tau(t)$ into meters by multiplying by the speed of light c results in the following pseudorange equation where the variable time has been removed for brevity and ρ is the range from the transmitter at transmission time to receive at reception time.

$$p = \rho + c\left(\Delta T_{RX} - \Delta T_{TX}\right) \tag{1.16}$$

1.4.1.1 Calculating code based solutions

The satellite's current clock biasΔT_{TX} is transmitted in the navigation data and therefore is a known value. The unknown values are therefore the receiver's position and clock bias; together these are $\mathbf{P} = [x, y, z, \Delta T_{RX}]$. This means the pseudorange is a function of these unknown variables. Obtaining one such pseudoranges for a satellite results in the following nonlinear equation.

$$p_n\left(\mathbf{P}\right) = \rho_n + c\left(\Delta T_{RX} - \Delta T_{TX}^{n}\right) \tag{1.17}$$

Because of the ease of solving linear equations, linearization of this equation using a first order Taylor expansion is sensible for deriving a generalized method of solving sets of pseudoranges of arbitrary sizes. We let an estimated solution be $\hat{\mathbf{P}} = \left[\hat{x}, \hat{y}, \hat{z}, \Delta\hat{T}_{RX}\right]$ for time of reception. Then, a first order Taylor expansion for p_n around $\hat{\mathbf{P}}$ is as follows.

$$\begin{aligned} p_n\left(\mathbf{P}\right) \approx p_n\left(\hat{\mathbf{P}}\right) &+ \left(x - \hat{x}\right) \frac{\partial}{\partial x} p_n \bigg|_{\hat{\mathbf{P}}} + \left(y - \hat{y}\right) \frac{\partial}{\partial y} p_n \bigg|_{\hat{\mathbf{P}}} \\ &+ \left(z - \hat{z}\right) \frac{\partial}{\partial z} p_n \bigg|_{\hat{\mathbf{P}}} + \left(\Delta T_{RX} - \Delta\hat{T}_{RX}\right) \frac{\partial}{\partial \Delta T_{RX}} p_n \bigg|_{\hat{\mathbf{P}}} \end{aligned} \tag{1.18}$$

Given a satellite's position $\mathbf{S}_n = [x_n, y_n, z_n]$ at time of transmission for a pseudorange p_n, then, the partial derivatives can be calculated given $\rho_n\left(x, y, z\right) = \sqrt{\left(x - x_n\right)^2 + \left(y - y_n\right)^2 + \left(z - z_n\right)^2}$. Therefore, upon evaluation, equation 1.18 can be written as follows.

$$p_n(\mathbf{P}) - p_n\left(\hat{\mathbf{P}}\right) \approx (x - \hat{x})\frac{(\hat{x} - x_n)}{\hat{\rho}_n} + (y - \hat{y})\frac{(\hat{y} - y_n)}{\hat{\rho}_n} + (z - \hat{z})\frac{(\hat{z} - z_n)}{\hat{\rho}_n} + \left(\Delta T_{RX} - \Delta\hat{T}_{RX}\right)c \quad (1.19)$$

Given a good estimate $\hat{\mathbf{P}}$, this equation has four independent unknowns, hence at least four equation just like it are needed to solve the unknowns. These four equations require the receiver's clock bias to be the same for all equations and the receiver's position to be the same for all equations. Therefore, the receiver has to obtain four pseudoranges simultaneously. A set of m such pseudoranges for different satellites obtained simultaneously are given below.

$$\begin{aligned} p_1(\mathbf{P}) &= \rho_1 + c\left(\Delta T_{RX} - \Delta T_{TX}^1\right) \\ &\vdots \\ p_m(\mathbf{P}) &= \rho_m + c\left(\Delta T_{RX} - \Delta T_{TX}^m\right) \end{aligned} \quad (1.20)$$

Because equation 1.19 is linear this set of pseudo ranges can be written in matrix form as follows.

$$\begin{bmatrix} \frac{(\hat{x}-x_1)}{\hat{\rho}_1} & \frac{(\hat{y}-y_1)}{\hat{\rho}_1} & \frac{(\hat{z}-z_1)}{\hat{\rho}_1} & c \\ \vdots & \vdots & \vdots & \vdots \\ \frac{(\hat{x}-x_m)}{\hat{\rho}_m} & \frac{(\hat{y}-y_m)}{\hat{\rho}_m} & \frac{(\hat{z}-z_m)}{\hat{\rho}_m} & c \end{bmatrix} \begin{bmatrix} (x-\hat{x}) \\ (y-\hat{y}) \\ (z-\hat{z}) \\ \left(\Delta T_{RX} - \Delta\hat{T}_{RX}\right) \end{bmatrix} \approx \begin{bmatrix} p_1(\mathbf{P}) - p_1\left(\hat{\mathbf{P}}\right) \\ \vdots \\ p_m(\mathbf{P}) - p_m\left(\hat{\mathbf{P}}\right) \end{bmatrix} \quad (1.21)$$

Using bold type notation vectors or matrices, this has the form of $\mathbf{A}\left(\mathbf{P} - \hat{\mathbf{P}}\right) \approx \mathbf{b}$ where $\mathbf{A}$, $\mathbf{b}$, and $\hat{\mathbf{P}}$ are known. Using LS and rearranging for the unknown yields $\mathbf{P} \approx \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b} + \hat{\mathbf{P}}$. The right-hand side of this equation when calculated yields only an approximation of the desired solution $\mathbf{P}$. Therefore $\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b} + \hat{\mathbf{P}}$ is another solution estimate and we denote it as $\hat{\mathbf{P}}_n = \left[\hat{x}_n, \hat{y}_n, \hat{z}_n, \Delta\hat{T}_{RX_n}\right]$. So we have seen a way of obtaining a new solution estimate from an old solution estimate. This process can be reiterated as the following algorithm describes.

Algorithm 1.1 Iterative LS solution using code observable

1. $\hat{\mathbf{P}}_0 = [0, 0, 0, 0]$, $n = 0$
2. increment n
3. estimate reception time $\hat{t}_{rx} = T_{RX} - \Delta\hat{T}_{RX}$
4. calculate $\mathbf{S}_m$ at time $\hat{t}_{rx}$
5. calculate $\mathbf{S}_m$ at time $\hat{t}_{rx} - \|\mathbf{S}_m - [\hat{x}_n, \hat{y}_n, \hat{z}_n]\| / c$ and reitterate a few times
6. if $\nexists \left(\mathbf{A^T A}\right)^{\mathbf{-1}}$ stop wih error
7. $\hat{\mathbf{P}}_n = \left(\mathbf{A}^T\mathbf{A}\right)^{-1} \mathbf{A}^T\mathbf{b} + \hat{\mathbf{P}}_{n-1}$
8. if $(\|[\hat{x}_n, \hat{y}_n, \hat{z}_n] - [\hat{x}_{n-1}, \hat{y}_{n-1}, \hat{z}_{n-1}]\| <$ less than desired error$)$ stop
9. if $(n >$ to big $)$ stop with error
10. goto 2

The correction made to the time of reception as believed by the receiver T_{RX} to produce an estimate of the reception time in step 3, for most receivers would be small in the order of less than a millisecond and could be conceivably ignored. The correction needed for calculating an estimate of the time of transmission as performed in step 5 is generally comparatively large, and is in the order of around $60\,ms$ being the approximate flight time from the satellite to the receiver; this step can't be ignored.

Figure 1.5 shows an example of algorithm 1.1 converging for a set of six satellites and code observations taken of them. As can be seen using the center of the Earth as the initial estimate within six iterations the algorithm has converged.

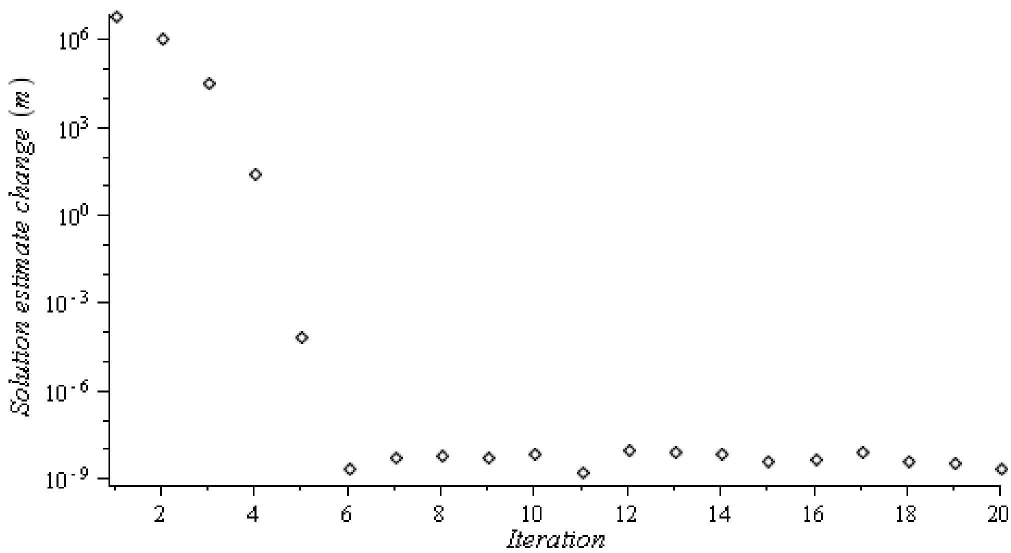

Figure 1.5: Example of convergence of algorithm 1.1

The output of this algorithm for the example in figure 1.5 resulted in a receiver clock bias estimate of $-116.644\,\mu s$. This compares to an estimate as computed by the receiver itself of $-116.738\,\mu s$ with a spatial solution discrepancy of $32\,m$ between the two.

1.4.1.2 Final code observable model

One reason why the discrepancy of $32\,m$ between our solution and the solution as calculated by the GPS receiver itself is we have neglected some things in our modeling of equation 1.16. The model in equation 1.16 can be extended by adding tropospheric delays T, ionospheric delays I, multipath M_r and miscellaneous errors e_r. Therefore, a more exact model of the phase observable can be written as in the following equation.

$$p = \rho + c\,(\Delta T_{RX} - \Delta T_{TX}) + T + I + M_r + e_r \qquad (1.22)$$

Of the sources of error ionospheric are typically the greatest. The ionosphere stretches from 50 km to 1000 km above the Earth consisting mainly of charged particles, charged atoms and charged molecules. A large part of the ionization is caused by *UltraViolet* (UV) rays from the sun and hence there is a large diurnal change in the *Total Electron Count* (TEC) which in turn effects the ionospheric correction term I. The ionosphere can produce a satellite range error as little as 1 m to as much as 100 m [1]. The ionospheric correction term is frequency dependent and with dual band receivers ionospheric free combinations of observables are possible. For singleband receivers no such combination is possible, instead Klobuchar ionospheric model is used for singleband receivers GPS. Klobuchar coefficients are transmitted in the navigation message so that the receiver can then estimate ionospheric correction terms. The Klobuchar algorithm corrects about 50% of the ionospheric errors [5].

The important thing for us in this section is that we are able to obtain an approximate solution, the spatial component with some sort of accuracy less than $100\,m$, and a time accuracy of some sort less than a $1\,\mu s$. This is all we are concerned about regarding the code observable.

1.4.2 The phase observable

The phase observable is a measured quantity taken by the GPS receiver for a particular satellite for a particular time. Phase observables allows higher accuracy GPS measurements to be made than compared to that of ones solely using code observables. This is due to the much shorter wavelength of GPS carrier than compared to the chip length of the code observable. The wavelength of L1 is approximately 20 cm compared to approximately 300 m length for the code chip of the C/A signal and can result in a correspondingly large increase in accuracy. The measurement comes from monitoring the phase difference between the received satellite carrier and a reference oscillator on the GPS receiver. The receiver accumulates this instantaneous phase difference by tracking and outputs this to the user as the phase observable. Figure 1.6 shows a block diagram of what the GPS receiver is doing when observing a satellite for the phase observable neglecting all signals sent on the carrier such as C/A, navigation and P(Y) code.

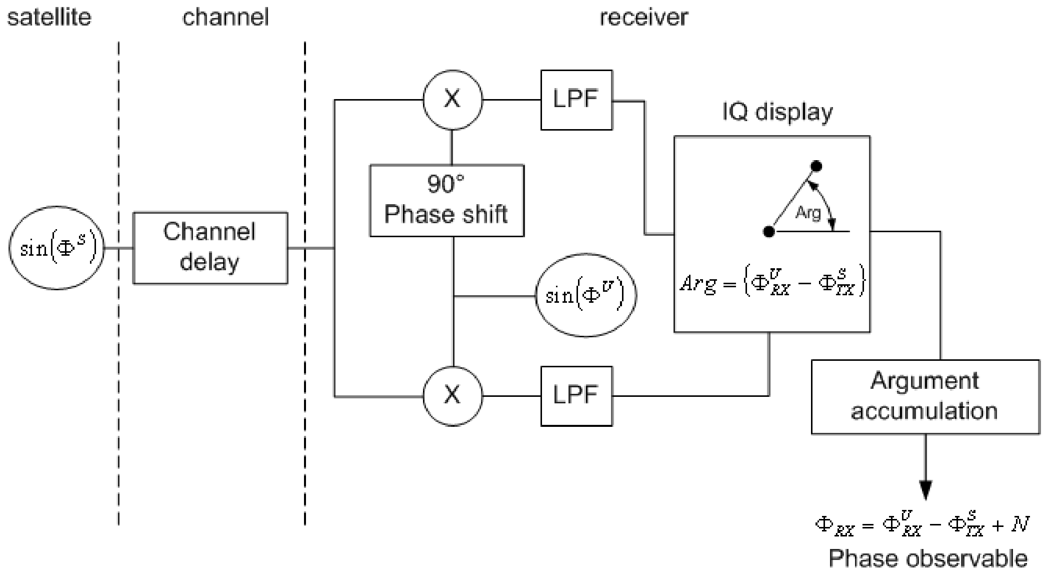

Figure 1.6: Simplified block diagram of phase measurement

For GPS, phase is customary in units of cycles rather than radians or degrees for GPS work. Phase is the argument inside a trigonometric function that accepts units of cycles. Φ^S is the phase of the carrier of the satellite while Φ^U is the receiver's reference phase. Both Φ^S and Φ^U can become arbitrarily large.

Inherent in accumulation of phase is an ambiguity N that depends on when you started accumulating phase. In addition to the ambiguity there is the possibility of missing some rotations. Counting the number of times a car tire rotates depends on when you started counting its rotations and also depends on whether you missed any rotations. While GPS receivers try to continuously track the phase, this is not always possible. Due to noise, loss of signal or turning the GPS receiver off and on again, the tracking of the phase can be lost resulting in an integer change in the value of N. This produces what is called a "cycle slip". So, ideally this phase ambiguity should be fixed while the satellite is being tracked but due to cycle slips occasionally it will change.

As we have seen satellites don't send out continuous waves, they are modulated with two BPSK signals, one in the quadrature phase and the other in the in-phase. Code tracking has the effect of wiping the C/A code from one of the BPSK signals but still leaves the navigation adding a level of complexity when trying to track it. When the BPSK data is not used to regenerate the original carrier wave a half cycle ambiguity in the carrier phase is introduced into the phase observable and it is said that the phase observable has a code factor of two $C_f = 2$. When the BPSK data is used to regenerate the original carrier wave, the original carrier wave can be fully regenerated with an ambiguity of one cycle and it is said that the phase observable has a code factor of two $C_f = 1$. Thus the ambiguity of the phase observable can be reduced when a code factor of one is used.

1.4.2.1 Final phase observable model

As we have seen tracking the carrier phase maximizes γ which in turn means the accumulated phase while tracking is $\Phi = t\triangle f + \varDelta t f_{TX} + \triangle\phi_0 + N/2$ for some fixed integer N when using a costas PLL. When considering the code factor this can be written as follows.

$$\Phi = t\triangle f + \varDelta t f_{TX} + \triangle\phi_0 + N/C_f \tag{1.23}$$

Assuming the receiver's clock is based around its local oscillator and its frequency keeps perfect time, true GPS time can be converted into the time as determined by the receiver as $T_{RX} = \left(t f_{RX} + \phi_0^{RX}\right)/f_0$. By definition true time plus clock bias is also the time as determined by the receiver $T_{RX} = t + \Delta T_{RX}$. Equating the two and rearranging yields the clock bias in term of true GPS time. Likewise this can be done for the satellite's clock.

$$\Delta T_{RX} = \left(t f_{RX} + \phi_0^{RX}\right)/f_0 - t \tag{1.24}$$

$$\Delta T_{TX} = \left(t f_{TX} + \phi_0^{TX}\right)/f_0 - t \tag{1.25}$$

Subtracting the two and multiplying by f_0 results in the following.

$$f_0\left(\Delta T_{RX} - \Delta T_{TX}\right) = t\Delta f + \Delta\phi_0 \tag{1.26}$$

Equating this with equation 1.23 we see that we can write equation 1.23 as follows

$$\Phi = f_0\left(\Delta T_{RX} - \Delta T_{TX}\right) + \varDelta t f_{TX} + N/C_f$$

$$\Rightarrow \Phi = f_0 \left(\Delta T_{RX} - \Delta T_{TX} \right) + \frac{1}{\lambda_0} \rho f_{TX} / f_0 + N / C_f \tag{1.27}$$

f_{TX} is the oscillator of the GPS transmitter and is an atomic clock being extremely close to f_0. Therefore $f_{TX}/f_0 = 1$ for our purposes and we can rewrite equation 1.27 as follows.

$$\Phi = f_0 \left(\Delta T_{RX} - \Delta T_{TX} \right) + \rho / \lambda_0 + N / C_f \tag{1.28}$$

This model in equation 1.28 can be extended by adding tropospheric delays T, ionospheric delays I, multipath M_ϕ and miscellaneous errors e_ϕ as was done with the code observable. However, the ionospheric correction for the phase while being of the same magnitude of that of the code observable is of the opposite sign. Therefore, a more exact model of the phase observable can be written as in the following equation.

$$\Phi = f_0 \left(\Delta T_{RX} - \Delta T_{TX} \right) + \left(\rho + T - I + M_\phi + e_\phi \right) / \lambda_0 + N / C_f \tag{1.29}$$

The range term ρ in equation 1.29 is for the receiver at reception time t_{RX} and the satellite at transmission time of t_{TX}. So the distance ρ is a measure of where you are to where the satellite was a short period of time ago because t_{TX} is an earlier time than the current time of t_{RX}; the difference between these two values is typically around the 60 ms mark and a satellite can move a few hundred meters in this time.

The phase observable was measured at GPS time t_{RX}, this variable itself has to be solved for, as you are not going to know exactly what time the measurement was performed; you know you performed a measurement now, but you don't know when now is. This can be obtained using the code observable as previously shown.

The receiver's clock bias is at t_{RX} while the satellite's clock bias is at t_{TX}, but these aren't so critical as these don't change rapidly over 60 ms and can safely be assumed to be constant over the short time periods.

Equation 1.29 is our final model for the phase observable. The left-hand side is what the receiver gives us, while the right-hand side is what we interpret it as.

Multiplying it by the satellite's nominal wavelength is still classified as the phase observable but rather than units of cycles, the units become meters.

1.5 Selected proofs

1.5.1 Received phase using flight time approximation.

Taylor expantion of flight time

$\mathit{\Delta} t\left(t\right)=\sum\limits_{k=0}^{\infty}\frac{\mathit{\Delta} t^{(k)}(t_m)}{k!}\left(t-t_m\right)^k$

$\mathit{\Delta} t\left(t\right)=\mathit{\Delta} t\left(t_m\right)+\left(t-t_m\right)\frac{v(t_m)}{c}+\left(t-t_m\right)^2\frac{1}{2}\frac{a(t_m)}{c}+\cdots$

First order linear approximation

$\mathit{\Delta} t\left(t\right)\approx\mathit{\Delta} t\left(t_m\right)+\left(t-t_m\right)\frac{v(t_m)}{c}$

Define received phase

$\Phi\left(t\right)=t\triangle f-\mathit{\Delta} t\left(t\right)f_{TX}+\triangle\phi_0$

$\rightarrow\Phi\left(t\right)\approx t\triangle f-\left(\mathit{\Delta} t\left(t_m\right)+\left(t-t_m\right)\frac{v(t_m)}{c}\right)f_{TX}+\triangle\phi_0$

$\rightarrow\Phi\left(t\right)\approx t\left(\triangle f-\frac{v(t_m)}{c}f_{TX}\right)+t_m\frac{v(t_m)}{c}f_{TX}-\mathit{\Delta} t\left(t_m\right)f_{TX}+\triangle\phi_0$

$\rightarrow\Phi\left(t\right)\approx t\left(\triangle f-\frac{v(t_m)}{c}f_{TX}\right)+\theta\left(t_m\right)$

$\rightarrow\Phi\left(t\right)\approx t\left(\triangle f-\Delta f_{TX}\right)+\theta\left(t_m\right)\quad\square$

1.5.2 τ First-order linear approximation

$\Delta t\left(t\right)\approx\frac{\Delta f_{TX}(t_m)}{f_{TX}}\left(t-t_m\right)+constant_1$

$\Delta T_{TX}\left(t\right)\approx\frac{f_{TX}(t_m)-f_0}{f_0}\left(t-t_m\right)+constant_2$

$\Delta T_{RX}\left(t\right)\approx\frac{f_{RX}(t_m)-f_0}{f_0}\left(t-t_m\right)+constant_3$

$\rightarrow\tau\approx\left(t-t_m\right)\left(\Delta f-\Delta f_{TX}\left(t_m\right)\right)\frac{-1}{f_0}+\mathit{\Xi}\left(t_m\right)$ as $f_{TX}\approx f_0$

1.5.3 Radial velocity with constant radial velocity offset ϵ

What happens with constant velocity offset.

If radial velocity is out by ϵ at all times then

$$\lim_{t\to t_m}\Phi\left(t\right)=t\triangle f-\left(\mathit{\Delta} t\left(t_m\right)+\left(t-t_m\right)\tfrac{v(t_m)}{c}\right)f_{TX}+t\tfrac{\epsilon}{c}f_{TX}+\delta+\triangle\phi_0$$

$$\therefore\lim_{t\to t_m}\Phi\left(t\right)=t\left(\triangle f-\Delta f_{TX}\right)+\theta'\left(t_m\right)+t\tfrac{\epsilon}{c}f_{TX}$$

Where $\theta'\left(t_m\right)$ and δ are some constants

1.5.4 Maximum radial velocity and acceleration of the satellite with respect to the receiver

Figure 1.7 is a simplified model of satellite orbiting the Earth while transmitting to a receiver. No relativistic effects are considered and it is assumes that the satellite's orbit is perfectly circular with constant tangential velocity and when the satellite is closest to the receiver the satellite is directly overhead.

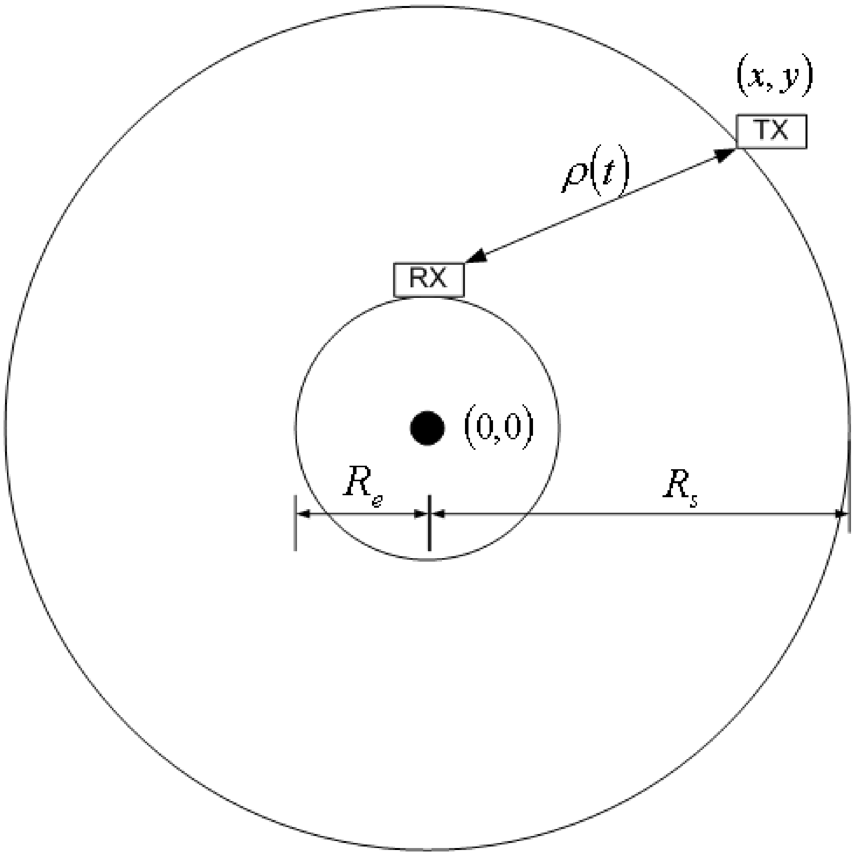

Figure 1.7: Simplified model of a satellite transmitting to a receiver while orbiting

From figure 1.7 we obtain the following equations to describe the system.

$$\begin{aligned} x(t) &= r_s \cos\left(\frac{2\pi t}{T}\right) \\ y(t) &= r_s \sin\left(\frac{2\pi t}{T}\right) - r_e \\ \rho(t) &= \sqrt{x^2(t) + y^2(t)} \end{aligned} \tag{1.30}$$

The following approximate generally recognized values for GPS satellite orbital period T , radius of the earth and radius of GPS satellite orbits were used. By differentiating the range equation 1.30 figure 1.8 was obtained.

$$
\begin{aligned}
r_e &= 6371000\,m \\
r_s &= 26600000\,m \\
T &= 43080\,s
\end{aligned}
\tag{1.31}
$$

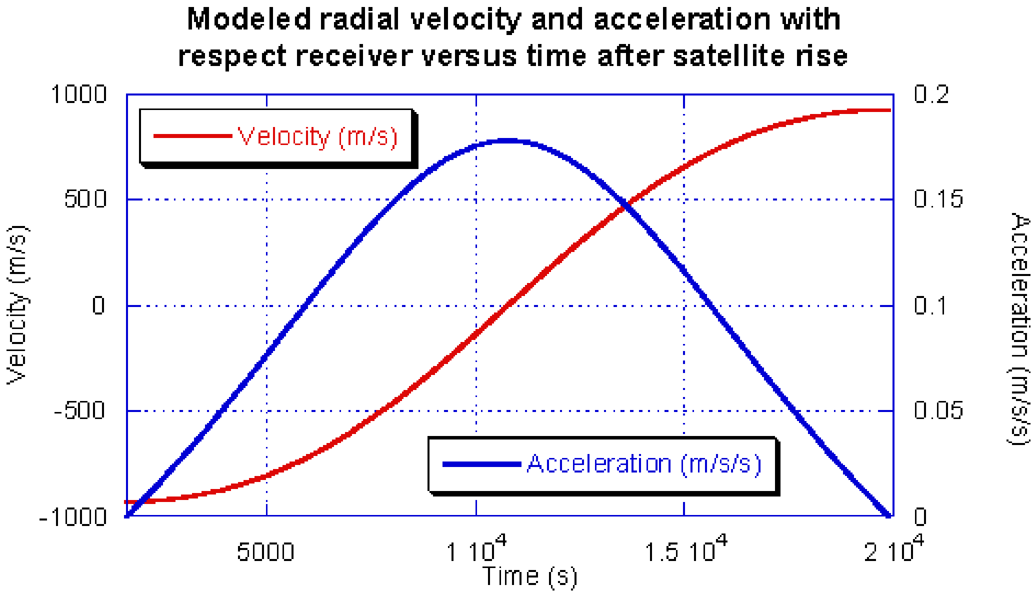

Figure 1.8: Modeled radial velocity and acceleration

The maximum radial velocity derived from the model was $929\,ms^{-1}$ at satellite rise time and set time. Maximum acceleration was $0.1178\,ms^{-2}$ and was when the satellite was directly overhead. [6] (pg 91) states that GPS satellites can have a radial velocity of up to $800\,ms^{-1}$ with respect to a stationary receiver on earth. This is consistent with our simplified model. On the L1 band an acceleration of $0.1178\,ms^{-2}$ is approximately a $0.9\,Hzs^{-1}$ doppler shift rate while $929\,ms^{-1}$ on the L1 band is approximately a 5 kHz doppler shift, therefore $f_0 \pm 5\,kHz$ must be searched for the carrier frequency.

Nomenclature

2D	two-dimensional
BPSK	Binary Phase Shift Keying
C/A	Coarse acquisition code
CDMA	Code Division Multiple Access
Chip	One bit of a PRN code
DLL	Delay Locked Loop
FFT	Fast Fourier Transform
FLL	Frequency Locked Loop
GPS	Global positioning system
HOW	Hand Over Word
LS	Least Squares
NAV	Navigation Data
P	Precise unencrypted code
PLL	Phase Locked Loop
PRN	Pseudo Random Number
RF	Radio frequency
TEC	Total Electron Count
UV	UltraViolet
Y	Precise encrypted code

Bibliography

[1] Klobuchar, j.a. ionospheric effects on gps. gps world, April 1991. Vol. 2, No. 4, pp. 48-51.

[2] Gps.gov: Gps modernization http://www.gps.gov/systems/gps/modernization/. Webpage, 2014. Modified: Monday, 15 September 2014 12:53:11 p.m.

[3] Neo-7,u-blox 7, gnss modules data sheet. http://www.u-blox.com/images/downloads/Product_Docs/NEO-7_DataSheet_(GPS.G7-HW-11004).pdf, May 2014.

[4] Gruber, C. B. Gps modernization and program update. In *Munich Satellite Navigation Summit, Munich, Germany* (2011).

[5] Klobuchar, J. Ionospheric time-delay algorithm for single-frequency gps users. *Aerospace and Electronic Systems, IEEE Transactions on AES-23*, 3 (May 1987), 325–331.

[6] Van Diggelen, F. *A-GPS: Assisted GPS, GNSS, and SBAS*. Artech House Gnss Technology and Applications Library. Artech House, 2009.

[7] Zhou, Y. Dsp in a satellite navigation receiver with a perspective of computational complexity. Internet, Nov 2013.