451-337 Satellite Positioning and Geodesy
Exam Solutions 2006

1. (a) Discuss the distinctions between coordinate conversion and coordinate transformation. (5 marks)

In geodesy, *coordinate conversion* is the process of changing the way in which 3D position can be expressed without changing the frame of reference. Typically, position can be expressed in one of three ways :

- Cartesian coordinates (X,Y,Z)
- Geodetic (geographic) coordinates (φ,λ,h)
- Map grid (UTM) coordinates (E,N,h)

Coordinate conversion is the process of converting from one form of coordinates to another, within a single frame of reference or datum. The process of coordinate conversion may be formulated as follows :

$$(X, Y, Z)_{Datum1} \Leftrightarrow (\varphi, \lambda, h)_{Datum1} \Leftrightarrow (E, N, h)_{Datum1}$$

The process is of course mathematical, utilising formulae such as Redfearn's formulae to convert between geographic and map grid coordinates (and vice versa) and Bowring's formulae to convert between cartesian and geographic coordinates (and vice versa).

Coordinate transformation on the other hand is the process whereby (usually) cartesian coordinates are moved from one datum to another. Typically, though not necessarily, this is achieved via a 7 parameter conformal transformation. The process of coordinate transformation may be formulated as follows :

$$(X, Y, Z)_{Datum1} \Rightarrow (X, Y, Z)_{Datum2}$$

Once in the desired datum, coordinate conversion can then change the form in which the transformed coordinates are expressed.

(b) Dilution of Precision (DOP) values are used to indicate the instantaneous quality of GPS positioning based on the C/A code. Describe how the various DOP factors are computed and how they can be interpreted.

(5 marks)

The various DOP factors used in GPS positioning are derived from the diagonal elements of the inverse of the normal matrix. The normal matrix is computed as part of standard GPS navigation solution. The navigation solution is based on measured C/A-code pseudoranges and solves for the 3D receiver coordinates (X,Y,Z) and the receiver clock offset (dT) using the least squares algorithm. In least squares solutions, the inverse of the normal matrix is, of course, the variance matrix of the estimated parameters and therefore takes the following form :

$$\mathbf{V}_{\hat{x}} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} & \sigma_{XZ} & \sigma_{XdT} \\ \sigma_{XY} & \sigma_Y^2 & \sigma_{YZ} & \sigma_{YdT} \\ \sigma_{XZ} & \sigma_{YZ} & \sigma_Z^2 & \sigma_{ZdT} \\ \sigma_{XdT} & \sigma_{YdT} & \sigma_{ZdT} & \sigma_{dT}^2 \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma_E^2 & \sigma_{EN} & \sigma_{Eh} & \sigma_{EdT} \\ \sigma_{EN} & \sigma_N^2 & \sigma_{Nh} & \sigma_{NdT} \\ \sigma_{Eh} & \sigma_{Nh} & \sigma_h^2 & \sigma_{hdT} \\ \sigma_{EdT} & \sigma_{NdT} & \sigma_{hdT} & \sigma_{dT}^2 \end{bmatrix}$$

The DOP factors are given by :

$$\mathrm{TDOP} = \sigma_{dT}$$

$$\mathrm{VDOP} = \sigma_h$$

$$\mathrm{HDOP} = (\sigma_E^2 + \sigma_N^2)^{1/2}$$

$$\mathrm{PDOP} = (\sigma_E^2 + \sigma_N^2 + \sigma_h^2)^{1/2} = (\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2)^{1/2}$$

$$\mathrm{GDOP} = (\sigma_E^2 + \sigma_N^2 + \sigma_h^2 + c\sigma_{dT}^2)^{1/2} = (\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2 + c\sigma_{dT}^2)^{1/2}$$

Where

TDOP relates to time
VDOP relates to height (vertical)
HDOP relates to 2D position (horizontal)
PDOP relates to 3D position (position)
GDOP cover all four parameters (geometric)

The DOP factors are used to quantify the instantaneous positioning quality in the component of interest. For example, VDOP is used to measure positioning quality when height is of importance. GDOP is the most comprehensive and therefore the most commonly used of the DOP indicators. A good GDOP is typically less than 2. As the GDOP value increase, positioning quality decreases. A GDOP greater than about 6 indicates fairly poor positioning quality.

(c) Describe the problem of GPS antenna phase centre offset and variation and elaborate on how this error can be modelled and/or eliminated.

(10 marks)

When doing GPS positioning, great care is taken to locate the *physical centre* of the GPS antenna in relation to the survey mark or the point for which coordinates are required. The assumption behind this procedure is that the physical centre of the antenna is the place where the data from the satellites is received and measurements are made. This assumption is however incorrect. The reality is that the antenna *phase centre* is the point of measurement and the phase centre and the physical centre do not (necessarily) coincide. To account for this non-coincidence, the spatial relationship between the physical centre and the phase centre must be known. However this relationship is not constant, in fact it is usual to break it down into two components, a constant offset and the spatially changing variation component. The phase centre offset is a constant displacement (in 3D) between the physical centre and the phase centre. But the phase centre also moves as a function of satellite geometry (particularly satellite elevation) and this is the variable component. The following diagram illustrates the situation in 2D, though in reality phase centre offset and variation is a 3D phenomenon and generally more significant in the vertical direction.

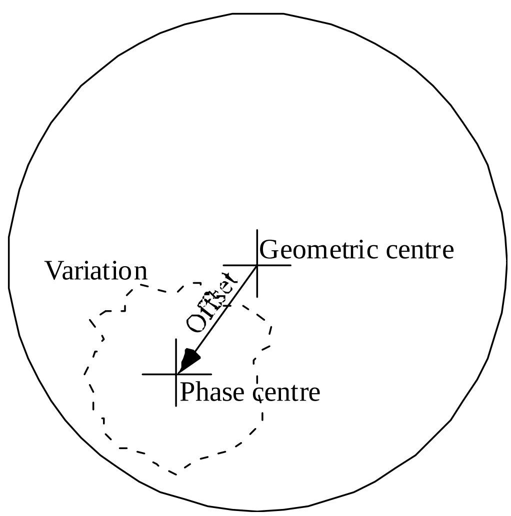


Figure 1 – The relationship between the geometric centre of the antenna and the phase centre, showing the offset and variation components

Dealing with antenna offset and variation can be done in two ways. If both antennas used in the measurement of a baseline are identical, parallel antenna orientation will result in a cancellation of the offset and variation errors through the process of measurement differencing. It is for this reason that GPS antennae are traditionally oriented to the north. The second option, which must be used in the case of using dissimilar antennae, is to employ antenna calibration models to correct the raw measurements for the influence of the error. The National Geodetic Survey (NGS) in the US has a web-site devoted to this problem and supplies relative calibration models for a very wide range of GPS antennae. With the advent of CORS networks, the need to accommodate for the use of different antennae is becoming more common.

2. The observation equation for the GPS carrier phase observable (in units of metres) is given below. Derive this equation, giving a full description of each term and illustrating the derivation with diagrams where appropriate.

$$\Phi = c(dt - dT) + R - d_{ion} + d_{trop} - \lambda N(t_0) + n$$

(20 marks)

Firstly, suppose we know the following information :

t the time of signal transmission from satellite
R the range between receiver and satellite

Assuming the carrier signal travels at the speed of light (c), we can deduce an expression for the time of signal reception at the receiver (T) :

T = t + R/c

Two assumptions have been made in deriving this expression :

- That the transmitted signal is unaffected by the earth's atmosphere
- That the satellite and receiver clocks are perfectly synchronised

Of course neither of these assumptions are valid. The above equation should therefore be modified to account for these assumptions :

$$(T + dT) = (t + dt) + (R - d_{ion} + d_{trop}) / c$$

where :

d_{ion} is the delay caused by the ionosphere on the C/A-code observation. It is negative here because that delay affects the carrier in an equal but opposite way (metres)
d_{trop} is the delay caused by the troposphere on the C/A-code, which is identical to the delay affecting the carrier phase (metres)
dT is the error in the receiver clock (seconds)
dt is the error in the satellite clock (seconds)

We re-arrange the above equation to get an expression for (T-t), such that :

$$(T - t) = (dt - dT) + (R - d_{ion} + d_{trop}) / c$$

For the second stage of the development, suppose we know :

t Time of signal transmission from satellite
f Frequency of the carrier wave
T Time of signal reception at the receiver

We can therefore write an expression for the range between the receiver and the satellite in units of cycles as :

$$\varphi_{total} = f(T - t)$$

Substitution gives:

$$\varphi_{total} = f(dt - dT) + \frac{f}{c}(R - d_{ion} + d_{trop})$$

In terms of what a GPS receiver actually measures, the total phase consists of :

$$\varphi_{total} = Fr(\varphi) + Int(\varphi; t_0, t) + N(t_0) = \varphi_{meas} + N(t_0)$$

where :

$Fr(\varphi)$	a *measured* fractional part
$Int(\varphi; t_0,t)$	a *measured* integer count of complete cycles since the initial epoch t_0
$N(t_0)$	an *unknown* number of integer cycles (*integer ambiguity*) between the satellite and receiver at the initial epoch

These last two equations are equivalent, therefore :

$$\varphi_{meas} + N(t_0) = f(dt - dT) + \frac{f}{c}(R - d_{ion} + d_{trop})$$

Re-arranging gives an expression for φ_{meas} as :

$$\varphi_{meas} = f(dt - dT) + \frac{f}{c}(R - d_{ion} + d_{trop}) - N(t_0)$$

Recall that λ=c/f

Multiply by λ to convert the measured carrier beat phase into units of metres :

$$\Phi = \lambda\varphi_{meas} = c(dt - dT) + R - d_{ion} + d_{trop} - \lambda N(t_0)$$

This equation is the basic GPS carrier phase observation equation. To be entirely rigorous, terms should be added to account for receiver multipath and noise :

$$\Phi = c(dt - dT) + R - d_{ion} + d_{trop} - \lambda N(t_0) + d_{mult} + n$$

Often noise and multipath are combined, leading to the final required form of the equation :

$$\Phi = c(dt - dT) + R - d_{ion} + d_{trop} - \lambda N(t_0) + n$$

3. (a) Give a definition for the *geoid* and explain its role in geodesy as a reference surface for height.

(5 marks)

The geoid is an equipotential surface of the earth's gravity field. An equipotential surface is a surface of equal gravity potential to which the direction of the instantaneous gravity vector is at all points perpendicular. Because the gravity vector is always at right angles to an equipotential surface, this surface is – by definition –a *level* surface. Thus water will not flow across an equipotential surface. The geoid is coincident with the average surface of the open oceans, once the influences of tides, wind, currents and sea surface topography have been accounted for (i.e. the Mean Sea Level surface).

In geodesy, we need a reference surface for height that can reliably determine directions and rates of fluid flow. Thus heights above the reference surface must implicitly map the behaviour of the gravity field since it is the force of gravity that leads to fluid flow. For this reason, heights above an equipotential surface (i.e. a *level* surface such as the geoid) are of most practical significance in surveying, geodesy and engineering applications. It generally follows that given two points with heights above the geoid, water will flow from the higher point to the lower point. It is for this reason that the Australian Height Datum is, in effect, a realisation of the geoid across Australia.

(b) Explain the importance of accounting for the geoid when carrying out precise GPS positioning.

(5 marks)

Precise GPS positioning produces a 3D baseline (vector) between a pair of receivers. The baseline components may be expressed in cartesian or geodetic form as follows :

$$\begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} \Leftrightarrow \begin{bmatrix} \Delta\varphi \\ \Delta\lambda \\ \Delta h \end{bmatrix}$$

Looking at the height component of the vector when expressed in geodetic form (Δh), it is clear that GPS yields the ellipsoidal height difference between the two end points. As explained above, in geodesy, surveying and engineering, orthometric rather than ellipsoidal height differences are what is generally needed. Certainly national heights systems such as the AHD are built on orthometric rather than ellipsoidal heights. Thus to use GPS to give useful and practically meaningful height information, the measured ellipsoidal height difference needs to be converted to an orthometric height difference. This can be done by accounting for the relative geoid undulation as shown by the following equation :

$$\Delta H = \Delta h - \Delta N$$

Rigorously accounting for ΔN will allow GPS to be used for levelling (i.e. to obtain orthometric heights) and will also allow GPS to be combined with conventional data (e.g. levelling) in a single network adjustment.

(c) Define the term *integer ambiguity*. What is *ambiguity resolution* and why is it important in GPS data processing?

(5 marks)

In GPS carrier phase positioning, the *integer ambiguity* (usually denoted by N) is the number of full cycles between the receiver and the transmitting satellite at the time the receiver begins to track the signal - i.e. at the initial epoch. This quantity appears in the GPS carrier phase observation equation developed for Question 2 as $N(t_0)$.

Ambiguity resolution refers to the process that allows the integer ambiguity to be determined computationally through the processing of GPS carrier phase data. While carrier phase measurements themselves are very precise, they are ambiguous in as much as the number of full cycles can never be measured. In order to turn the carrier phase measurement into an non-ambiguous range requires that the integer ambiguity be determined. This can only be done mathematically as part of the solution to determine receiver coordinates. The quality of the derived receiver coordinates will be explicitly linked to the success or otherwise of the ambiguity resolution process. Reliable ambiguity resolution implies the determination of accurate and precise receiver satellite ranges and therefore optimum receiver coordinates. If uncertainty exists as to the integer ambiguities, the ranges will be inaccurate and the receiver coordinates will be compromised accordingly.

The real key to ambiguity resolution is to estimate the ambiguities as integer not as float or real numbers. This is known as the fixed solution. The fixed solution yields optimum receiver coordinates. The objective is to resolve the integer ambiguities as reliably and as efficiently (minimum data) as possible.

(d) When it comes to dealing with GPS errors, what is meant by the term *spatial correlation*? What benefit does spatial correlation bring to the problem of GPS error mitigation? Provide a list of the spatially correlated errors that affect GPS positioning and a second list of those errors which are not spatially correlated?

(5 marks)

Spatial correlation is the term that is used to explain the fact that in relative GPS positioning (whether using the C/A-code or the carrier phase) a number of the errors that affect the measurement process are "similar" between both receivers. The degree of similarity between many of these spatially correlated errors is a function of the distance of separation. Thus the closer together the receivers are, the more correlated (similar) the errors will be and the more they will tend to cancel in relative positioning.

The fact that many GPS errors are spatially correlated is of great practical significance when attempting to mitigate the influence of these errors. By positioning one receiver with respect to another (i.e. by doing relative rather than absolute positioning) the influence of spatially correlated errors can be either totally eliminated (in the case that the errors are identical) or minimised (in the case that they are similar). It is on account of the strong spatial correlation in GPS errors that significant accuracy improvements can be achieved in GPS positioning by using a relative rather than an absolute approach. For example, using the C/A-code, absolute positions are typically accurate to ±10 m. In relative mode, through the partial mitigation of spatially correlated errors, accuracy improves to ±1 m or better.

Spatially correlated errors affecting GPS positioning include :

- Satellite orbit
- Satellite clock
- Ionospheric refraction
- Tropospheric refraction

Those errors that are not spatially correlated include :

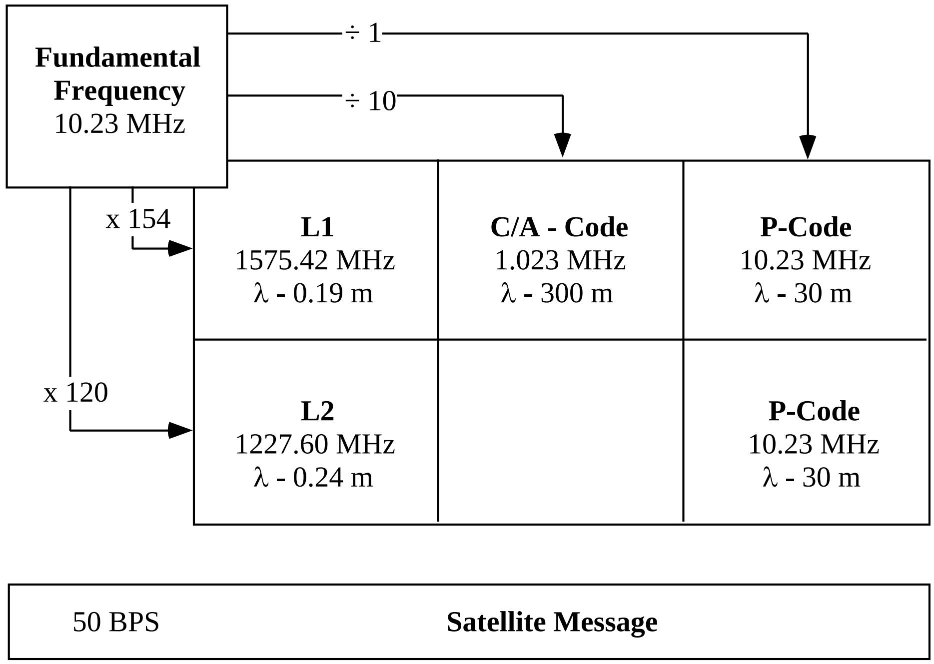
- Receiver clock
- Receiver multipath
- Antenna phase offset and variation (presuming different antenna types)
- Receiver noise

4. (a) Provide details of the GPS signal structure including codes, carriers and satellite message. Describe the use of each component of the signal in terms of positioning applications.

(10 marks)

The structure of the signals transmitted by each GPS satellite is shown in the following chart. The following comments can be made in relation to the details in the chart :

- The "Fundamental Frequency" relates to the on-board satellite oscillator from which all carriers and codes are generated. The oscillator runs at 10.23 MHz and each derived signal is a integer of fractional multiple of this fundamental frequency
- Each satellite transmits two sinusoidal L-band carrier signals known as L1 and L2. The L1 carrier has a wavelength of about 0.19 metres and is derived from the fundamental frequency through a multiplier of 154. Similarly, the L2 carrier has a wavelength of 0.24 metres and is derived from the fundamental frequency through a multiplier of 120.
- The carriers are used for precise relative positioning, mainly for surveying and geodetic purposes. Measurement of the carriers provides ambiguous ranges, requiring the resolution of the integer ambiguities (described above) to attain high accuracy positioning results
- The carriers have modulated onto them pseudo-random noise (PRN) codes and the Satellite Message. The PRN codes are binary streams that appear to be random (hence the name), but are in fact very well defined. The C/A-code (coarse acquisition) is modulated onto the L1 carrier and the P-code (precise or protected) is modulated onto both L1 and L2.
- The C/A-code has a wavelength of approximately 300 metres and a repetition period of 1 millisecond (300 km). It is derived from the fundamental frequency through a multiplier of 0.1. The C/A-code forms the basis of the Standard Positioning Service (SPS) intended for civilian use, providing an autonomous accuracy of about ±10 metres. Since each satellite has its own C/A-code, this code allows the identification of each satellite and the measurement of unambiguous (though somewhat coarse) ranges
- The P-code has a wavelength of approximately (30 metres) and a repetition period of 7 days. In fact each satellite has a unique 7-day segment of the P-code which has a repetition period of 266 days. The P-code forms the basis of the Precise Positioning Service (PPS) which is restricted to authorised users of the system. The PPS offers autonomous positioning accuracy at the 1-2 metre level. To prevent the generation of copy-cat signals and restrict access to the P-code to authorised users, the P-code is encrypted with a W-code to produce the Y-code. This encryption system is known as Anti-Spoofing (AS).
- The Satellite Message is a 50 bps stream of data that allows the space segment to communicate fundamentally important information to users. The satellite message consists of a 30 second frame, made up of 5 sub-frames. The first three sub-frames contain information unique to the transmitting satellite such as health status, clock coefficients and broadcast ephemeris. Sub-frames 4 and 5 contain constellation-wide information such as the satellite almanac, clock parameters, ionospheric model parameters and the UTC-GPS time offset. 25 frames are required to transmit information for the full constellation. of 24 satellites.



Fundamental
Frequency
10.23 MHz
÷ 1
÷ 10
x 154
x 120
L1
1575.42 MHz
λ - 0.19 m
C/A - Code
1.023 MHz
λ - 300 m
P-Code
10.23 MHz
λ - 30 m
L2
1227.60 MHz
λ - 0.24 m
P-Code
10.23 MHz
λ - 30 m
50 BPS
Satellite Message

(b) Describe the function and operation of the GPS Control Segment.

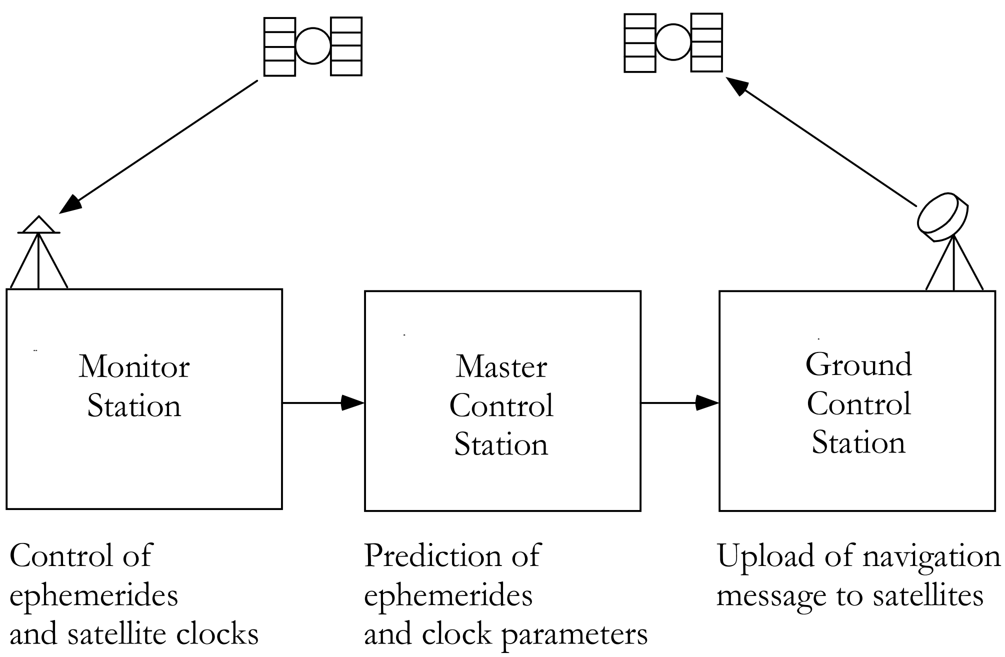
(10 marks)

The GPS Control Segment provides the operational heart of the Global Positioning System. The Control Segment continuously monitors the status of orbiting satellites, collecting data to allow the satellite clocks to be mathematically synchronised to GPS system time, satellite orbit parameters to be calculated, satellite health to be verified and ionospheric model information to be determined and disseminated. The Control Segment also allows satellite orbit manoeuvres and maintenance to be carried out as required.

The GPS Control Segment consists of five ground based tracking stations at Colorado Springs (Falcon Air Force Base), Diego Garcia, Hawaii, Kwajalein and Ascension Island. The station at Colorado Springs is the Master Control Station (MCS). Data from all other sites is sent to the MCS where all data manipulation and computations are performed. Data for uploading to the satellites is then sent back to the three up-link stations for transmission to the GPS satellites. This data is provided to users as part of the Satellite Message.

From a user's perspective, the most important function filled by the GCS is to compute and disseminate satellite orbit and clock parameters. Thus the broadcast ephemeris (consisting of the 16 parameters that define the perturbed Keplerian orbit for each satellite) must be determined and uploaded to satellites on a regular basis. User positioning quality (particularly in autonomous mode) is inherently linked to the quality of the satellite ephemerides and satellite clock parameters.

The following diagram shows the flow of data within the GCS and back to the satellites :



5. (a) Explain why *measurement differencing* is used to aid the processing of GPS carrier phase data. Include a discussion of the advantages and disadvantages of measurement differencing. Develop the double difference carrier phase equation to demonstrate the key points of your explanation.

(10 marks)

As can be seen from the GPS carrier phase observation equation (developed for Question 2), there are many errors influencing the measurement of the carrier phase observable. Here we will re-write the equation in a slightly modified and simplified form, noting that (i) is the receiver and (p) is the satellite, noise and multipath have been combined and the ionospheric and tropospheric delays have been denoted by (I) and (T) respectively rather than d_{ion} and d_{trop}.

$$\Phi_i^p = cdt^p - cdT_i + R_i^p - I_i^p + T_i^p - \lambda N_i^p + n_i^p$$

Because many of the errors in carrier phase positioning are spatially correlated, measurement differencing is introduced as a strategy to eliminate or at least minimise the influence of these errors on the solution. For example, we will introduce a second receiver (j) also tracking satellite (p). The observation equation can be written as :

$$\Phi_j^p = cdt^p - cdT_j + R_j^p - I_j^p + T_j^p - \lambda N_j^p + n_j^p$$

Forming the ***between receiver*** single difference gives :

$$\Delta\Phi_{ij}^p = -c\Delta dT_{ij} + \Delta R_{ij}^p - \Delta I_{ij}^p + \Delta T_{ij}^p - \lambda\Delta N_{ij}^p + \Delta n_{ij}^p$$

If the same two receivers track a second satellite (q), a second ***between receiver*** single difference can be formed as follows :

$$\Delta\Phi_{ij}^q = -c\Delta dT_{ij} + \Delta R_{ij}^q - \Delta I_{ij}^q + \Delta T_{ij}^q - \lambda\Delta N_{ij}^q + \Delta n_{ij}^q$$

Forming these single differences has eliminated the satellite clock errors dt^p and dt^q and to reduce the impact of the ionosphere and troposphere due to spatial correlation.

The ***double difference*** is the difference between two single differences. Thus if we take the above between-receiver single differences, we can form a double difference as follows (note the same could be done with two between-satellite single differences) :

$$\Delta\nabla\Phi_{ij}^{pq} = \Delta\nabla R_{ij}^{pq} - \Delta\nabla I_{ij}^{pq} + \Delta\nabla T_{ij}^{pq} - \lambda\Delta\nabla N_{ij}^{pq} + \Delta\nabla n_{ij}^{pq}$$

Taking the double difference has effectively eliminated the receiver clock errors and has further reduced the impact of the ionosphere and troposphere.

Development of the double difference observation equation shows both the advantages and disadvantages of measurement differencing. These will now be summarised below :

Advantages :

- ☞ Receiver clock errors are eliminated
- ☞ Satellite clock errors are eliminated
- ☞ Ionospheric errors are reduced (minimised)
- ☞ Tropospheric errors are reduced (minimised)

Disadvantages :

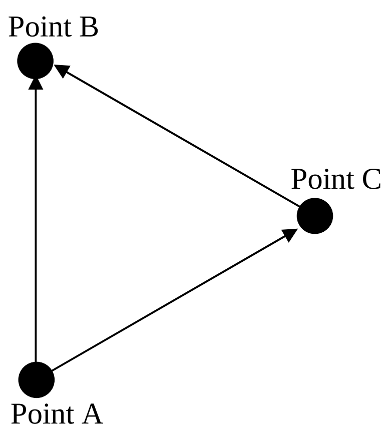
- ☞ Noise propagates (increases) because it is a random quantity
- ☞ Four individual carrier phase observations are needed for one DD (thus redundancy is reduced or more data must be collected)

(b) As an employee of a large surveying company, you have been given four GPS receivers and have been told that you should compute six baselines from every observation session. Obviously the issue of *trivial baselines* has never been explained to your employer. You must prepare a written report explaining the concept of trivial baselines and justifying why they should not be used in the adjustment of GPS networks.

(10 marks)

For a configuration of n simultaneously observing receivers, it is possible to compute ½(n (n-1)) baselines. Thus with four receivers, a set of six different baselines can be computed. However, as will be shown, only three (n-1) of these baselines will be truly independent observations, the other three are known as *trivial baselines* and should not be used.

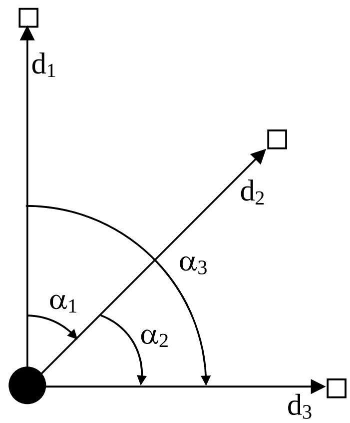
The issue of trivial baselines is easiest to explain by considering a configuration of three simultaneously observing receivers as shown in the diagram below.



Though correlated through the use of common data from A, vectors AB and AC are independent measurements. However CB is not independent, it can be derived from AB and BC as follows :
CB = -AC + AB

We must only use independent observations in any least squares adjustment. Thus to rigorously incorporate vector CB would cause the adjustment to fail due to a singular (non-invertible) measurement variance matrix.

This issue will now be further explained by considering the example of correlated angles derived from a set of observed directions :



From the three measured directions, three angles can be derived as follows :

$\alpha_1 = d_2 - d_1$
$\alpha_2 = d_3 - d_2$
$\alpha_3 = d_3 - d_1$

However, α_3 is not an independent observation since :

$\alpha_3 = \alpha_1 + \alpha_2$

The variance matrix for the three derived angles can be computed by propagation of variances as follows :

$$\mathbf{V}_\alpha = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_d^2 & 0 & 0 \\ 0 & \sigma_d^2 & 0 \\ 0 & 0 & \sigma_d^2 \end{bmatrix} \begin{bmatrix} -1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2\sigma_d^2 & -\sigma_d^2 & \sigma_d^2 \\ -\sigma_d^2 & 2\sigma_d^2 & \sigma_d^2 \\ \sigma_d^2 & \sigma_d^2 & 2\sigma_d^2 \end{bmatrix}$$

This variance matrix is singular since the last row is not independent. If incorporated into a least squares adjustment, the adjustment would fail when attempting to invert this matrix.

The same is true for GPS baselines. If, using the three baseline example above, a rigorous (9x9) variance matrix was produced for the three baselines, the variance matrix would be singular and any adjustment including that data would fail.

This is the reason why trivial baselines should not be included. If a non-rigorous variance matrix was used (as is often the case) the adjustment will succeed, but the adjustment regards the measurements as independent. The statistics will be overly optimistic as a result. Great care needs to be taken if this approach is adopted. By far the best and most rigorous solution is not to use trivial baselines.